

RESEARCH

Open Access



How teacher talk guidance during Invention activities shapes students' cognitive engagement and transfer

Catherine C. Chase^{*} , Jenna Marks, Laura J. Malkiewich and Helena Connolly

Abstract

Background: A key question in K-12 STEM education is how best to guide students as they engage in exploratory learning activities so that students develop transferable knowledge. We investigated this question in a study of teacher talk guidance of an exploratory activity called Invention. In this study, teachers worked one-on-one with students, guiding them as they attempted to invent ratio-based equations of physical science phenomena. We applied the interactive, constructive, active, and passive (ICAP) framework as a theoretical lens through which to explore different forms of teacher talk guidance and resulting student talk. The ICAP hypothesis predicts that constructive engagement leads to greater learning than active engagement, which in turn leads to greater learning than passive engagement. However, students do not always enact the type of cognitive engagement that teachers prompt. In this paper, we work towards three goals: (1) to explore the forms of cognitive engagement prompted by teachers and enacted by students in their talk, (2) to test the ICAP hypothesis in the novel context of teacher-student dialog during Invention, and (3) to identify effective forms of teacher talk guidance for Invention activities and other exploratory STEM learning tasks.

Results: While the majority of student talk was active, teachers produced an even distribution of constructive, active, and passive prompts. Teacher and student talk types tended to align, such that students often responded with the type of cognitive engagement teachers invited, with the exception of passive talk. In general, teacher talk showed the most robust relationship with students' abilities to transfer, while teacher-student dialog demonstrated a weaker relationship with transfer, and student talk was not significantly related to transfer. Some evidence for the ICAP hypothesis was found, most prominently in teacher talk, where constructive prompts positively predicted transfer, active prompts were not related to transfer, and passive prompts negatively predicted transfer.

Conclusions: This research implies that teachers should use a large proportion of constructive prompts and relatively few passive ones when guiding students through Invention tasks, when the goal is to provoke transfer of learning to novel contexts. This work also extends the CAP portion of the ICAP hypothesis to teacher-student dialog and underscores the teacher's critical role in encouraging students to cognitively engage with exploratory STEM tasks in effective ways.

Keywords: Transfer, Dialog, Teacher guidance, STEM learning, Tutoring, Exploration

* Correspondence: chase@tc.columbia.edu

Department of Human Development, Teachers College, Columbia University,
525 West 120th St, New York, NY 10027, USA

Introduction

A broad question in K-12 STEM (science, technology, engineering, and math) education is how best to support exploratory learning activities such as those focused on discovery (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; de Jong & van Joolingen, 1998; Mayer, 2004) and construction (Kafai & Resnick, 1996; Papert, 1991). While there are many ways to support exploratory learning activities, such as technology-enabled scaffolds (Quintana et al., 2004) or collaborative structures (Puntambekar & Hubscher, 2005), the teacher likely plays a critical role in guiding learners through the wheel-spinning, impasses, and frustration that often accompany these types of minimally structured activities (Kolodner, Gray, & Fasse, 2003). In our view, the efficacy of many exploratory learning activities often boils down to the astuteness of teachers' questions, hints, explanations, and other forms of verbal scaffolding (Furtak, 2006; Lobato, Rhodehamel, & Hohensee, 2012; Roth, 1996). Thus, a key question is which types of teacher talk guidance optimally support students as they engage with exploratory STEM learning activities? In this paper, we investigate teacher talk guidance in an exploratory task called Invention (Schwartz & Bransford, 1998; Schwartz & Martin, 2004), during which students attempt to invent representations to describe a set of data or examples. Research on effective guidance for Invention activities has yielded conflicting results (Chase, Connolly, Lamnina, & Alevan, 2019; Holmes, Day, Park, Bonn, & Roll, 2014; Kapur, 2011; Loibl & Rummel, 2014; Roelle & Berthold, 2016); however, no work has explored teacher talk guidance of Invention.

We adopt the ICAP framework as a theoretical lens through which to explore different forms of teacher talk guidance and resulting student talk, as learners are guided through Invention activities. A core premise of the ICAP framework is that deep learning occurs when learners construct novel inferences. Thus, the ICAP framework seems highly relevant for Invention, given the inherently constructive nature of Invention tasks. According to the ICAP hypothesis (Chi, 2009; Chi & Wylie, 2014), the mode of cognitive engagement learners adopt determines how much they will learn. Interactive modes of cognitive engagement yield greater learning than constructive modes, which are better for learning than Active modes, which in turn yield greater learning than passive modes of engagement ($I > C > A > P$). However, much of the support for ICAP has been culled from studies which classify fairly *traditional* learning activities along ICAP lines. In this paper, we extend the body of ICAP research to the non-traditional context of Invention. Moreover, we look beyond the learning activity to explore how teachers use their *talk* to prompt students for constructive, active, and passive modes of

engagement, how students respond, and how these categories of engagement relate to transfer outcomes.

A key goal of STEM education is to help students transfer their understanding of fundamental principles and structures to novel situations. Unfortunately, transfer across highly variant contexts is notoriously difficult to achieve, in any domain (Anolli, Antonietti, Crisafulli, & Cantoia, 2001; Detterman, 1993; Gick & Holyoak, 1983; Lave, 1988), and there is no shortage of examples of failed transfer in science and math domains (Adey & Shayer, 1993; Clough & Driver, 1986; Georghiadis, 2000; Novick, 1988; Ross, 1989; Vattam & Kolodner, 2008). Thus, it is important to identify forms of teacher-student dialog that can promote the transfer of learning.

This paper works towards two broad goals. One goal is to contribute to research on ICAP, expanding our understanding of how teachers prompt various forms of cognitive engagement and testing the CAP portion of the ICAP hypothesis in the novel context of teacher-student dialog during Invention. A second goal is to understand teacher guidance of Invention activities, with an eye towards identifying effective forms of guidance for Invention and other exploratory STEM learning tasks.

The ICAP framework

The theoretical framework applied in this paper is the ICAP framework of cognitive engagement (Chi, 2009; Chi & Wylie, 2014). The ICAP framework differentiates four modes of cognitive engagement which can be reasonably inferred from learners' overt behaviors. According to Chi and Wylie (2014), *passive* engagement occurs when learners are "oriented toward and receiving information" such as paying attention to a lecture or reading a passage. *Active* engagement occurs when learners "manipulate some part of the learning materials" such as copying solution steps, underlining portions of a text, and repeating or rehearsing information. *Constructive* engagement occurs when learners "generate... externalized outputs" that go beyond the information given, requiring learners to make novel inferences, such as drawing a concept map, generating predictions, or self-explaining. *Interactive* engagement occurs when two learners both construct ideas during joint dialog, such as debating an issue, asking and answering one another's questions, and jointly explaining. Note that the ICAP categories are hierarchical, such that higher modes of engagement subsume lower modes. For example, being interactive also involves being constructive, being constructive also involves active engagement, and so on.

The ICAP hypothesis predicts that learning outcomes (in many domains) depend on the learner's mode of cognitive engagement, such that generally, interactive engagement will yield the greatest amount of learning, followed by constructive, then active, and finally, passive

engagement, which should yield the least learning. A caveat is that the ICAP hypothesis applies to deep learning and other outcomes that require deep, flexible knowledge, such as transfer. For instance, passive engagement may be just as good as constructive engagement for shallow forms of learning like rote memorization of facts.

In an extensive review of the existing literature, Chi (2009) classified the learning activities of many prior studies into ICAP categories and found extensive support for the ICAP hypothesis. There is also more explicit evidence for the ICAP hypothesis which has compared different instructional *activities* explicitly designed by researchers to promote various modes of cognitive engagement (Menekse, Stump, Krause, & Chi, 2013). However, this raises many questions about the *teacher's* role in facilitating these modes of engagement. In the latest work on ICAP from Chi et al. (2018), they explored how teachers translate ICAP into the design of their own lessons. They found that teachers were competent at developing active lessons but struggled to develop constructive and interactive ones. Lessons they designed to be constructive or interactive tended to call for more active engagement, according to an analysis of teachers' worksheet questions. Moreover, while students were slightly more likely to respond with constructive writing in response to constructive or interactive prompts, the most frequent response to any kind of written prompt was active. Thus, teachers had difficulty designing activities to provoke the intended mode of engagement, and even when they were successful, students had difficulty responding with the intended form of engagement. This work underscores an interesting tension between intended and enacted modes of engagement. In our study, we explore to what extent students enact the type of cognitive engagement that teachers prompt, in the context of teacher-student dialog.

The Chi et al. (2018) study explored how effectively teachers translated the ICAP theory into the design of learning activities and worksheets. However, what remains unexplored is how teachers elicit various forms of student engagement with their *talk*, which is arguably the most common form of teacher guidance. While there is some research relating to the ICAP hypothesis in the context of tutor-tutee dialogs (Chi, Roy, & Hausmann, 2008) and other work comparing more and less interactive collaborative talk (Chi & Menekse, 2015), this work did not explicitly code teacher and student talk into constructive, active, or passive categories. Moreover, most work on ICAP has experimentally manipulated the type of lesson, rather than exploring *spontaneously-arising* forms of constructive, active, and passive types of engagement. This ignores the fact that within a single lesson, teachers are likely to provoke, and students are likely to enact, many forms of engagement.

In the current study, we directly examine the teacher's role in facilitating various forms of cognitive engagement in students, in the context of one-on-one teacher-student dialog. Thus, we classify teacher and student talk into constructive, active, and passive (CAP) modes. To explore the alignment between intended and enacted forms of engagement, we ask whether teachers' prompts for various forms of cognitive engagement correspond to the type of cognitive engagement students enact in their responses. In addition, we relate constructive, active, and passive forms of student and teacher talk to students' ability to flexibly transfer their learning to novel contexts, in order to test the CAP portion of the ICAP hypothesis.

We do not examine interactive engagement in this paper because we focus on how teachers *prompt* students for various forms of engagement. To be truly interactive, both partners must (1) share constructive ideas that go beyond the given materials and (2) engage with each other's ideas (Chi et al., 2018). Moreover, interactive dialog is only more beneficial than constructive talk because it allows the learner to infer from new knowledge provided by the partner. Thus, an interactive teacher prompt in the context of teacher-student interaction would provide some knowledge that is new to the student, then prompt the student to make an inference from that knowledge (e.g., by asking the student to elaborate on a given explanation or inviting the learner to reason about some novel information). This occurs in the context of "guided construction" when a more knowledgeable partner (teacher, tutor, expert) provides hints and elaborative feedback that students draw upon, when constructing knowledge (Chi, 2009). In our dataset, there were very few instances of truly interactive prompts, in which the teacher prompted the student to engage with some knowledge that the teacher produced. When teachers provided new information, such as content explanations, they were rarely accompanied by prompts for the student to build on them. In contrast, when teachers prompted learners to construct novel inferences, these prompts were often content-free (e.g., "Explain your answer") or posed a question about information the student had already provided (e.g., "You told me this is more crowded than this. Why?"). Thus, truly interactive prompts so rarely occurred in our data set that we could not effectively explore the interactive category, within the context of teacher prompts. This is perhaps why Chi et al.'s latest definition of interactive engagement confines it to collaborative interactions between peers (2018).

Evidence of the CAP hypothesis in teacher and student talk research

Thus far, we have reviewed work conducted within the ICAP frame, which largely examines learning activities.

In this section, we look beyond ICAP work to review some existing research on teacher and student talk in math and science tasks. The research reviewed in this section does not explicitly examine CAP categories, but we have interpreted the researcher's categories along CAP lines.

Generally, there is a paucity of research exploring the connection between teacher talk and formally-assessed student learning or transfer outcomes (Howe & Abedin, 2013; Kyriacou & Issitt, 2008). Moreover, research on the efficacy of teachers' constructive talk is somewhat mixed. Teachers' constructive prompts in the form of "higher cognitive questions" have shown conflicting effects on student learning across meta-analyses (Redfield & Rousseau, 1981; Samson, Strykowski, Weinstein, & Walberg, 1987; Winne, 1979). More recent qualitative research suggests that constructive forms of questions may be beneficial for learning, particularly when compared to active forms (Hiebert & Wearne, 1993; Wolf, Crosson, & Resnick, 2005). Passive prompts, which largely take the form of teacher explanations, seem to bear little relationship to deep learning. For instance, some work has found that the number of explanations students received was unrelated to their math learning (Webb, 1991). Other work has demonstrated small positive relationships between tutor explanations and learning, but only for shallow learning (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001) or learning of certain science topics (Vanlehn, Siler, Murray, Yamauchi, & Baggett, 2003).

We now discuss evidence relating to the CAP hypothesis in student talk research. Extensive research shows that constructive forms of student talk are beneficial for learning. For instance, students' self-explanations (Bielaczyc, Pirolli, & Brown, 1995; Chi, de Leeuw, Chiu, & LaVancher, 1994; Rittle-Johnson, 2006; Wong, Lawson, & Keeves, 2002), explanations to others (Webb, 1991), and deep questions (Davey & McBride, 1986; Graesser & Person, 1994; King, 1994) have been associated with positive learning and transfer outcomes in several STEM domains. In addition, several studies have shown that students' constructive talk (e.g., interpreting and generating) is associated with greater task success or learning compared to various forms of active or passive student talk (e.g., describing, affirming, etc., Chi et al., 2008; Coleman, Brown, & Rivkin, 1997; Fuchs, Fuchs, Mathes, & Simmons, 1997; Teasley, 1995).

In sum, while there is good evidence suggesting that constructive *student* talk yields learning and transfer, the evidence relating constructive *teacher* prompts to student outcomes is mixed. Studies of constructive teacher prompts may yield inconsistent findings either because teacher talk codes were not designed for CAP categories or because of inconsistent alignment between teacher

and student engagement, such that in some studies, the form of engagement provoked by teacher prompts was not enacted by the students. There is also very little work that distinguishes between active and passive forms of either teacher or student talk. Thus, a study that tests the CAP hypothesis by coding explicitly for CAP categories in both teacher and student talk is needed. Moreover, examining the relationship between aligned teacher-student exchanges and student outcomes may yield stronger evidence of the CAP hypothesis.

Invention activities

Much of the work that supports the ICAP framework involves fairly traditional learning activities, such as reading from a text, note-taking, well-structured problem-solving, and drawing concept maps (Atkinson, Renkl, & Merrill, 2003; Bauer & Koedinger, 2007; Chi et al., 1994; Czerniak & Haney, 1998; Griffin, Wiley, & Thiede, 2008; Trafton & Trickett, 2001; Yin, Vanides, Ruiz-Primo, Ayala, & Shavelson, 2005). There is some work supporting the ICAP hypothesis in non-traditional activities, such as jigsaw collaboration or learning with a pedagogical agent (Doymus, 2008; Moreno, Mayer, Spires, & Lester, 2001). We know of one recent study that coded student talk during collaborative Invention tasks using ICAP categories, but findings relating forms of engagement to outcomes were mixed (Mazziotti, Rummel, Deiglmayr, & Loibl, 2019). Thus, more research is needed to determine whether the ICAP theory extends to exploratory problem-solving activities, such as Invention. In particular, Invention activities may provide a fertile testing ground for the CAP hypothesis. Because the goal of Invention is inherently constructive (i.e., to invent a representation that is novel to the student) we would expect a more balanced distribution of constructive, active, and passive forms of engagement, compared to many traditional learning activities studied in the context of ICAP, which tend to evoke mainly active forms of engagement (Chi et al., 2018).

In the current study, teachers guided students through Invention activities. Invention is a type of exploratory problem-solving that has successfully enhanced learning and transfer in many science and math domains (Roll, Alevin, & Koedinger, 2009; Schwartz & Bransford, 1998; Schwartz, Chase, Oppezzo, & Chin, 2011; Schwartz & Martin, 2004; Shemwell, Chase, & Schwartz, 2015; Taylor, Smith, van Stolk, & Spiegelman, 2010). During Invention, students attempt to construct a representation of the "deep structure" of a domain. A deep structure is a unifying principle or relational structure—a fundamental construct that underlies a content area. Inventions can take many forms, but students often invent an equation, chart, or diagram. Inventors are frequently guided by a set of "contrasting cases" (Fig. 2). These

cases are seeded with pre-designed contrasts that highlight deep features, providing student inventors with clues to the fundamental principles that underlie them (Bransford, Franks, Vye, & Sherwood, 1989; Bransford & Schwartz, 1999).

A key component of the Invention instructional method is that the Invention activity *precedes* direct instruction. While students often produce incorrect inventions, the process of generating and evaluating candidate equations helps them begin to notice deep features of the problem, uncover gaps in their understanding, and develop well-differentiated prior knowledge of the domain (Loibl, Roll, & Rummel, 2017)—all of which prepares them to learn the canonical principles from traditional expositions (Kapur, 2008; Schwartz & Bransford, 1998). The teacher-student dialogs we explore in this study occur only during the Invention phase of this method, not during the lecture phase. This is because we are particularly interested in identifying effective guidance of the exploratory, unstructured portion of this instructional method.

Evidence on the effect of guidance during Invention has been mixed. Loibl and Rummel (2014) found that guidance in the form of contrasting cases that conflict with learners' invented solutions did not impact learning, while Kapur (2011) found that guidance in the form of teachers' on-demand individual help, mini-lectures, and whole-class discussions actually hindered learning, relative to an unguided Invention condition. However, other work suggests that guidance during Invention can be beneficial. One study found that guiding students through a series of subgoals during Invention activities led to greater conceptual learning than a no-guidance condition (Holmes et al., 2014). Another study found that computerized "problematizing" (Reiser, 2004) guidance enhanced transfer over two minimal guidance conditions (Chase et al., 2019), and another study found that comparison activities facilitated learning from Invention activities for some students (Roelle & Berthold, 2016). We suspect that differences across studies may be due to variation in the specific type of guidance provided (ranging from contrasting cases to problematizing activities to teacher lessons). Moreover, in many studies where students successfully learn and transfer from Invention activities, teachers and experimenters provide some verbal guidance during the Invention tasks themselves (Schwartz et al., 2011; Schwartz & Martin, 2004). However, surprisingly little research has examined how specific types of teacher talk affect learning or transfer from Invention tasks. To our knowledge, no studies have measured teachers' talk guidance of Invention tasks, and no studies have explored guidance of Invention using the ICAP framework.

Another question we pose in this work is what forms of guidance teachers will spontaneously provide during

Invention activities. On the one hand, we might expect teachers' guidance to align with the activity, and since Invention is a constructive activity by nature, we might expect teachers to produce mostly constructive prompts. On the other hand, there is work suggesting that even when provided with problems that require learners to generate novel inferences, American math teachers frequently guide students to engage in more active forms of plug-and-chug problem-solving (Stigler & Hiebert, 2004), which would likely be accompanied by mainly active and passive teacher prompts. Moreover, work in the tutoring literature in the context of traditional learning activities suggests that many tutors dominate and control the dialog, largely explaining and providing feedback, rarely providing opportunities for extensive student construction (Chi et al., 2001; Vanlehn, 2018). Thus, we set out to explore the types of cognitive engagement teachers would invite in the context of Invention.

The current study

In the present study, teachers worked one-on-one with students, guiding them through two Invention tasks. We analyzed teacher-student dialog and student performance on assessments of basic learning and transfer of ratio structures in scientific formulas, with an eye towards answering four main questions.

Research question 1: What forms of engagement will teachers prompt students to use, while guiding them through Invention tasks? What forms of engagement will students enact in their talk, while being guided through Invention tasks? Given Chi et al.'s (2018) results, we predicted that the most prevalent type of student talk would be active. However, without prior research on teacher talk in the context of Invention, we did not pose a specific hypothesis about the predominant form of teacher talk.

Research question 2: Will students enact the form of cognitive engagement that teachers prompt? We predicted that teacher prompts and student responses would tend to align, such that constructive prompts would beget constructive responses, active prompts would beget active responses, and passive prompts would beget passive responses. However, given Chi et al.'s (2018) finding, we did not predict perfect alignment.

Research question 3: Which teacher and student talk types are associated with students' ability to transfer? We hypothesized that the proportion of students' talk that is constructive would show the strongest positive relationship with transfer, followed by active, then passive talk. Similarly, for teachers' talk, we predicted that teachers' constructive prompts would show the strongest positive relationship to

transfer, followed by active, then passive prompts. However, this is a weaker prediction, since the ICAP framework ties learning outcomes to the learner’s level of cognitive engagement, and teachers’ prompts may not always elicit the type of student talk they intend (Chi et al., 2018).

Research question 4: Which kinds of aligned teacher-student dialog exchanges are associated with students’ ability to transfer? We speculated that mixed results on the efficacy of teacher prompts might be due to the fact that sometimes these prompts did not provoke the intended form of engagement in students (e.g., Chi et al., 2018). Thus, it is possible that only aligned teacher-student exchanges (e.g., constructive teacher prompts that successfully elicit constructive student responses, etc.) would demonstrate the CAP pattern of results. We predicted that aligned constructive exchanges would show the strongest positive relationship to transfer, followed by aligned active, and then aligned passive exchanges.

Method

Participants

Teachers were three middle school science teachers with more than 5 years of teaching experience (one female, two male). Prior to the study, teachers received a 1-h informational session to familiarize them with study logistics and the Invention tasks. We explained the goals and constraints of each Invention task and told the teachers not to tell students the final answer (otherwise it would not be an Invention task). However, we asked teachers to guide students as they saw fit, and we explicitly refrained from giving any other instruction or guidelines because we wanted to observe untrained teacher guidance of Invention. Students were 18 seventh- and eighth-grade students (50% African American, 28% Latino, 22% multiracial; 14 female, 4 male) recruited from an afterschool program targeted at low-income, urban youth. Students were randomly assigned to teachers, and each teacher guided 5–7 students.

Procedure

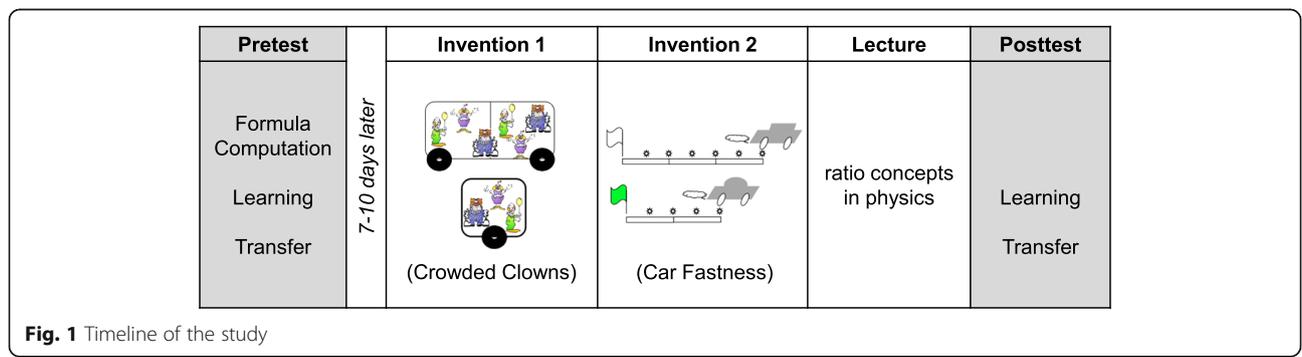
Figure 1 shows the full timeline of the study. Students first took a 15-min pretest in a group setting. Then, about 7–10 days later, depending on scheduling, students completed a one-on-one session with a teacher in a quiet room at the afterschool program. During this session, students were given up to 15 min to complete each Invention task with verbal guidance from the teacher. However, if students generated the fully correct solution to an Invention task before the end of 15 min, they moved on to the next task. After both Invention tasks, teachers gave a brief 6-min lecture on the target learning. Immediately afterward, students completed a 20-min posttest. The posttest was given after the lecture to determine learning and transfer outcomes after the full suite of Invention instruction, which combines both an exploratory Invention experience with follow-up direct instruction.

Instructional materials

Invention tasks

Students completed two Invention activities, guided one-on-one by a teacher. Students spent an average of 25 min completing both tasks (SD = 6.2). The two Invention tasks were a Crowded Clowns task (focusing on density, $D = m/V$) and a Car Fastness task (focusing on speed, $s = d/t$), both adapted from Schwartz et al. (2011) (see Fig. 2). Common to both tasks is the deep structure of ratio, which was the target learning that we hoped students would transfer to other physical science equations (e.g., pressure = force/area). Ratio structures are embedded in many physical science phenomena, yet they are quite challenging for students to grasp (Howe, Nunes, & Bryant, 2010; Piaget & Inhelder, 1975).

Before beginning each task, students watched a 3-min video explaining the task goals and constraints. For instance, in the Crowded Clowns Invention task, students were asked to invent a numerical “index” to describe how crowded the clowns are in each set of busses. By asking students to create an index, we were essentially asking them to generate an equation or procedure for determining clown crowdedness. The video introducing



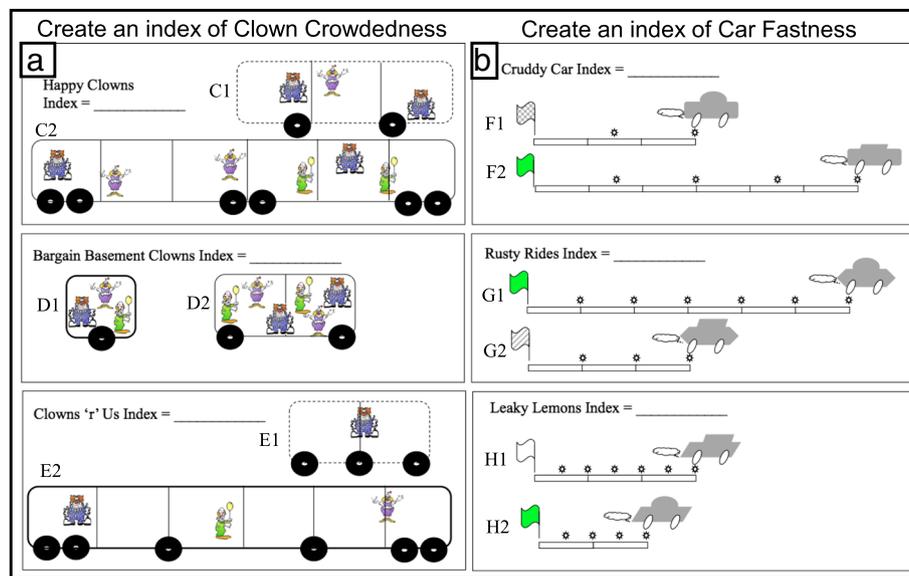


Fig. 2 Invention tasks, adapted from Schwartz et al. (2011). **a** Clown crowdedness. **b** Car fastness

this task gave examples of other indexes (e.g., grades, star ratings). It also gave four simple rules (constraints) that the index must follow:

1. Come up with one number to stand for each company's Crowded Clown Index. The index shows *exactly* how crowded the clowns are.
2. A big index number means the clowns are more crowded. A small index number means the clowns are less crowded.
3. Use the exact same method to find the index for each company.
4. Some companies are more crowded than other companies. But busses from the same company are equally crowded and should have the same index.

Students then began the Invention task, working with a sheet that listed the above rules and a paper version of Fig. 2a. The busses in Fig. 2a are carefully designed contrasting cases that highlight critical features of "crowdedness." For their first attempt at an index, most students often count the clowns in each bus. However, by contrasting cases such as busses C1 and D1, students often begin to realize that the clowns alone cannot explain how crowded a bus is, because bus size matters too. The correct solution is to divide the number of clowns by the number of "boxes" in each bus, relating clowns to bus size in a ratio structure (similar to density). Subsequently, students were asked to invent an index to describe car "fastness" (Fig. 2b), a proxy for speed. In this scenario, cars drip oil as they drive, at the rate of one drop per second. Again, students often begin by counting the number of oil drops, which are highly salient,

but eventually realize the importance of distance traveled (number of segments) as well. The correct solution, dividing the number of segments by the number of oil drops, relates distance to time in a ratio structure. A major goal of this instruction is to get students to begin noticing the ratio structures that exist in many scientific phenomena.

Lecture

After the Invention tasks, teachers gave students a brief lecture using a provided PowerPoint presentation, which explained the correct solutions to the two Invention tasks and showed several worked examples of solving for crowdedness and fastness. The lecture also introduced and defined the concepts of density and speed and connected them to clown crowdedness and car fastness. Then, the lecture emphasized the concept of ratio by comparing the density and speed equations and discussing how both rely on a systematic, mathematical relationship between two quantities. The lecture also described ratios in a qualitative way as a comparison of two variables that have opposite effects on an outcome. Finally, the lecture discussed how ratio structures are common in the world and gave some examples (e.g., shooting percentages in basketball, test scores).

Outcome measures

Invention task performance

To assess learners' success on the two Invention tasks, we counted the number of correct indices on their final Invention worksheets. For each task, there were a total of three possible correct indices, comprising the index numbers they compute for each of three examples

(Fig. 2). Performance across the two Invention tasks was averaged.

Pretest and posttest

The pretest had two formula computation items, two basic learning items and one transfer item. The posttest had isomorphic versions of the basic learning items, the identical transfer problem, and three new transfer problems. Figure 3 shows example test items. *Formula computation* items were designed to test students' initial knowledge of the density and speed equations and their ability to successfully compute them using division. These items were very similar to standard word problems found in physical science textbooks. *Basic learning* items assessed students' qualitative understanding of the ratioed nature of density and speed (two concepts that were explicitly taught during the lecture). This basic learning measure was used mainly as a check to determine whether students had paid attention to the lecture. *Transfer* items assessed transfer across sub-domains within science (Barnett & Ceci, 2002). Transfer items required

students to apply the deep ratio structure to novel physical science concepts, such as pressure and the spring constant, which were not covered in the Invention tasks or lecture. Transfer items were adapted from Schwartz et al. (2011).

Formula computation and basic learning items were coded on a 3-point scale, while transfer items were coded on a 5-point scale, ranging from incorrect to fully correct. Two coders scored 25% of the data and achieved good reliability (intra-class correlations or ICCs for each item are shown in Fig. 3). The remainder of the data were coded by a single master coder. Posttest item subscales demonstrated good internal consistency, Cronbach's $\alpha_{\text{learning}} = 0.60$ and $\alpha_{\text{transfer}} = 0.67$, which is reasonable given the small number of items per subscale. Items for each subscale were scaled from 0 to 1. Average item scores for each subscale were computed for pre- and posttest (see Table 5). We also computed a post-only transfer score, which included only the three new transfer items that were on the posttest because, by definition, transfer items are only those that are novel to participants.

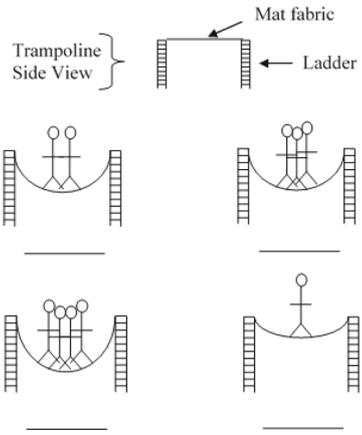
Example Items	Scoring Procedures
<p>Formula computation item: Imagine a huge chocolate heart that is 20 cubic meters big. If it has a mass of 20,000 kg, what is the density of chocolate? Include a formula and show your work.</p> <p style="text-align: right;">ICC's for each item: .89, .92</p>	<p>0: neither correct formula nor correct numerical answer</p> <p>1: correct formula or correct numerical answer</p> <p>2: correct formula and correct numerical answer</p>
<p>Learning Item: Mel and Jenna are trick-or-treating. Throughout the night, they stuff their treat bags with candy. Their treat bags are the same size, but Jenna gets more candy than Mel. Which statement is true?</p> <p>a. Mel's bag is more dense b. Jenna's bag is more dense c. Mel and Jenna's bag are equally dense</p> <p>How do you know?</p> <p style="text-align: right;">ICC's for each item: .79, .83</p>	<p>0: neither correct multiple choice selection nor correct explanation</p> <p>1: correct multiple choice selection or correct explanation</p> <p>2: correct multiple choice selection and correct explanation</p>
<p>Transfer Item: Trampolines are made with mats using different fabrics. Stiffer mats make the trampoline bouncier.</p> <p>Determine the <u>stiffness</u> of the mat fabric for each trampoline.</p> <p style="text-align: right;">ICC's for each item: .98, .97, .95, .98</p> 	<p>0: no correct numerical solutions</p> <p>1: correct numerical answer for one trampoline</p> <p>2: correct numerical answer for two trampolines</p> <p>3: correct numerical answer for three trampolines</p> <p>4: correct numerical answer for four trampolines</p>

Fig. 3 Example test questions and scoring. Transfer items were adapted from Schwartz et al. (2011)

Dialog coding

Sessions were videotaped and dialog was transcribed, including some relevant gestures (e.g., pointing to the cases, writing down a number). All teacher and student utterances were coded at the statement level. A statement corresponds to a single idea. Statement level segments were parsed by a single human transcriber, using indicators such as grammar, inflection, and cadence to demarcate the end of a statement, as recommended in Chi (1997). Afterward, two additional coders read through the statements and marked cases where they disagreed with the segmentation. All segmenting disagreements were then discussed and adjudicated by the two coders. Two coders then classified a random 25% sample of the data into CAP categories. Satisfactory agreement for both teacher and student coding schemes was achieved by the second round of coding ($\kappa_{\text{teacher}} = 0.70$, $\kappa_{\text{student}} = 0.73$). A single master coder coded the remaining teacher and student statements.

The teacher talk coding scheme captures four main kinds of teacher talk: constructive, active, and passive prompts, as well as irrelevant talk. Codes were assigned to statements in a mutually exclusive fashion (Table 1). *Constructive prompts* are statements that prompt students to construct a new idea or explain their reasoning. *Active prompts* ask students to give simple responses such as yes/no or simple numerical calculations which the student has already demonstrated competence in (e.g., counting clowns). *Passive prompts* invite students to pay attention, and these largely occur when the teacher is providing explanations about the task or related concepts. Statements that do not apply directly to

the content or task process are coded as irrelevant with respect to our hypotheses. *Irrelevant statements* include motivational remarks, monitoring of the teacher's own process, and off-topic talk. Effects of these irrelevant statements were not directly analyzed because the ICAP model only applies to engagement with content or task-relevant activities (Chi & Wylie, 2014).

The student talk coding scheme captured four mutually exclusive categories: constructive, active, passive, and irrelevant talk (see Table 2). *Constructive statements* indicate that the student is generating a novel idea, problem solution, substantive question, explanation, or novel mathematical reasoning. *Active statements* demonstrate that the learner is actively manipulating information but not generating a novel inference, such as repeating a phrase the teacher said, doing simple math for which the student has already generated competence (e.g. counting), or making simple yes/no judgments. *Passive statements* imply that the learner is paying attention but not taking any action, such as simple continuers and agreements. Student statements that are not relevant to the task or content are coded as *irrelevant*. Note that while students may have been engaging in various forms of mental activity without verbalizing it, overt student talk was our best measure of students' cognitive engagement.

When assigning codes, coders considered dialog that came before (but not after) each statement, which is necessary for coders to determine whether a prompt is constructive or active with respect to each student. This helped to determine whether the teacher was asking the learner to construct an idea that was new to the student

Table 1 Teacher talk codes

Code	Description	Examples
Constructive prompt	A question or prompt for constructive engagement. Asking the learner to explain or generate new ideas, reasoning, or mathematical thinking that has not been mentioned in previous talk.	<ul style="list-style-type: none"> - I'll give you a minute to look at the information and see if you can come up with a car fastness index. - How do you know that this one [points to F2] is going faster than this one [points to H2]? - Why did you use division to make your index?
Active prompt	A question or prompt for active engagement. Asking the learner for information with a finite, known set of answers or responses, such as yes/no questions, or asking the student to do simple math that the student has demonstrated competence in through previous talk.	<ul style="list-style-type: none"> - From here [points to F1 flag] to here [points to first oil drop in F1], how many drops did you see? - Six what? - Why don't you write that down here [points to right of C1]?
Passive prompt	A statement that requires no more than passive engagement, requiring the learner only to attend but not inviting the learner to be active/constructive, such as when the teacher provides an explanation, elaboration, or revoicing of a concept, idea, problem features, or task constraint.	<ul style="list-style-type: none"> - Now what we want to do is make an index, and remember that the lowest number means the least crowded, and the highest number means the most crowded. - So now you know how much time each of these is because you know how many oil drops [points across F2] in that period of time, right. - So you said this bus has three clowns.
Irrelevant talk	Statements that do not apply to content or process. These include motivational remarks, monitoring of the teacher's process, and off-topic talk.	<ul style="list-style-type: none"> - I can see the gears turning in your brain! - Let me just check something on the worksheet.

Table 2 Student talk codes. Student statements are italicized, while teacher statements are not

Code	Description	Example
Constructive talk	Talk indicating that the student is constructing novel outputs and inferences. These include explanations about the topic or process, meaningful questions, generating or applying a novel invention, complex mathematical reasoning, and substantial observations about the task or cases, which were not mentioned in previous talk.	- <i>I know that this is the least crowded [points to E] because even though there's like people there [points to clowns in C2], it depends on how much space it is.</i> - <i>I think this one is one third [points to D] because in each space [points to second room in D2] there's three clowns [points to D2].</i> - <i>These boxes below [points to segments in F2] show the distance, right?</i>
Active talk	Talk indicating the student's active engagement with the task but not involving novel inferences. These included simple math that the student has already generated competence in (e.g., counting), simple decisions (e.g., yes/no), or repeating back what the teacher said.	So there's how many people for this space [points to C1]? - <i>One.</i> You want to give this a lower number? - <i>Yeah.</i> OK, so maybe we can just call that something like a dividing line? - <i>Yeah, dividing line</i>
Passive talk	Talk indicating that the learner is paying attention but not engaging actively. These include simple continuers and agreements.	- <i>M-hmm.</i> - <i>OK.</i> Why? - <i>Because there's... umm...</i>
Irrelevant talk	Off-topic talk that does not relate to the content or task. These include motivational remarks, monitoring statements that do not imply constructive/active/passive engagement, and other irrelevant talk.	- <i>I'm pretty good at math.</i> - <i>This is hard.</i> - <i>How long will this take?</i>

or whether they were simply asking the student to rehearse what the student already knew. Relevant gestures were also used to interpret the talk, such as discerning referents in the dialog (e.g., “this one” [points to specific case]).

To investigate the alignment of teacher-student talk, we also coded dialog at the turn level. A single turn was comprised of all the talk made by a single speaker before the next speaker talked. Thus, three teacher statements followed by two student statements were collapsed into one teacher turn and one student turn. Turns were coded using the same codes as statement-level codes (See Tables 1 and 2). A simple algorithm was used to convert statement-level codes to turn-codes. Single-statement turns received the same codes as the statement itself. Turns composed of multiple statements received the code that occurred most frequently. If all codes appeared equally, the code of the last statement in the turn was used, since this typically conveyed the main point of each turn. If irrelevant talk was the most frequent or last statement in the turn, the next best option was used. To check whether this algorithm accurately identified the intent of the turn, a single coder scored a random 25% selection of the data using the overall gist of each turn and achieved strong reliability with the algorithm on the first round ($\kappa = 0.91$).

Pairs of sequential teacher-student turns were concatenated to produce teacher-student dialog exchanges. This yielded nine types of teacher-student exchanges of interest: constructive-constructive, constructive-active, constructive-passive, active-constructive, active-active,

active-passive, passive-constructive, passive-active, and passive-passive. We chose to look at teacher-student turn pairs, rather than student-teacher turn pairs, because teacher-student dialogs were largely initiated and driven by the teachers. Analysis of teacher-student exchanges enabled us to explore the degree of alignment in teacher-student dialog and the relationship between aligned teacher-student exchanges and student transfer outcomes.

Dialog measures

Proportions of various types of statements or teacher-student exchanges (out of all statements/exchanges) were computed. This provided a way to standardize the relative contributions of each type of talk across students, since the total amount of talk varied drastically by student (ranging from 81 to 383 statements of teacher talk and 54 to 310 statements of student talk). This is because teachers were instructed to go on to the next problem or phase of the study as soon as students produced a correct invention. Thus, students who solved the problem quickly had far less teacher-student dialog. Given this, proportions of talk are a more appropriate measure than talk frequencies, since they give us a feel for the relative amounts of each type of talk students and teachers engaged in during an Invention session, regardless of the total amount of dialog.

Teacher and student talk statements

To describe the relative amounts of talk statements of various types, we computed proportions of constructive,

active, passive, and irrelevant statements out of total statements. These proportions were computed separately for teacher talk and for student talk (see Tables 3 and 4).

Teacher-student turn exchanges

To explore alignment between teacher prompts and student responses, we computed (1) for each type of teacher turn, the proportions of each type of student turn that followed (e.g., of all constructive prompts, the proportion that were followed by constructive, active, and passive student responses), and (2) for each type of student turn, proportions of each teacher turn that preceded it (e.g. of all constructive student responses, the proportion that was preceded by constructive, active, and passive student responses, see Fig. 4).

Aligned teacher-student exchange predictors

To measure aligned teacher-student turns, we computed proportions of constructive-constructive, active-active, and passive-passive teacher-student exchanges, out of all teacher-student exchanges.

Results

Because all talk variables in our data set are proportions (not frequencies), issues of absolute multicollinearity arise. To avoid this problem, we did not conduct statistical analyses on the irrelevant talk category, for which we had no hypotheses. Thus, statistical analyses were only conducted on proportions of constructive, active, and passive forms of talk. Despite this, we did run into one issue with collinearity, in which proportions of constructive and passive teacher prompts were highly correlated. We address this further in our interpretation of the teacher talk regression results.

In all relevant analyses, pretest performance on formula computation items was used as a covariate, to control for prior knowledge. We used pretest formula computation as our measure of prior knowledge because it was positively related to transfer outcomes ($r = .50$, $p = .03$), while neither the basic learning pretest nor the pretest transfer item were significantly correlated with posttest transfer scores, p 's $> .48$. Pretest formula computation items assessed students' initial knowledge of the relevant density and speed equations and their ability to apply them accurately in simple word problems.

Table 3 Mean frequencies and proportions of *teacher* statements (with SD)

Teacher talk code	Frequency	Proportion (out of total teacher talk)
Constructive prompt	67.1 (41.0)	0.28 (0.11)
Active prompt	66.3 (26.7)	0.29 (0.06)
Passive prompt	57.9 (28.0)	0.28 (0.13)
Irrelevant talk	35.8 (19.2)	0.15 (0.04)

Table 4 Mean frequencies and proportions of *student* statements (with SD)

Structure code	Frequency	Proportion (out of total student talk)
Constructive talk	57.7 (20.6)	0.33 (0.05)
Active talk	83.7 (35.4)	0.48 (0.09)
Passive talk	14.1 (7.0)	0.08 (0.04)
Irrelevant talk	17.8 (13.6)	0.10 (0.06)

What mode of engagement was invited by teacher talk?

Teachers uttered an average of 227 statements (SD = 81.8), summed across both Invention tasks. Table 3 shows both raw frequencies and proportions for each type of teacher talk (out of total teacher statements). Teachers apportioned their talk equally across each type of talk, posing similar proportions of constructive (28%), active (29%), and passive prompts (28%). The remaining 15% of talk was irrelevant.

A mixed ANCOVA on proportions of teacher talk statements, with the type of talk as a within-subjects factor, teacher as a between-subjects factor, and pretest as covariate, found no significant teacher effect, $F(1, 14) = 1.75$, $p = .21$, nor teacher x talk type interaction, $F(4, 28) = 1.87$, $p = .19$, indicating that talk type proportions did not differ across individual teachers. There were no significant effects of either pretest, $F(1, 14) = 0.27$, $p = .62$, or pretest x talk type interaction, $F(2, 28) = 1.28$, $p = .29$, suggesting that students' prior knowledge did not influence proportions of various types of teacher talk. Most importantly, there was no significant effect of talk type, $F(2, 28) = 1.66$, $p = .21$, confirming that the proportions of various teacher talk types were similar.

What mode of engagement was demonstrated by student talk?

Students uttered an average of 173 statements (SD = 59.9), summed across both Invention tasks. The largest proportion of student statements were active (48%), followed by constructive (33%), then passive (8%), which occurred relatively infrequently. An additional 10% of talk was coded as irrelevant. Table 4 shows the mean frequencies and proportions (out of total student talk) of each type of student statement.

A mixed ANCOVA was run on student talk proportions, with talk type as the within-subjects factor, teacher as the between-subjects factor, and pretest as a covariate. There were no significant effects of either teacher, $F(2, 14) = 1.24$, $p = .32$, or teacher x talk type interaction, $F(4, 28) = 0.85$, $p = .51$, demonstrating that specific teachers did not yield different proportions of student talk types. There were also no effects of student pretest, $F(1, 14) = 0.03$, $p = .86$, or pretest x student talk type interaction, $F(2, 28) = 0.0002$, $p = 1.0$, indicating that students' prior knowledge did not affect the proportions of different talk

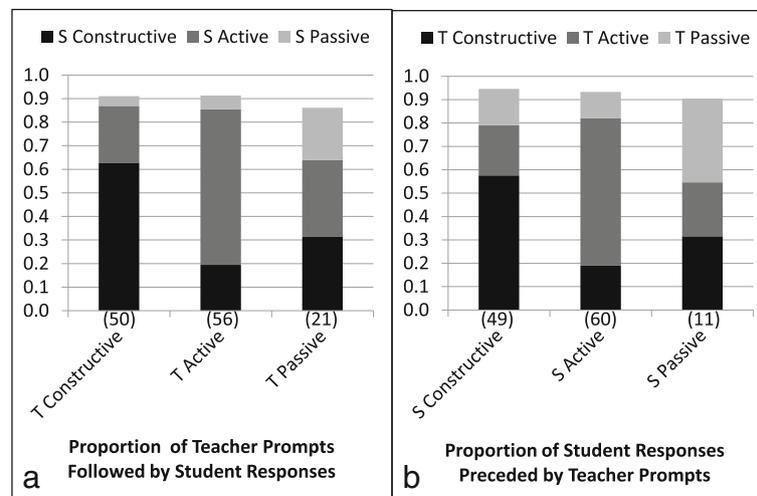


Fig. 4 Proportion of teacher-student exchanges from **a** teacher perspective and **b** student perspective (with mean total turns)

types they produced. However, there was a significant main effect of talk type, $F(2, 28) = 48.16, p < .001, \eta_p^2 = .78$, indicating that the proportion of statements differed by talk type. Post hoc tests with Bonferroni correction revealed that all three proportions of student talk types differed significantly from one another, p 's $< .001$. Thus, the highest proportion of student talk was active, followed by constructive, then passive.

Do students enact the form of engagement that teachers prompt?

In general, we found fairly good alignment between teacher prompts and student responses, except in the case of passive prompts, which yielded a variety of kinds of student responses. Figure 4a shows the proportion of each type of teacher prompt that resulted in each type of student response. For instance, in the left-most column of the graph, we see that on average, 62% of a teacher's constructive prompts were followed by students' constructive responses, 24% of constructive prompts were followed by active responses, and only 4% were followed by passive responses. Figure 4b looks at the same data from the student perspective. It shows the proportion of each type of student talk that was preceded by each type of teacher talk. The left-most column of Fig. 4b shows that on average, 58% of a student's constructive responses were preceded by constructive prompts, 22% of constructive responses were preceded by active prompts, and 16% were preceded by passive prompts.

To statistically test for differences in teacher-student turn transitions, a mixed ANCOVA was run on the proportion of each type of teacher turn that was followed by each student turn (i.e., the numbers shown in Fig. 4a). Type of teacher prompt and type of student response were both treated as within-subjects factors, individual

teacher was a between-subjects factor, and pretest was a covariate. There were no significant teacher or teacher \times talk type effects, nor any pretest or pretest \times talk type effects, p 's $> .21$. Most importantly, there was a significant teacher prompt \times student response interaction, $F(4, 56) = 15.02, p < .001, \eta_p^2 = 0.52$. Post hoc tests with Bonferroni correction demonstrated that constructive teacher prompts were most likely followed by constructive student responses, then by active responses, and least likely followed by passive responses, p 's $< .001$. Active teacher prompts were most likely followed by active student responses, then by constructive, and least likely by passive responses, p 's $< .001$. However, the proportion of responses to passive teacher prompts did not differ from one another, such that students were equally likely to respond to a passive teacher prompt using constructive, active, or passive language, p 's $> .87$.

A similar repeated measures ANCOVA was conducted on the proportion of each student response that is preceded by each type of teacher prompt (i.e., the numbers of Fig. 4b), with the same factors. Neither teacher nor teacher \times talk type effects were significant, nor were any pretest or pretest \times talk type effects, p 's $> .24$. Most importantly, there was a significant student talk type \times teacher talk type interaction, $F(4, 56) = 16.36, p < .001, \eta_p^2 = .54$. Post hoc tests with Bonferroni correction demonstrate that constructive student responses were most frequently preceded by constructive prompts, p 's $< .001$. Constructive responses were less likely to occur after active and passive prompts; however, these proportions did not differ from one another, $p = .08$. Active responses were most likely to occur after active prompts and least likely to occur after passive prompts, p 's $< .003$. Passive responses were equally likely to occur in response to any kind of prompt, p 's $> .58$.

To summarize, constructive prompts were most likely to be followed by constructive responses, and active prompts were most likely to be followed by active responses. However, there was no single most common response to passive prompts. Looking at it from the perspective of student responses demonstrated the same pattern. It is interesting to note that while teachers and students tended to be in sync on constructive and active talk, the alignment is far from perfect (percentages never reach higher than 66%). Thus, the type of cognitive engagement that teachers solicited was not always what students chose to enact. Moreover, passive prompts and responses did not align, such that passive prompts were equally likely to yield any form of response.

Invention task performance

Students were fairly successful at the Invention tasks, with an average of 2.2 out of 3 correct indices ($SD = 1.0$) across the two Invention tasks they completed. Only 2 of the 18 participants had fewer than 1.5 correct indices. Individual teachers did not differ in their ability to guide students towards correct inventions, $F(2, 15) = 0.56$, $p = .59$. Bivariate correlations showed that Invention task performance was not significantly related to either basic learning or transfer performance on the posttest ($r_{\text{learning}} = 0.23$, $p = .36$; $r_{\text{transfer}} = 0.01$, $p = .96$). Thus, we did not attempt to relate teacher and student talk to Invention task performance.

Basic learning and transfer

Students improved their scores from pre- to posttest (see Table 5). On basic learning items, students improved by 38%, while on the one transfer item that was common to pre- and posttest, students more than doubled their scores. A repeated measures ANOVA with time, item type, and teacher as factors confirmed that pre- and posttest scores differed significantly, $F(1, 15) = 14.96$, $p = .002$, $\eta_p^2 = .50$, but there was no interaction of item by time, $F(1, 15) = 1.04$, $p = .33$. There was also no effect of teacher \times time, $F(2, 15) = 1.37$, $p = .28$, nor any interaction of teacher \times time \times item, $F(2, 15) = 1.86$, $p = .19$. Post hoc tests with Bonferroni correction confirmed that students made significant gains on both

Table 5 Mean pre- and posttest learning and transfer scores on 0–1 scale (with SD)

Item type	Pretest	Posttest
Formula computation	0.36 (0.31)	–
Basic learning*	0.50 (0.33)	0.69 (0.27)
Transfer**	0.24 (0.25)	0.56 (0.32)
Post-only transfer items	–	0.49 (0.25)

Significant gain at * $p < .05$ and ** $p < .01$

basic learning ($p = .04$, $d = 0.56$) and transfer items ($p = .004$, $d = 0.73$).

Relating teacher and student talk to transfer outcomes

We used regression analyses to test the relationship between talk variables and transfer outcomes, while controlling for prior knowledge. Three separate sets of regression analyses were run, predicting transfer: one with teacher talk variables, another with student talk variables, and one with teacher-student exchange variables.¹ Proportions of various types of statements (out of all statements) were used as predictors. The post-only transfer score was used as the outcome variable of all regression analysis presented in the next section. We did not explore the relationship between talk variables and basic learning, since the ICAP hypothesis makes predictions about deep learning and transfer, but not shallow learning (Chi & Wylie, 2014).

In all regression analyses, pretest performance was used as a covariate, to control for prior knowledge. We did not control for measures of task performance or the number of student errors, since neither was significantly correlated with posttest transfer scores (p 's $> .15$). We also did not control for individual teachers in this regression because neither the proportions of teacher statements nor transfer scores differed by individual teacher, and when dummy variables for teacher were added to the regression, they were not significant.

We used hierarchical regression models to test the relationship between constructive, active, and passive forms of talk and transfer outcomes. For each analysis, our control variable was entered first (pretest), then each talk variable was added in order from the strongest to weakest predicted effects. We predicted that constructive talk would have the strongest positive relationship with transfer and passive talk would have the weakest effect on transfer. Thus, we first added constructive, followed by active, and then passive talk.

Teacher talk predictors of transfer

Table 6 shows the results of our teacher talk regression analysis. The main predictors of interest are the proportion of each type of teacher statement (constructive, active, and passive) out of all teacher statements (mean values shown in Table 3).

In Model 1, pretest formula computation scores were entered alone and explained 28% of the variance in posttest scores. Because the proportions of teachers' constructive and passive prompts were highly collinear ($r = -.91$, $p < .001$), neither predictor was significant when simultaneously entered into a regression equation. Thus, we ran two sets of models (Model 2a/3a and Model 2b/3b) to test the effect of

Table 6 Regression models predicting transfer scores from proportions of *teacher* talk types

	<i>B</i>	SE <i>B</i>	β	<i>F</i> (df)	ΔF (df)	<i>R</i> ²	ΔR^2
Model 1				6.16 (1, 16)*	–	.28	–
Pretest*	0.42	0.17	0.53				
Model 2a				11.67 (2,15)***	12.68 (1, 15)**	.61	.33
Pretest***	0.55	0.13	0.69				
T Constructive**	1.33	0.38	0.60				
Model 3a				7.41 (3, 14)**	0.18 (1, 14)	.61	.01
Pretest***	0.56	0.14	0.71				
T Constructive**	1.37	0.39	0.61				
T active	– 0.31	0.75	–0.07				
Model 2b				2.92 (2, 15)	0.04 (1, 15)	.28	.002
Pretest*	0.41	0.18	0.52				
T active	0.20	0.97	0.05				
Model 3b				7.19 (3, 14)**	11.61 (1, 14)**	.61	.33
Pretest**	0.55	0.14	0.69				
T active	– 1.08	0.83	– 0.25				
T passive**	– 1.22	0.36	– 0.65				

Significant at * $p < .05$, ** $p < .01$, *** $p < .001$

constructive and passive prompts, independent of one another.

In model 2a, the proportion of teachers' constructive prompts positively predicted transfer, accounting for an additional 33% of variance in transfer scores beyond what pretest scores predicted alone. In model 3a, active prompts were added to the model, but they did not explain additional variance.

In model 2b, we removed constructive prompts such that only pretest and active prompts were in the model. Model 2b was not significant and active prompts did not significantly predict transfer scores. In model 3b, passive prompts were added, which negatively predicted transfer scores and accounted for an additional 33% of the variance.

Because the proportion of teachers' constructive and passive prompts were collinear, it is impossible to tell if teachers' passive talk was hurting transfer or if teacher's constructive talk was helping transfer. However, it makes sense that when teachers prompt students to generate and construct on their own that teachers themselves do less of the generating (passive prompts largely consist of teachers explaining to students). Thus, our interpretation of these data is that controlling for prior knowledge and active teacher prompts, students who received both a higher proportion of constructive prompts and a lower proportion of passive prompts were more successful at transferring their knowledge to novel situations. An alternative interpretation is that lower-ability students received a greater proportion of explanations (which are coded as passive prompts) from teachers and

also transferred less. However, since this analysis controls for pretest scores (which were not significantly correlated with the prevalence of various teacher talk types), this alternative explanation seems unlikely.

Student talk predictors of transfer

We next conducted a similar hierarchical regression analysis relating the proportions of student talk types (means shown in Table 4) to transfer outcomes, following our planned order: pretest, constructive, active, and then passive prompts. Table 7 shows the regression results.

In model 1, pretest formula computation scores were entered alone. Pretest scores positively predicted transfer, explaining 28% of the variance in transfer scores. Additional models did not explain significantly more variance than model 1.

While none of the predictors of interest were significant in the final model, the beta coefficients are in the predicted direction of the general CAP hypothesis ($C > A > P$), demonstrating a positive relationship between constructive talk and transfer ($\beta = 0.22$), a weaker positive relationship between active talk and transfer ($\beta = 0.18$), and a negative relationship between passive talk and transfer ($\beta = -0.23$).

Aligned teacher-student exchange predictors of transfer

Finally, we explored the relationship between aligned teacher-student exchanges (turn pairs) and students' transfer scores (Table 8). While there are nine types of teacher-student exchanges, we only had hypotheses

Table 7 Regression models predicting transfer scores from proportions of student talk types

	<i>B</i>	<i>SE B</i>	β	<i>F</i> (df)	ΔF (df)	<i>R</i> ²	ΔR^2
Model 1				6.16 (1, 16)*	–	.28	–
Pretest*	0.42	0.17	0.53				
Model 2				3.72 (2, 15)*	1.20 (1, 15)	.33	.05
Pretest*	0.42	0.17	0.53				
S constructive	1.14	1.04	0.23				
Model 3				4.00 (3, 14)*	3.40 (1, 14)	.46	.13
Pretest*	0.42	0.16	0.53				
S constructive	1.41	0.98	0.29				
S active	1.00	0.54	0.37				
Model 4				2.97 (4, 13)	0.40 (1, 13)	.48	.02
Pretest*	0.42	0.16	0.53				
S constructive	1.06	1.15	0.22				
S active	0.50	0.98	0.18				
S passive	– 1.20	1.91	– 0.23				

Significant at **p* < .05

about constructive-constructive, active-active, and passive-passive exchanges. Thus, we maintained the same basic hierarchical model by first adding pretest as our control variable, and then testing our variables of interest by next adding constructive-constructive, active-active, and then passive-passive exchanges.

Adding constructive-constructive teacher-student exchanges in model 2, explained an additional 28% of the variance in transfer, over a model with pretest scores only. Constructive teacher-student exchanges positively predicted transfer. Adding active-active exchanges and passive-passive exchanges in models 3 and 4 did not add any predictive power. However, constructive-constructive

exchanges lost their significance once active exchanges were added to the model, likely because they explained a large proportion of overlapping variance² (constructive and active proportions were correlated but not collinear, *r* = .64, *p* = .004). One interpretation of this result is that when teacher-student pairs engaged in a higher proportion of constructive teacher-student exchanges, students were better able to transfer their knowledge to novel contexts, regardless of their prior abilities, while adding active and passive prompts did not explain significant additional variance in transfer outcomes. Another interpretation is that constructive exchanges did not explain unique variance above that of active or passive teacher-student

Table 8 Regression models predicting transfer scores from proportions of teacher-student exchanges

	<i>B</i>	<i>SE B</i>	β	<i>F</i> (df)	ΔF (df)	<i>R</i> ²	ΔR^2
Model 1				6.16 (1, 16)*	–	.28	–
Pretest*	0.42	0.17	0.53				
Model 2				9.35 (2, 15)**	9.44 (1, 15)**	.56	.28
Pretest**	0.46	0.14	0.58				
T-S constructive**	1.57	0.51	0.53				
Model 3				6.65 (3, 14)**	1.11 (1, 14)	.59	.03
Pretest**	0.43	0.14	0.55				
T-S constructive	1.10	0.67	0.37				
T-S active	0.65	0.62	0.24				
Model 4				5.03 (4, 13)*	0.66 (1, 13)	.61	.02
Pretest**	0.45	0.14	0.57				
T-S constructive	1.03	0.69	0.35				
T-S active	0.35	0.73	0.13				
T-S passive	– 1.29	1.58	– 0.19				

Significant at **p* < .05, ***p* < .01

exchanges. Either way, it seems that the relationship between teacher talk and transfer is more robust than that of teacher-student exchanges and transfer.

It is interesting to note that even though the teacher-student predictors were not significant in the final model, the pattern of results is the same as those found in the separate regressions of student and teacher talk. Constructive-constructive exchanges were positively related to transfer ($\beta = 0.35$), active-active exchanges were positively associated with transfer to a lesser degree ($\beta = 0.13$), and passive-passive exchanges were negatively related to transfer ($\beta = -0.19$).

Discussion

In this study, we asked experienced teachers to provide one-on-one verbal guidance to students as they worked on Invention tasks, which are a form of exploratory activity for science and math learning. Students were fairly successful by measures of task performance, basic learning gains, and transfer gains. Almost all students invented the canonical solutions for at least some of the cases, and students made sizeable learning gains after only 30 min of instruction ($d = 0.56$). More importantly, students made large transfer gains ($d = .73$), demonstrating their ability to recognize and implement ratio structures in novel science domains. This is impressive, given that transfer is hard to achieve, and there are many empirical demonstrations of transfer failure (Detterman, 1993; Lave, 1988). Moreover, the length of instruction was much shorter than in previous Invention studies, in which students completed Invention tasks over several class periods (Schwartz et al., 2011; Schwartz & Martin, 2004). Thus, students learned and, in particular, transferred successfully from these guided Invention activities. We suspect that the one-on-one teacher guidance led to these impressive learning and transfer gains (though we cannot know for sure without testing an unguided condition). This is supported by research showing that human tutoring is an effective form of instruction (Cohen, Kulik, & Kulik, 1982; Lepper, Drake, & O'Donnell-Johnson, 1997).

What forms of engagement did teachers prompt and what forms of engagement did students enact in the context of a guided Invention task? We created and applied explicit codes for teacher prompts and student talk related to constructive, active, and passive forms of cognitive engagement. Teachers used their talk to guide students through Invention tasks by prompting students to construct their own ideas and explanations, to actively rehearse simple math or repeat information, and to passively attend to the teacher, in roughly equal proportions. In contrast to experimental studies of the ICAP framework (Chi & Menekse, 2015; Menekse et al., 2013; Menekse & Chi, 2018), which manipulate cognitive

engagement across activities (e.g., comparing a constructive activity to an active activity), in our study, teacher talk guidance prompted an even mix of all forms of engagement, within a single activity. Moreover, it is interesting that teachers did not produce mainly active or passive prompts, which might be the case if teachers had treated the Invention task as a more traditional, tell-and-practice type of activity. It is also interesting to note that despite the constructive nature of the task, teachers did not provide a majority of constructive prompts. It is likely that extended teacher-student dialog would be challenging to carry on without some active and passive prompts. For instance, passive prompts can be useful, such as when the teacher wants to encourage the student to pay attention or to direct her attention to a particular case. Likewise, some active prompts may be necessary to give a student a sense of mastery or to scaffold student progress on the task, such as asking a student to apply a procedure she has mastered for one case to another case or encouraging her to write down her answer so she will not forget it.

In contrast to teachers, students generated mostly active talk, followed by constructive talk, while passive talk occurred relatively infrequently. Thus, while students were doing more than merely attending most of the time, they naturally engaged in a high proportion of rehearsing and repeating and a slightly lower share of generative activity. This is somewhat consonant with Chi et al.'s (2018) findings, in which she found that the majority of student writing, in response to any kind of prompt, indicated active engagement. However, a far higher proportion of student constructive talk was found in our study, perhaps due to the inherently constructive nature of the Invention task and the greater proportion of teacher constructive prompts.

How did teacher and student talk categories align? An analysis of teacher-student turn exchanges demonstrated that teachers' prompts and students' responses tended to align along constructive and active modes but not passive. Constructive prompts were most often followed by constructive responses, and active prompts tended to yield active responses. However, passive prompts were equally likely to yield constructive, active, and passive responses. This partially aligns with Chi et al.'s (2018) study, in which posing a constructive worksheet question led to a higher proportion of constructive student responses compared to when teachers posed an active question. However, they found that students were highly biased to engage in active modes such that the majority of responses to constructive questions were still active. It is possible that our study produced better alignment between teachers and students because our study examined teacher-student *dialog* in a *one-on-one* setting, while their study examined teacher-student *writing* in a

whole-class setting, where there is less pressure for students to respond according to the prompt. However, as in Chi et al. (2018), teacher-student talk alignment in our study was not perfect (across CAP categories, the alignment was never greater than 66%). This suggests the need to investigate other contextual factors that influence students' cognitive engagement.

Why did passive prompts show poor alignment, resulting in a mix of response types? Most of what we coded as "passive prompts" were not explicit prompts for attention like "look here" or "listen up" (though there were some of these). Instead, passive prompts largely took the form of teacher explanations. Explanations are implicit passive prompts because the implication is that students should listen when the teacher is explaining, but they do not explicitly direct students to pay attention. Constructive and active prompts, on the other hand, were more explicit in asking students to generate a specific type of output, such as asking them to explain their reasoning or state a fact (e.g., "How did you come up with that number?" or "How many clowns are in the picture?"). Without this explicit direction, passive prompts give students the freedom to respond in many ways. In addition, passive student talk can occur in a range of situations. For instance, when a student is stumped by a constructive or active question, it is common for her to respond with only "Ummm..." which is coded as passive talk. In these situations, it is possible that our code for passive engagement in student talk was not very accurate, since students may be thinking many constructive or active thoughts, but when they do not have a specific or confident answer to a teacher's prompt, they only voice passive responses like "um." Finally, it is possible that by coding largely for verbal responses (though we did take gestures into account), we may have underestimated the occurrence of passive student responses. While students often produced passive utterances to signify that they were paying attention, others may have paid attention silently, which is a type of passive response that was not captured in our data. It is possible that the alignment between passive teacher prompts and passive student responses would have improved had we measured non-verbal passive engagement. Nonetheless, we view the poor alignment between passive teacher prompts and passive student responses as an interesting and novel finding (note that Chi et al.'s (2018) study did not code for passive worksheet prompts or responses).

Which forms of talk were associated with students' ability to transfer from Invention tasks? First, we found that teachers' individual talk showed the most robust relationship to transfer, while aligned teacher-student dialog demonstrated some relationships but less robust effects. Moreover, student talk did not significantly predict transfer outcomes. This work underscores the

importance of the teacher's role in supporting students with verbal prompts for the appropriate form of cognitive engagement, when student transfer is the end goal. While previous ICAP research has largely focused on how specific kinds of learning *activities* lead students to adopt different forms of cognitive engagement, which in turn influences learning, our work shows that the *verbal guidance* teachers use to provoke cognitive engagement can shape what learners transfer from an activity. Even within a highly constructive activity like Invention, where the goal is for learners to generate their own equations for physical science phenomena, how teachers use their talk to evoke different forms of student cognitive engagement matters.

Which kinds of teacher prompts predicted transfer? When controlling for pretest scores and active prompts, the proportion of teachers' constructive prompts was positively related to transfer scores, while the proportion of teacher passive prompts was negatively related to transfer scores. Given that teachers' constructive and passive prompts are collinear, it is difficult to tell which one is contributing to transfer outcomes. That is, when teachers predominantly ask students to construct explanations and solutions, they do relatively less passive prompting, during which teachers largely explain to students. Our interpretation of these data is that when teacher guidance contained relatively more constructive prompts and fewer passive prompts within an Invention task, students demonstrated greater transfer. It is interesting that the proportion of teachers' passive prompts may have actually hindered students' ability to transfer their knowledge to novel contexts, while previous research suggests that explanations have little to no effect on learning in more traditional learning activities (which often yields transfer, Chi et al., 2001; Vanlehn et al., 2003; Webb, 1991). It may be that teacher explanations are counterproductive in exploratory learning activities like Invention, where the goal is to engage students in productive exploration. When teachers prompt students in mostly passive ways, they do not invite them to continue exploring the problem and domain space. Moreover, when teachers do the explaining, they may eliminate students' need to explore or explain for themselves (Schwartz et al., 2011; Schworm & Renkl, 2002).

What kinds of student talk predicted transfer? We were surprised to find that forms of student talk were not significantly associated with transfer. This is puzzling, since the ICAP theory argues that the student's mode of engagement (and attendant cognitive processes) is most important for influencing learning and transfer, and there is extensive research linking forms of student constructive talk to learning and transfer outcomes (Chi et al., 1994; King, 1994; Webb, 1991). We offer several potential explanations for the lack of significant student

talk-transfer relations. One possibility is that the CAP coding scheme was more accurately applied to teacher talk than student talk, particularly when it comes to distinguishing between constructive and active talk categories. Note that in our coding scheme, a statement is coded as constructive when the student is demonstrating some novel inferences or novel output, while active statements are those in which the student is largely rehearsing known facts or already-mastered math. Thus, the coder would have to determine from the problem context and the talk that came before, whether the student was demonstrating some new substantive thinking or just rehearsing a known math fact, and some statements were bound to be misclassified. For this reason, Chi and Wylie (2014) note that distinguishing between constructive and active *student* engagement is highly challenging in problem-solving contexts (such as the one in this study). On the other hand, *teachers'* constructive and active prompts may have been more easily distinguished. Constructive prompts tended to ask for general explanations ("How did you come up with that index number?" or "Why?") while active prompts were more specific and focused on eliciting a simple yes/no or multiple choice answer ("You just counted six. Six what?"). Finally, teachers tend to express themselves more clearly than adolescents, which also makes their talk easier to code with greater accuracy.

Another possibility is that teachers' constructive prompts may provoke some constructive processing in the mind of the learner, even when students do not respond with an overt constructive verbal response. In this way, the teachers' constructive prompts may be more predictive of transfer than students' constructive talk if students are doing some mental construction without verbalizing it. For instance, in one study, participants who heard a deep-level-reasoning question before being shown the answer learned more than those who were simply shown the answer, even though participants who received the questions were not forced to explicitly answer them (Craig, Sullins, Witherspoon, & Gholson, 2006).

What kinds of teacher-student exchanges predicted transfer? We found a positive relationship between the proportion of constructive teacher-student exchanges and transfer outcomes, when controlling for prior knowledge. Given that student talk predictors were not significantly related to transfer outcomes, these results suggest that constructive teacher-student dialog may be more important than students' individual cognitive engagement. Both the ICAP framework and other theories of tutor-tutee or teacher-student interaction suggest that joint dialogs can have a strong influence on learning and transfer (Chi et al., 2001, 2008). However, compared to the relationship between teacher talk and transfer, the association between teacher-student exchanges and

transfer is more tenuous, as it did not predict unique variance in transfer scores when other predictors were added to the model.

In sum, our data generated some evidence for the CAP portion of the ICAP hypothesis in the form of teacher talk, teacher-student exchanges, and to some degree in student talk as well. First, teachers' constructive prompts positively and significantly predicted transfer, while passive prompts negatively and significantly predicted transfer. Second, constructive teacher-student exchanges positively and significantly predicted transfer, when controlling for prior knowledge. Third, while there were no significant student talk predictors of transfer, the coefficients in the final model were in the general direction predicted by the CAP hypothesis: $C > A > P$, in that constructive talk was positively related to transfer, passive talk was negatively related to transfer, and active talk fell in between. In fact, this general $C > A > P$ pattern occurred in all three sets of regressions. While these coefficients were not significant in some cases, we speculate that with a larger sample size and greater power, these effects might have been significant.

Limitations and future research

One limitation of this study is the small sample size, which raises issues of power and generalizability. For instance, it is possible that regressions of student talk and teacher-student exchanges would have yielded more significant findings, had we included more students in our study, increasing our power. However, because we had low power, in cases where we did achieve significance, the effects were sizeable (e.g., significant predictors explained 28–33% of unique variance in transfer scores). Another issue is whether these effects are likely to generalize to a broader population of students and teachers. First, it should be noted that our student population had low socioeconomic status and was very racially diverse, and our teachers were fairly experienced science teachers. Thus, these results may not generalize beyond this specific population. Also, with small sample sizes, the possibility that effects were due to chance increases. However, it seems unlikely that our effects were entirely due to chance, given that many of them align with prior ICAP research. Nonetheless, all findings and implications should be viewed as tentative, and future work should attempt replication with a larger sample of students and teachers.

Another limitation of this work is that it may only generalize to Invention activities or exploratory learning activities that are followed by expository instruction. Future work should explore cognitive modes of engagement in teacher and student talk, and their relationships to transfer, in other forms of exploratory learning activities like discovery or problem-based learning. It would

also be interesting to test whether these effects generalize to a classroom setting. In this study, teachers worked one-on-one with students. While we believe this is similar to a situation where students engage in individual or group seatwork with teachers circling the room and providing individualized guidance, we cannot be sure that these results would generalize to the broader classroom setting, particularly in larger classes with high student-teacher ratios.

Future research could also test the causal contributions of teachers' constructive, active, and passive prompts in a controlled experimental study that explicitly manipulates these forms of teacher guidance. This would give us (1) more definitive evidence of the impact of teacher talk on student learning and transfer and (2) help us disentangle the effects of constructive and passive teacher prompts on transfer outcomes.

Another interesting line of follow-up research would explore when and why student and teacher exchanges do not align. For instance, why do students sometimes respond to constructive prompts with active or passive responses? Why do students sometimes respond to passive prompts with constructive or active responses? What contextual variables account for these effects? This would further inform our understanding of how to solicit productive modes of cognitive engagement in students.

Finally, it would be interesting to explore the I portion of the ICAP hypothesis in the context of Invention and other exploratory STEM tasks. Our work focused on teacher *prompts* and student *responses*, and because teachers generated so few interactive prompts in our data set, we did not analyze them. However, an analysis that focuses on how teacher ideas were taken up by students (whether or not this was explicitly prompted by teachers) could uncover evidence of *interactive dialog*. Future work could also examine the relationship between transfer outcomes and interactive talk exchanges between collaborative peers as they engage in Invention tasks (see Mazziotti et al., 2019).

Implications

Our study has explored the role of teacher talk in influencing students' mode of cognitive engagement and their ability to transfer STEM concepts. The work makes several contributions to the ICAP framework. First, while most of the work supporting the ICAP hypothesis has been done through experimental manipulations of *task types*, we have provided evidence supporting the CAP portion of the ICAP hypothesis in the context of *teacher-student dialog*. Our study found that teachers' constructive prompts positively relate to transfer, while active prompts are unrelated to transfer, and passive prompts negatively predict transfer (with the caveat that

we cannot distinguish the relative effects of constructive and passive prompts). This suggests that the ICAP framework may be effectively extended to teacher talk and can provide a useful framework for classifying forms of teacher talk guidance. Second, in our data, evidence for the CAP hypothesis was strongest in the context of teacher talk and was found to a lesser extent in both student talk and teacher-student exchanges. While issues of low power may have hindered our ability to find significant relationships in student talk or teacher-student exchanges, this finding underscores the important role of teachers' verbal prompts for cognitive engagement in shaping student transfer outcomes. Third, our data sheds light on the tension between the kind of cognitive engagement students enact and the kind that teachers intend in teacher-student dialog. Students tended to respond with the form of cognitive engagement they were prompted for by teachers. However, alignment between teacher and student talk categories was not perfect, and the majority of student talk was active. Thus, instructional designers may need to consider other factors that could contribute to learners' mode of engagement, such as the task, learners' motivation, and other potential contextual variables. Moreover, it should be noted that passive prompts may yield a variety of student responses, and passive student talk may arise in a variety of situations. Fourth, we have generated a novel coding scheme for constructive, active, and passive teacher and student talk, which other researchers could apply. Fifth, while the ICAP framework has been effectively applied to many forms of more traditional learning activities, such as note-taking and reading (Chi & Wylie, 2014), in this study, we found reasonable support for the CAP hypothesis in the context of an exploratory Invention task.

In addition, our findings contribute to our understanding of guidance during Invention activities. First, we discovered that teachers spontaneously produced an even distribution of constructive, active, and passive forms of talk rather than privileging one particular form of prompt. This shows that while guiding Invention tasks, teachers did not favor active and passive prompts, which are common in more traditional teaching paradigms. Teachers also did not produce a majority of constructive prompts, which might be expected given the constructive nature of the task. Second, this study sheds light on how teacher guidance affects a student's ability to transfer from Invention tasks and whether Invention tasks should be guided at all. While research findings on the efficacy of guidance during the Invention phase is mixed (Chase et al., 2019; Holmes et al., 2014; Kapur, 2011; Loibl & Rummel, 2014), little research on Invention has explored teacher talk guidance. Our work shows that guidance in the form of teacher talk is associated with learners' developing transfer abilities, when the guidance

takes the form of constructive prompts rather than passive ones. While causality cannot be inferred from our correlational findings, the fact that certain teacher prompts predict student transfer performance even when controlling for prior knowledge (and when prior knowledge does not predict those teacher prompts) suggests that teacher guidance plays a role in facilitating students' transfer from Invention tasks.

Lastly, our findings have practical implications for the design of guidance for Invention. We suggest that a productive way to guide Invention tasks is for teachers to prompt students to construct novel ideas, explanations, and reasoning (e.g., How did you come up with that number? Explain your solution. Why did you decide to divide? Why did you think this bus is more crowded than this bus?). These prompts are relatively more productive than active forms of questions, which invite learners to demonstrate known facts, mastered mathematical procedures, or to repeat or paraphrase information (e.g., Count the clowns in this bus. Six what? What does this say?). Finally, prompts that engage students passively, such as teacher explanations or statements that only invite the student to agree, look, or listen, should be suppressed. While some active and passive prompts and responses are probably necessary to move the dialog along (and indeed, no teachers in our study gave only one type of prompt), we suggest that teachers attempt to use a larger proportion of constructive than active prompts and relatively few passive ones. Takeaways from this research have already impacted the design and implementation of the Invention coach, a computer-based Invention space (Aleven et al., 2017; Chase et al., 2019; Marks, Burnett, & Chase, 2016). Finally, applying the ICAP framework may be a productive way to identify forms of guidance that could facilitate transfer from other exploratory STEM learning activities.

Conclusion

In this study, we examined teacher-student dialog as teachers guided students through Invention activities, with an eye towards understanding spontaneous teacher talk guidance of Invention and identifying effective guidance of exploratory STEM learning activities. We applied the ICAP theoretical framework to explore teacher-student dialog and to classify talk into constructive, active, and passive teacher prompts, student responses, and teacher-student exchanges. While findings should be replicated with larger student and teacher populations before strong conclusions can be drawn, there were some promising findings. Results showed that teachers produced an even distribution of constructive, active, and passive prompts; however, students produced mostly active forms of talk. Although students tended to respond with the form of engagement teachers prompted, alignment was

not perfect, particularly for the passive category. Compared to student talk and teacher-student exchanges, teacher talk showed the strongest relationship to students' abilities to transfer knowledge of ratio structures to novel scientific contexts. Higher proportions of constructive teacher prompts and lower proportions of passive prompts were associated with greater transfer. Finally, teacher talk, student talk, and teacher-student exchanges all produced evidence of the CAP portion of the ICAP hypothesis (albeit descriptively, in some cases), whereby the relationship between talk and transfer outcomes followed this general trend: $C > A > P$. This work implies that the CAP portion of the ICAP framework can be extended to teacher-student dialog in exploratory STEM learning activities and that teacher talk plays a key role in eliciting student cognitive engagement. Moreover, this research suggests that teacher talk guidance for Invention should contain many prompts for students to construct—generating novel ideas, inferences, and explanations—and few prompts for passive attention—such as asking students to listen to explanations or telling them where to look, when the goal is to promote students' abilities to adapt their knowledge for novel contexts.

Endnotes

¹Given our small sample size, we conducted Cook's, Mahalanobis, and Leverage tests for outliers. We did find two outliers according to Cook's criteria. However, given that removing these outliers only magnified our effects, we chose to go with the more conservative option of keeping these two participants in the analysis. Thus, the results are slightly under-estimated.

²However, unlike constructive and passive *teacher* talk, which were each significant predictors only when the other was not contained in the model, active teacher-student exchanges did not predict transfer, even when entered into a model with pretest only.

Abbreviations

ANCOVA: Analysis of covariance; ANOVA: Analysis of variance; ICAP: Interactive, constructive, active, and passive; S: Student; SD: Standard deviation; STEM: Science, technology, engineering, and math; T: Teacher

Acknowledgements

We thank Bryan Keller for providing advice on our statistical analyses and Vincent Aleven for his help in the early stages of study design.

Funding

We thank the National Science Foundation (grant #1361062) for funding this research.

Availability of data and materials

The datasets analyzed in the current study are available from the corresponding author on reasonable request.

Authors' contributions

CC conceptualized the work, oversaw all aspects of the research, and wrote the paper. JM, LM, and HC contributed to the design of our coding scheme,

data analysis, and interpretation. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Received: 21 November 2018 Accepted: 11 April 2019

Published online: 07 May 2019

References

- Adey, P., & Shayer, M. (1993). An exploration of long-term far-transfer effects following an extended intervention program in the high school science curriculum. *Cognition and Instruction*, 11(1), 1–29.
- Aleven, V., Connolly, H., Popescu, O., Marks, J., Lammina, M., & Chase, C. C. (2017). An adaptive coach for invention activities. In *International conference on artificial intelligence in education* (pp. 3–14). Cham: Springer.
- Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, 103(1), 1–18.
- Anolli, L., Antonietti, A., Crisafulli, L., & Cantoia, M. (2001). Accessing source information in analogical problem-solving. *The Quarterly Journal of Experimental Psychology Section A*, 54, 237–261.
- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology*, 95(4), 774.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612.
- Bauer, A., & Koedinger, K. R. (2007). Selection-based note-taking applications. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 981–990). New York: ACM.
- Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction*, 13(2), 221–252.
- Bransford, J. D., Franks, J. J., Vye, N. J., & Sherwood, R. D. (1989). New approaches to instruction: Because wisdom can't be told. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 470–497). New York: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511529863.022>.
- Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, 24(1), 61–100.
- Chase, C. C., Connolly, H., Lammina, M., & Aleven, V. (2019). Problematising helps! Classroom study of an adaptive computer-based coach for invention activities. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-019-00178-y>
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences*, 6(3), 271–315.
- Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73–105.
- Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchock, C., Kang, S., Lancaster, M., et al. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive Science*, 42(6), 1777–1832.
- Chi, M. T. H., de Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439–477.
- Chi, M. T. H., & Menekse, M. (2015). Dialogue patterns that promote learning. In L. B. Resnick, C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 263–274). Washington, DC: AERA.
- Chi, M. T. H., Roy, M., & Hausmann, R. G. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, 32(2), 301–341.
- Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, 25(4), 471–533.
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243.
- Clough, E. E., & Driver, R. (1986). A study of consistency in the use of students' conceptual frameworks across different task contexts. *Science Education*, 70(4), 473–496.
- Cohen, P. A., Kulik, J. A., & Kulik, C. L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19(2), 237–248.
- Coleman, E. B., Brown, A. L., & Rivkin, I. D. (1997). The effect of instructional explanations on learning from scientific texts. *The Journal of the Learning Sciences*, 6(4), 347–365.
- Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The deep-level-reasoning-question effect: The role of dialogue and deep-level-reasoning questions during vicarious learning. *Cognition and Instruction*, 24(4), 565–591.
- Czerniak, C. M., & Haney, J. J. (1998). The effect of collaborative concept mapping on elementary preservice teachers' anxiety, efficacy, and achievement in physical science. *Journal of Science Teacher Education*, 9(4), 303–320.
- Davey, B., & McBride, S. (1986). Effects of question-generation training on reading comprehension. *Journal of Educational Psychology*, 78(4), 256.
- De Jong, T., & Van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68(2), 179–201.
- Detterman, D. K. (1993). The case for the prosecution: Transfer as an epiphenomenon. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 99–167). Norwood: Ablex Publishing.
- Doymus, K. (2008). Teaching chemical equilibrium with the jigsaw technique. *Research in Science Education*, 38(2), 249–260.
- Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34(1), 174–206.
- Furtak, E. M. (2006). The problem with answers: An exploration of guided scientific inquiry teaching. *Science Education*, 90(3), 453–467.
- Georghiades, P. (2000). Beyond conceptual change learning in science education: Focusing on transfer, durability and metacognition. *Educational Research*, 42(2), 119–139.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15(1), 1–38.
- Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31(1), 104–137.
- Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, 36(1), 93–103.
- Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal*, 30(2), 393–425.
- Holmes, N. G., Day, J., Park, A. H., Bonn, D. A., & Roll, I. (2014). Making the failure more productive: Scaffolding the invention process to improve inquiry behaviors and outcomes in invention activities. *Instructional Science*, 42(4), 523–538.
- Howe, C., & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education*, 43(3), 325–356.
- Howe, C., Nunes, T., & Bryant, P. (2010). Intensive quantities: Why they matter to developmental research. *British Journal of Developmental Psychology*, 28(2), 307–329.
- Kafai, Y., & Resnick, M. (1996). *Constructionism in practice: Designing, thinking and learning in a digital world*. Mahwah: Lawrence Erlbaum Associates, Inc.
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379–424.
- Kapur, M. (2011). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science*, 39(4), 561–579.
- King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, 31(2), 338–368.
- Kolodner, J. L., Gray, J., & Fasse, B. (2003). Promoting transfer through case-based reasoning: Rituals and practices in learning by design™ classrooms. *Cognitive Science Quarterly*, 3(2), 183–232.
- Kyriacou, C., & Issitt, J. (2008). What characterises effective teacher initiated teacher-pupil dialogue to promote conceptual understanding in mathematics lessons in England in key stages 2 and 3? A systematic review. Technical report. In *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. New York: Cambridge University Press.
- Lepper, M. R., Drake, M. F., & O'Donnell-Johnson, T. (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds.), *Scaffolding student*

- learning: *Instructional approaches and issues* (pp. 108–144). Cambridge: Brookline Books.
- Lobato, J., Rhodehamel, B., & Hohensee, C. (2012). "Noticing" as an alternative transfer of learning process. *Journal of the Learning Sciences*, 21(3), 433–482.
- Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review*, 29(4), 693–715.
- Loibl, K., & Rummel, N. (2014). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instructional Science*, 42(3), 305–326.
- Marks, J., Bennett, D., & Chase, C. C. (2016). The Invention Coach: Integrating data and theory in the design of an exploratory learning environment. *International Journal of Designs for Learning*, 7(2), 74–92.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, 59(1), 14.
- Mazziotti, C., Rummel, N., Deiglmayr, A., & Loibl, K. (2019). Probing boundary conditions of productive failure and analyzing the role of young students' collaboration. *NPI Science of Learning*, 4(1), 2.
- Menekse, M., & Chi, M. T. (2018). The role of collaborative interactions versus individual construction on students' learning of engineering concepts. *European Journal of Engineering Education*, 1–24. <https://doi.org/10.1080/03043797.2018.1538324>.
- Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education*, 102(3), 346–374.
- Moreno, R., Mayer, R. E., Spire, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, 19(2), 177–213.
- Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 510–520. <https://doi.org/10.1037/0278-7393.14.3.510>.
- Papert, S. (1991). Situating constructionism. In S. Papert & I. Harel (Eds.), *Constructionism* (pp. 1–11). Norwood: Ablex.
- Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. London: Routledge.
- Puntambekar, S., & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, 40(1), 1–12.
- Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., Kyza, E., Edelson, D., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The Journal of the Learning Sciences*, 13(3), 337–386.
- Redfield, D. L., & Rousseau, E. W. (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research*, 51(2), 237–245.
- Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences*, 13(3), 273–304.
- Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, 77(1), 1–15.
- Roelle, J., & Berthold, K. (2016). Effects of comparing contrasting cases and inventing on learning from subsequent instructional explanations. *Instructional Science*, 44(2), 147–176.
- Roll, I., Alevin, V., & Koedinger, K. R. (2009). Helping students know 'further'—increasing the flexibility of students' knowledge using symbolic invention tasks. In *Proceedings of the 31st annual conference of the cognitive science society* (pp. 1169–1174). Austin: Cognitive Science Society.
- Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(3), 456.
- Roth, W. M. (1996). Teacher questioning in an open-inquiry learning environment: Interactions of context, content, and student responses. *Journal of Research in Science Teaching*, 33(7), 709–736.
- Samson, G. K., Strykowski, B., Weinstein, T., & Walberg, H. J. (1987). The effects of teacher questioning levels on student achievement: A quantitative synthesis. *The Journal of Educational Research*, 80(5), 290–295.
- Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, 16(4), 475–5223.
- Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, 103(4), 759–775.
- Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, 22(2), 129–184.
- Schworm, S., & Renkl, A. (2002). Learning by solved example problems: Instructional explanations reduce self-explanation activity. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 816–821). Mahwah: Erlbaum.
- Shemwell, J. T., Chase, C. C., & Schwartz, D. L. (2015). Seeking the general explanation: A test of inductive activities for learning and transfer. *Journal of Research in Science Teaching*, 52(1), 58–83.
- Stigler, J. W., & Hiebert, J. (2004). Improving mathematics teaching. *Educational Leadership*, 61(5), 12–17.
- Taylor, J. L., Smith, K. M., van Stolk, A. P., & Spiegelman, G. B. (2010). Using invention to change how students tackle problems. *CBE—Life Sciences Education*, 9(4), 504–512.
- Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology*, 31(2), 207–220. <https://doi.org/10.1037/0012-1649.31.2.207>.
- Trafton, J. G., & Trickett, S. B. (2001). Note-taking for self-explanation and problem solving. *Human-Computer Interaction*, 16(1), 1–38.
- VanLehn, K. (2018). What do human tutors do? In K. A. Gluck & J. E. Laird (Eds.), *Interactive task learning: Agents, robots, and humans acquiring new tasks through natural interactions, Strüngmann forum reports*. Cambridge: MIT Press.
- Vanlehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, 21(3), 209–249.
- Vattam, S. S., & Kolodner, J. L. (2008). On foundations of technological support for addressing challenges facing design-based science learning. *Pragmatics & Cognition*, 16(2), 406–437.
- Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education*, 22(5), 366–389.
- Winne, P. H. (1979). Experiments relating teachers' use of higher cognitive questions to student achievement. *Review of Educational Research*, 49(1), 13–49.
- Wolf, M. K., Crosson, A. C., & Resnick, L. B. (2005). Classroom talk for rigorous reading comprehension instruction. *Reading Psychology*, 26(1), 27–53.
- Wong, R. M., Lawson, M. J., & Keeves, J. (2002). The effects of self-explanation training on students' problem solving in high-school mathematics. *Learning and Instruction*, 12(2), 233–262.
- Yin, Y., Vanides, J., Ruiz-Primo, M. A., Ayala, C. C., & Shavelson, R. J. (2005). Comparison of two concept-mapping techniques: Implications for scoring, interpretation, and use. *Journal of Research in Science Teaching*, 42(2), 166–184.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)