UC Berkeley

UC Berkeley Previously Published Works

Title

Why Engaging in Mathematical Practices May Explain Stronger Outcomes in Affect and Engagement: Comparing Student-Driven With Highly Guided Inquiry

Permalink

<https://escholarship.org/uc/item/94h3g7dk>

Journal

Journal of the Learning Sciences, 24(4)

ISSN

1050-8406

Authors

Sengupta-Irving, T
Enyedy, N

Publication Date

2015-10-02

DOI

10.1080/10508406.2014.928214

Peer reviewed

Why Engaging in Mathematical Practices May Explain Stronger Outcomes in Affect and Engagement: Comparing Student-Driven With Highly Guided Inquiry

Tesha Sengupta-Irving

*School of Education
University of California, Irvine*

Noel Enyedy

*Graduate School of Education & Information Sciences
University of California, Los Angeles*

This article investigates why students reported liking a student-driven learning design better than a highly guided design despite equivalent gains in knowledge assessments in both conditions. We created two learning designs based on the distinction in the literature between student-driven and teacher-led approaches. One teacher assigned each of her two 5th-grade classes to one design ($n = 52$); both designs addressed the mathematical concepts for the same amount of time. Data were collected using written assessments, surveys, and video. Students in both classes improved equivalently on assessments. On surveys, students in the student-driven condition were significantly more positive about learning. Video was coded to examine why students were more positive in the student-driven design. This analysis showed students engaging more frequently with data in discussing strategies, questioning peers, and aligning outcomes with prior experiences. We trace the association between students' positive responses to learning and richer engagement in mathematical practices to specific features of the student-driven design. Furthermore, we reject competing explanations (e.g., amount of off-task behavior or adult intervention). We conclude that designing learning opportunities that promote mathematical

---

Correspondence should be addressed to Tesha Sengupta-Irving, University of California, Irvine, Education, 3000B Education, Irvine, CA 92697-5500. E-mail: tsengupt@uci.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hlms.

practices affords opportunities to cultivate students' disciplinary interest. We discuss implications for teachers and curriculum designers who are responding to new mathematics standards.

Current mathematics education policy characterizes proficiency as grasping mathematical ideas and developing fluency in practices like argumentation and reasoning (Common Core Standards Initiative, 2011; Kilpatrick, Swafford, & Findell, 2001). In this article, we develop the conjecture that an emphasis on mathematical practices is important not only for proficiency but also for promoting greater enjoyment of learning.

Mathematical proficiency develops over many years of schooling, and whether or not students *like* mathematics often determines whether they continue working toward developing deep proficiency. Unfortunately, students generally lose interest in mathematics as schooling goes on (Fredricks & Eccles, 2002; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Köller, Baumert, & Schnabel, 2001). International and comparative studies attest that even as students gain proficiency in mathematics, they often lose interest in the subject, and that high-achieving students often express below-average interest in the subject (e.g., Beaton et al., 1996; Mullis et al., 2000; Trautwein, Ludtke, Marsh, Köller, & Baumert, 2006).

We see this as a problem: Promoting mathematical proficiency without developing students' interest (or better yet, passion) marks a failure in the design of learning opportunities. The learning sciences have, for a number of years now, explored the value of promoting *productive disciplinary* engagement as an integral part of mathematics teaching and learning (Ball, 1993; Cobb, Gravemeijer, Yackel, McClain, & Whitenack, 1997; Lampert, 1990; National Council of Teachers of Mathematics, 2000). In these formulations, students' mathematical practices—creating strategies, questioning peers, or tying abstract ideas to concrete situations—are largely analyzed in terms of cognitive processes or resulting learning outcomes.

In the present study, although there was certainly learning, there were no differences in learning outcomes between students in a highly guided or student-driven approach, as reflected in written assessments. This left us with a dilemma. We had the sense, both in the class and later when we were analyzing the surveys, that there was an important difference between the two approaches that was not evident in the written assessment. Was there value to organizing student learning around deep engagement in mathematical practices beyond the written outcomes? After all, the learning sciences have recently expanded the focus on learning to include the importance of positive dispositions and domain identities in mathematics learning (e.g., Esmonde, 2009; Gresalfi, 2009; Gresalfi & Cobb, 2006; Langer-Osuna, 2011). Thus, we were compelled by the idea that designs for instruction might increase proficiency and also increase how much students *like* learning mathematics.

We develop support for this conjecture by investigating three claims: (a) that different designs for learning the same mathematics can have different affordances for disciplinary engagement; (b) that the affordances are realized during instruction as differences in mathematical practices such as questioning, meaning making, and strategy talk; and (c) that overall differences in students' affective response to the designs are more plausibly linked to the differences in mathematical practices than to competing hypotheses. To investigate the claims, we analyze data from a teaching experiment that contrasted two pedagogical approaches to fifth-grade statistics. In the highly guided approach, the teacher managed student inquiry, with mathematical concepts formalized in a predetermined sequence early on. In the student-driven approach, students were in charge of their own inquiry process at first; later, students and the teacher formalized concepts based on the students' explorations and discussions. (This distinction between highly guided and student-driven instruction, what we call *Guided* and *Open*, respectively, is operationalized further in our Methods section.) Both approaches featured equal amounts of instructional time and used the same materials.

We collected data using three instruments: a written assessment of student knowledge, a survey of students' affective response, and video recordings of the teaching experiment. Our analysis of the written assessments found that students in both approaches showed statistically significant gains and that there was no interaction between the type of approach and the amount of gain. Our analysis of the survey data found that students in the student-driven approach, however, were significantly more positive about their learning. Through video analyses, we investigated why students in the student-driven design may have liked learning mathematics more. Our findings are that they engaged more deeply in mathematical activity, worked more consistently with data to ground assertions, and strived to make mathematical meaning. These findings suggest a link between engagement in disciplinary practices and liking mathematics.

We discuss the results of our investigation in relation to broader shifts and challenges in mathematics curriculum. As educators and designers take up the call to create learning environments that foster productive disciplinary engagement (e.g., Common Core Standards Initiative, 2011; Engle & Conant, 2002; Kilpatrick et al., 2001), we examine the important implications for advancing disciplinary enjoyment. We examined these implications by first nuancing current discussions of highly guided and student-driven pedagogies. We then operationalized this distinction, engaged a teacher in several discussions of these distinctions, and worked with her to implement a teaching experiment organized around this difference. We then use this distinction to analyze the differences in the impact of the two approaches on student-related outcomes of learning and interest. In particular, we develop the idea that the construct of mathematical practices may mediate the relationship between varied instructional designs (teacher led vs. student driven) and outcomes related to disciplinary interest while also contributing to outcomes related to mathematical proficiency. Because the level of mathematics

proficiency desired in society develops over many years, cultivating and sustaining disciplinary interest is important to longitudinal outcomes of schooling in mathematics.

LITERATURE REVIEW

To study disciplinary engagement, we draw broadly on notions of affect and disciplinary interest (Ainley, 2012) as well as on more specific principles of fostering productive disciplinary engagement (Engle & Conant, 2002). We found these principles useful as a framework for both our design *and* analyses, and thus Engle and Conant's (2002) principles connect all facets of this teaching experiment. In deciding how to vary teaching approaches, we also drew on the well-established distinction between teacher-led and student-driven pedagogies to inform and frame our contrasting instructional design.

Disciplinary Interest

As previously discussed, students lose interest in mathematics across years of schooling. Proficiency and interest have a plausible relationship. For example, in 1913, John Dewey wrote that if educators were able to "secure interest in a given set of facts or ideas, [they] may be perfectly sure that the pupil will direct his energies toward mastering them" (p. 1). Dewey suggested that learning should involve more than mastery of facts or ideas—it should also promote *interest* in the facts or ideas to be mastered.

Despite this early 20th-century call, attention to conceptually related constructs like interest, emotion, or affect in learning generally declined during the era of behaviorism and the cognitive revolution that followed (Boekaerts, 1999; Eppstein, 1990; Frenzel, Goetz, Pekrun, & Watt, 2010; Krapp, 1999). Hidi (2006) observed that in the past 25 years, psychologists and neuroscientists have reinvigorated attention to the importance of emotions, feelings, and other affective variables in learning (e.g., Ainley, Corrigan, & Richardson, 2005; Meyer & Turner, 2002; Panksepp, 2003). This has drawn attention to *disciplinary* interest (Ainley, 2012; Hidi & Harackiewicz, 2000; Hidi & Renninger, 2006; Hidi, Renninger, & Krapp, 2004), in which inquiries on interest "adopt a domain-specific approach" (Frenzel et al., 2010, p. 509). We treat disciplinary interest as characterized by increased attention, concentration, and affect during interactions with others and objects in the learning environment (Hidi, 2006, p. 70). Developing disciplinary interest is thus understood to promote student engagement and affective responses to tasks (Ainley et al., 2005; Ainley, Hidi, & Berndorff, 2002). This in turn could be important to desired societal outcomes, such as students who want to pursue studies and careers that involve doing mathematics.

For the purpose of this study of how varied instructional designs relate to affect, we have chosen to collapse differences in the psychological constructs of interest, emotion, affect, enjoyment, and so on, while recognizing that others do not. Turner and Silvia (2006), for example, argued that enjoyment is not a proxy for interest, whereas Ainley and Ainley (2011) would argue that enjoyment and interest are distinct but intertwined constructs (for relevant overviews, see Ainley & Hidi, 2014; Pintrich & Schunk, 2002; Renninger & Hidi, 2011). Still others would argue that the construct of interest is conceptually very close to those of values and enjoyment (Frenzel et al., 2010). We collapse such terms because our focus is not on identifying *what* accounts for the interrelationship of cognition and emotion but rather on illustrating *how* the interrelationship becomes evident in light of a contrasting pedagogical design. We therefore interchangeably use terms like *enjoyment* and *liking* (the latter being the literal language of students' self-reports) without laying any claim to these terms as more fundamental than the others.

Productive Disciplinary Engagement

In addition to the broad research literature on interest described previously, we have found a specific framework to be particularly useful in structuring this study. In 2002, Engle and Conant articulated four principles for fostering productive disciplinary engagement that have since served as design principles or as the basis for analyses in numerous studies, including this one (see also, e.g., Barron, 2003; Reiser, 2004; Zhang, Scardamalia, Reeve, & Messina, 2009). Learning environments embodying the four principles would

1. encourage students to take on intellectual problems (*Problematising*),
2. offer students the authority to address the problems (*Authority*),
3. hold students' work accountable to others and disciplinary norms (*Accountability*), and
4. make sufficient resources available to facilitate the three prior principles (*Resources*). (pp. 400–401)

Students who embody these principles in mathematical activity—directing inquiry to capably address problems in meaningful, collective, and disciplined ways—are more likely to be engaged learners. As we later discuss, Engle and Conant's principles provided important structure and language to video analyses of students' mathematical activity—analyses that, for example, compare how often students pursued their own problem-solving strategies (*Authority*) or asked one another questions to account for their mathematical reasoning (*Accountability*).

Learning Design: Guided Versus Open Approaches

In designing the pedagogical contrast of our study, we drew on a well-established distinction in the literature. The Guided and Open approaches were designed

to embody the principles of teacher-led and student-driven inquiry, respectively. In particular, prior research distinguishes among “traditional” and “reform” or “skills-oriented” and “inquiry-oriented” teaching (e.g., Boaler, 2002; Hiebert et al., 1996; Sherin & Drake, 2009). In this section, we discuss prior work that broadly informed our designed contrast; the contrast is operationalized, however, in the Methods section.

Mathematical reforms have prioritized opportunities for students to engage with authority in problem solving. The Dutch approach known as Realistic Mathematics Education (Freudenthal, 1979) has spread globally and thus serves as a good example. Realistic Mathematics Education relies on richly contextualized problems to promote conceptual understanding through inquiry (Case, 2005; Gravemeijer & Doorman, 1999; Hadi, 2012). We note that the reforms align with the principles proposed by Engle and Conant (2002), which serve as our guiding design and analytic framework.

In commonplace usage, many people see a Guided approach as featuring explicit instruction with a small amount of directed exploration, whereas they see Open as designed for students to generate strategies and engage in their own explorations (e.g., Klahr, 2009; Klahr & Nigam, 2004). Although some would characterize the former approach as learning by acquisition and the latter as learning through participation, Sfard (1998) warned that single-minded devotion to the metaphor—acquisition *or* participation—can lead to theoretical excess and impoverished practices. Thus, our design avoids strict adherence to a dichotomy.

Prior research on the effectiveness of student exploration and invention in shaping conceptual reasoning (e.g., Boaler, 2002; Greeno, 2006; Hiebert et al., 1996; Horn, 2006) was also influential in our thinking. An invented strategy or modeling approach posits that the iterative process of creating and refining strategies leads to increasingly sophisticated mathematical understandings (Enyedy, 2003; Lehrer, Strom, & Confrey, 2002). By refining strategies and the ideas implicit in their graphic and symbolic representations, students construct a complex set of personally meaningful relationships between the model and what it is intended to represent (Lehrer & Pritchard, 2002). Moreover, the modeling approach addresses one of the most pernicious beliefs in math teaching and learning: When encountering a problem, students should have a known strategy to produce an answer quickly (Carpenter, as cited in Schoenfeld, 1992). Kapur’s (2008) study of *productive failure*, however, offers an important counter to such beliefs. Kapur demonstrated how students engaged in ill-structured (minimally supported) collaborative tasks who then work individually on well-structured tasks benefit in later learning over those who only engage in well-structured tasks. As Kapur (2010, p. 546) wrote, the idea of productive failure challenges the basic argument used against student-driven inquiry: “Why waste time letting learners make mistakes when you could give them the correct understandings?” In fact, despite the appearance of inefficiency of learners in invention activities, doing such explorations before direct instruction can promote achievement

(e.g., diSessa, Hammer, Sherin, & Kolpakowski, 1991; Schwartz & Martin, 2004) and motivation (e.g., Samarapungavan, Patrick, & Mantzicopoulos, 2011). In our study, we used these ideas by thinking of the most salient distinction in the pedagogical approaches in terms of structure and support: To maintain a contrast between Open and Guided meant being aware of when to provide (in Guided) or withhold (in Open) scaffolding for students' inquiry and problem solving.

In contrast to student-driven inquiry, some see highly guided instruction as more efficient. Kirschner, Sweller, and Clark (2006), for example, argued that minimally guided forms of instruction are likely to be ineffective based on how human cognitive structures are organized. The authors argued that learning, defined as changes in long-term memory, is best accomplished when instruction fully explains concepts and procedures to students. Fully explaining concepts and procedures—hallmarks of highly guided instruction—is in part why direct instruction often supports teachers' sense of efficacy (Smith, 1996), although we describe the opposite effect elsewhere (Sengupta-Irving, Redman, & Enyedy, 2013). The sense of efficacy gleaned from telling students what they should understand, however, is not the same as students actually learning more effectively. In response to Kirschner et al., Kuhn (2007) argued that learning is improved when students identify a question or problem for inquiry and that the act of defining one's own inquiry is foundational to motivation within a discipline. Likewise, Schwartz and Bransford's (1998) experimental design comparing undergraduate psychology students' abilities to learn concepts by either exploring contrasting cases first or reading explicit comparisons of cases revealed that active engagement created its own "time for telling," whereas explicit instruction was simply "more telling." Schwartz and Bransford argued that "telling" (formalizing concepts) is more effective *after* students have the opportunity to develop their own relevant understandings; that is, telling is not antithetical to constructivist models of instruction, but the timing or sequence of telling is what matters. The work of Schwartz and Bransford is closely related to Kapur's notion of productive failure discussed earlier. In our study, these considerations play out in the contrast between conditions. In the Open approach, formalization *followed* students exploring and strategizing around differences in data sets. In the Guided approach, formalization *led* mathematical inquiry, as comparisons were made explicit first.

In summary, the literature suggests two key dimensions in designing a contrast between an Open and a Guided approach: (a) how structure and support are provided and (b) the timing of when formalization occurs.

METHODS

This study examines the following three claims: (a) that different designs for learning the same mathematics can have different affordances for disciplinary engagement; (b) that the affordances are realized during instruction as differences

in mathematical practices such as questioning, meaning making, and strategy talk; and (c) that overall differences in students' affective responses to the designs are more plausibly linked to the differences in mathematical practices than to competing hypotheses (e.g., off-task behavior, teacher intervention).

The main contrast in the study is between an Open and a Guided teaching approach. Despite the contrasting design principles (operationalized below), it was important that the teaching approaches maintain certain similarities to facilitate a valid comparison. First, we wanted both treatments to cover the same mathematical content, although we designed for more control in Guided. Second, we wanted the time students spent in whole-class conversation and small groups to be about even. If one treatment had more group work, for example, that alone might position one treatment as more favorable to students than another. Third and finally, we wanted the overall time spent on the curriculum to be even. This would ensure a fair comparison of assessment results (i.e., neither treatment received more time for discussion).

Design of the Mathematical Task

For this study, we chose a task in the mathematics curriculum strand about data analysis and statistics. The task, which has been used extensively in prior research, focuses on how to characterize the central tendency in a varied data set. The concept of central tendency will be familiar to readers as mean, median, and mode. A classic mathematical task for students dealing with this concept asks students to find a quantitative comparison between two data sets with the sample number of data points (sample size), where the data set is constructed so that qualitative perceptual comparisons between contrasting data sets generate conflicting results.

In particular, the task common to both approaches was focused on measures of central tendency with equal sample sizes to support additive reasoning (Cobb, 1999; Konold, 1995; Konold et al., 2002). This choice was made to align our instructional objectives with the California State Content Standards for 5th Grade Data, Statistical Analysis, and Probability (California Department of Education, 1999), which requires students to compute and compare statistical measures of center (mean, median, mode) and, to a lesser degree, spread (minimum, maximum, shape of distribution).

We focused on some of the common challenges associated with deceptively simple statistical concepts. For example, a mostly consistent finding in the literature is that students have difficulty reasoning about the data set as a whole. Instead, they are inclined to reason from individual, meaningful data points (e.g., their friend's favorite color or a data point they personally observed; Konold & Higgins, 2003). As late as the eighth grade, students are found to use individual data points (often opposite, extreme values) to make comparisons between two

groups (Bright & Friel, 1998). Even when taught how to calculate the mean, students struggle when the value for the mean has no concrete correspondence to a point in the data set (Hancock, Kaput, & Goldsmith, 1992). Cobb (1999) argued that because students expect the value of the center to be represented in the data, and to be meaningful in some way, mode and median are often more intuitive starting places for students.

The lessons were organized around the task for five sessions. Students were asked to justify which of two search engines (bing or Google) was the better choice for marketing an imaginary product in a 5-day advertising campaign. The students were told that they had no control over which 5 days the ads would appear, including the possibility of nonconsecutive days. Working in teams of three or four, students were provided 30 days of website traffic data for each site (i.e., number of hits) in a table and in multiple graphs that were intended to elicit different statistical analyses (see Figures 1–3). The table of values, not depicted, was expected to prompt students in determining the total number of hits for each data set (bing vs. Google). Column graphs drew attention to a data set's minimum and maximum values, range, and mode (see Figures 1 and 2), whereas a bar chart was meant to draw attention to mean and median (see Figure 3). In all, students were provided with one table and three graphs for each website.

Although modeled after research-based data and statistics lessons administered to seventh-grade students (Cobb & McClain, 2001, 2004), the data set was created to produce whole-number results for most numeric measures (e.g., total, mean, range). Moreover, the data sets were created to yield a nearly split decision, with

FIGURE 1 Column graph representing 30 consecutive days of bing website traffic.

FIGURE 2   Column graph of frequency for Bing website traffic over 30 days.

FIGURE 3   Bar chart ordering 30 days of Bing website traffic from lowest to highest value.

four of the expected measures indicating Google as the better site and three of the measures indicating Bing (see Table 1).

The statistical split decision reflects the problem inherent to the task: Students may infer very different solutions from the data set based on their (highly guided

TABLE 1
Comparison of Websites by Statistical Measure

<i>Measure</i>	<i>Better Site</i>	<i>Possible Rationale</i>
Minimum	Google	Has higher minimum (22,000 vs. 17, 000)
Maximum	bing	Has higher maximum (75,000 vs. 70,000)
Total	bing	Has higher total (1,368,000 vs. 1,329,000)
Range	Google	Lower range (48,000 vs. 58,000), which may be proxy for “consistency” of hits
Mean	bing	Higher mean (45,6000 vs. 44,300)
Mode	Google	Modal value is higher (57,000 vs. 52,000)
Median	Google	Median value is higher (50,000 vs. 45,000)

or self-guided) inquiry, which could offer opportunities for extended discussion of problem-solving strategies, for the questioning of peers on their reasoning and justifications, and for students to strive in making meaning of their mathematical results.

Operationalization of the Contrast Between the Guided and Open Approaches

The Guided and Open approaches were designed to embody the principles of teacher-led and student-driven inquiry, respectively. This distinction does not suggest that our conception of Open, as Hmelo-Silver, Duncan, and Chinn (2007) noted, is learning *without* guidance, as in *discovery* learning. Rather, the nature of scaffolding and its timing differentiate these approaches. Our development and contrast of ecologically valid pedagogical approaches is akin to the work of Eysink et al. (2009). In their study of multimedia learning approaches, the authors delineated four kinds of teaching: (a) hypermedia, (b) observational learning, (c) self-explanation-based learning, and (d) inquiry learning. What we term *Guided* instruction approximates their description of *observational learning*, whereas our *Open* instruction approximates what they describe as *inquiry learning*. The investigators argued that learners in the latter two teaching approaches (self-explanation-based and inquiry learning) had significantly higher assessment outcomes than students in the other environments (hypermedia and observational learning).

Some of the pedagogical design principles are unique to an approach, whereas others are common to both but appear in a different sequence. We provide an overview of the design principles in Table 2, followed by further explanation of their strategic use in each treatment.

The first two principles in the Guided approach (*problem given* and *guided seeing*) distinguished this approach from Open at the outset. Giving students the

TABLE 2
Design Principles of the Guided and Open Teaching Approaches

No.	Guided Approach to Teaching	Open Approach to Teaching
1	<i>Problem Given</i> Problem is explicitly defined (students are told what to explore and how to explore it)	<i>Problem Articulation</i> Students explore data displays and debate why the answer may not be easy to see; students define the problem to address
2	<i>Guided Seeing</i> Students are led to and through the math concepts (e.g., “What value occurs most frequently?”)	<i>Invention</i> Students invent the solution path (e.g., “What matters to us?”)
3	<i>Formalization</i> Students are given the formalized name for what they have been led to see (e.g., “The value that occurs most often is called the mode”)	<i>Exploration and Discussion</i> Students debate, experiment, and refine their approach to the problem until reaching (near) consensus
4	<i>Explanation and Ownership</i> Students explain the concepts in their own words and relate them back to the problem statement	<i>Best Inference</i> Students produce a solution and justification to answer the problem statement
5	<i>Best Inference</i> Students produce a solution and justification to answer the problem statement	<i>Formalization</i> Student pathways are named for their formal mathematical foundations (mode, mean, median, range)
6	<i>Practice and Application</i> Students review the mathematical concepts by applying them to new problems as concepts are introduced daily	<i>Practice and Application</i> Students review the mathematical concepts by applying them to new problems at the conclusion of all lessons
7	<i>Assessment</i>	<i>Assessment</i>

problem (i.e., telling them that data representations and statistical measures will not uniformly identify the same site as best) meant guiding their attention toward particular graphs and features of the graphs. Given that representations offer different kinds of information (Stenning & Oberlander, 1995), which makes some information seem more salient than others (Larkin & Simon, 1987), this kind of scaffolding is similar to the idea of providing representational guidance (Suthers & Hundhausen, 2003). In this study, these principles were embodied when, for example, the teacher distributed the table of values *before* proposing that groups consider how to compare the lists and what the comparison might suggest about the task.

The next two principles in Guided (*formalization* and *explanation and ownership*) are related. Formalization appears near the onset of students’ inquiry; the timing of this principle near the start furthers the distinction between highly

guided and self-guided inquiry. For the latter, in contrast, formalization appears *after* extended opportunities for students to pursue their own lines of inquiry. In Guided, this principle was embodied when groups shared their ways of comparing the lists of values and the teacher provided the formal name and procedure (e.g., “the difference between highest and lowest values—highest minus lowest—is called the range”). The teacher would then send students back into groups to talk about range, how it is calculated, and what it means in their own words (i.e., explanation and ownership) before having all students pursue its calculation. This is akin to scaffolding meta-cognitive and problem-solving processes related to the task and the determination of its solution (e.g., Ge & Land, 2003; Lin, Hmelo, Kinzer, & Secules, 1999).

Like formalization, the fifth principle (*best inference*) is not unique to Guided, but its timing again provides an important contrast to Open. After directing students to features of the graphs, introducing concepts formally, and allowing for students to reconsider formal definitions in their own words and ways, the teacher guides students to the logical conclusion of their guided inquiry: “Which website does this statistical measure suggest is best?” As was discussed previously, each representation and statistical measure maps to a particular best inference regarding the websites. In the end, comparing inferences across measures could lead to a final recommendation.

Participants

The data reported in this article were collected during a teaching experiment at a progressive elementary school in southern California. The school was ethnically and racially diverse, reporting a student population that was 36% Caucasian, 32% Latino/a, 14% Asian, 10% African American, and 8% other.

The teaching experiment involved all fifth-grade students at the school, who were taught by Ms. Parker and periodically aided by two assistants. At the time of the study, Ms. Parker had 15 years of experience teaching math and science at the elementary and middle school levels. Ms. Parker’s typical approach to teaching math involved introducing a concept in relation to prior knowledge, demonstrating a new procedure or algorithm, providing individual and group opportunities for practice, and assessing students on written tests. Ms. Parker self-identified as an equity-minded educator and was seeking opportunities to develop her teaching, which led her to engage in the study. Of her 54 students divided in two sections, 52 consented to participate in all aspects of the study (two could not be videotaped or interviewed).

Procedure

Assignment of students to condition. One section of students ($n = 25$) experienced the Guided approach, whereas the other ($n = 27$) was taught

in the Open format. Students were told that the experiment was focused on their mathematical thinking and collaborations; students were not explicitly told which teaching approach they were experiencing or that there was a difference between sections.

Specification of the teacher's role. The teacher's role in successfully maintaining the principles and objectives of a teaching experiment is critical. In preparation, we worked formally with the teacher during two 55-min sessions. In the first session we discussed the design principles of each teaching approach, highlighting overlaps, sequencing, and key distinctions. We also discussed how her response to student questions would vary by treatment, such that the Guided approach may elicit a definitive response whereas the Open approach would likely involve redirecting students' attention to the data and discussions with peers. In the second session we presented the task, representations, and suggested daily lesson plans for each treatment. Each lesson plan included the learning objectives, materials needed, time allotment for activities, content to be covered, and potential student questions with possible teacher responses. Ms. Parker provided feedback on time allotments—suggesting what might take more or less time than we had imagined—and the representations, encouraging two separate sets of data (one for Bing, one for Google) rather than our initial draft depicting data from both websites on each graph. In addition to these two formal sessions, we informally exchanged e-mails and discussed the design and implementation of the experiment throughout. Subsequent to these sessions Ms. Parker shared the design of the teaching experiment with her teaching assistants, who also interacted with students during the study.

During the experiment, Ms. Parker was charged with two main responsibilities: (a) maintaining the designed differences between Open and Guided and (b) teaching the mathematical concepts. For the first, how well students experienced Open as exploratory was related to Ms. Parker's redirection of students seeking a single formal answer to having them develop strategies, debate their merits, and explore their outcomes. For Guided, something Ms. Parker felt more familiar with, she had to adapt her usual approach to include more opportunities for student ownership of concepts and the articulation of their understanding. In terms of teaching mathematics, in whole-class discussions she highlighted particular strategies, made their mathematical underpinnings transparent, and led students in reflecting on the elegance and efficiency of their ideas throughout. In small-group discussions, Ms. Parker would validate students' approaches while pushing them to think through the data, often saying, "Okay, that makes sense from what you know and feel emotionally, but what do the data tell you?" Finally, Ms. Parker's knowledge of students' personalities and prior math experiences was as important in her response to students as any of the suggestions we offered.

Maintaining equal time for both approaches. All 50 hr of video were coded to mark when students engaged in whole-class discussions, in peer-based small-group discussions, or in conversation with the teacher (i.e., the teacher working directly with a group). In all of these respects, the two teaching approaches were closely matched compared to the total time of just over 5 hr (see Table 3).

Expected contrast between approaches. The first principle of the Open approach (*problem articulation*) allows students to determine the problem themselves. This principle was embodied when the teacher began the class by providing groups with all of the representations at once and told them to explore the information in light of the task. The goal was to take advantage of prior research that speaks to the benefits of students contrasting cases—here, comparing representations—before a more formal explanation is provided (Schwartz & Bransford, 1998).

The second and third principles, *invention* and *exploration and discussion*, are effectively the hallmarks of this treatment: Together they refer to students generating ways of working with the data in response to the problem they have articulated for themselves and negotiating how, why, and which inquiries to further pursue. This notion of students pursuing their own investigations is quite common in established problem-based mathematics curricula (e.g., Interactive Mathematics Program or Connected Mathematics Project). In Open, these principles were embodied in small-group discussions of potential solution pathways, use of particular representations, and even use of prior knowledge of the websites to reason through various strategies. Students' problem-solving paths, the representations they drew on (or not), were largely left to the group; the teacher provided support as needed and at the students' pacing.

The fourth and fifth principles of Open (*best inference* and *formalization*) have been briefly addressed in relation to the Guided approach because, though they differ in timing, they are common across treatments. In the Open treatment, for example, best inferences develop based on students' chosen inquiries in relation to the task. Unlike Guided, this provides greater space for novel problem-solving

TABLE 3
Comparison of Time Spent in Various Activities by Treatment

Activity	Guided	Open	Difference
Whole class	3:02:02	3:19:08	(0:17:06)
Small group	2:02:50	1:51:45	0:10:05
Adult intervention	1:24:49	1:15:40	0:09:09
Total time	5:04:52	5:10:53	(0:06:01)

Note. Data are hours:minutes:seconds.

approaches and for development of unique explanations to particular conclusions. In Open, best inference is followed by formalization, which is when the teacher reviews students' problem-solving pathways and identifies their connection to formal statistical procedures and measures (e.g., "The cutoff point that you are working with is formally called the median, which is the value that separates the higher half of data from the lower half").

Finally, students in the Guided approach completed practice problems daily and in alignment with the target concepts of the day. For Open, because the order in which students would discuss or explore concepts was asynchronous, practice problems were completed at the end of the lessons once the concepts had been formalized. All students, regardless of treatment group, were provided the same pre- and postassessments.

Instruments

Three instruments were used to collect data: (a) an assessment of students' knowledge administered before and after the intervention; (b) a survey, administered after the intervention; and (c) video recording throughout the intervention.

Assessment. The assessment was administered before and after the experiment, and all student work was collected. The pre- and postassessment had four multiple-choice and three free-response questions. The multiple-choice questions were duplicated or modeled after items released from the National Assessment of Educational Progress online tool. Items were written for eighth-grade mathematics (three grades above of our students' grade) in the area of data analysis and probability. The items selected were limited to the National Assessment of Educational Progress's Medium and Hard levels of difficulty (not Easy) in order to prevent ceiling effects. One free-response question was a simple adaptation of a multiple-choice question in order for students to explicitly share their calculations. The remaining two free-response questions were adapted from prior research (Cobb & McClain, 2001, 2004). The first question asked students to determine which of two cars was safer based on their braking distances shown in a bar graph. The other question asked students to evaluate the effectiveness of a recycling program based on scatter plots showing the amount of trash produced before and after the intervention. These items were selected for free response to make student reasoning with data more evident, which aligned with their original use as instructional activities in a seventh-grade statistics unit.

Survey. Students completed free-response written surveys asking them to reflect on aspects of the lessons that they liked and disliked, how well they worked in groups, and their impressions of the task. Twenty-four of the 25 students in Guided completed the survey and 27 of 27 students did so in Open. The surveys

were intended to provide a more detailed record of what students were responding to in the experiment, as well as the nature of their responses (i.e., positive affect or negative affect).

Video analysis. All lessons during the teaching experiment were videotaped. The unit lasted 5 days, and each session was approximately 55 min long. Five groups of three or four students were videotaped in each condition ($n = 39$), which yielded approximately 50 hr of video total. We purposefully sampled three groups from each condition based on standardized mathematics test scores. We wanted heterogeneous—though not highly stratified—groupings, wherein the perceived status of the students might affect the way in which the students engaged one another. On the one hand, we prioritized heterogeneity in prior achievement because heterogeneous grouping in mathematics yields academic and social benefits (e.g., Burris, Heubert, & Levin, 2006; Cohen & Lotan, 1997; Linchevski & Kutscher, 1998). On the other hand, we wanted to examine groupings for which the perceived differences in ability were not so great that students dismissed others' ideas out of hand or acquiesced to another's idea without examining it first (see, e.g., Sengupta-Irving et al., 2013). This led us to eliminate groups that had either too little or too much variation in test scores. In the Open condition the standard deviation of students' standardized test scores by group showed a nearly bimodal pattern ($SD = 7.9, 7.9, 10.8, 20.8, 26.4$); the upper two groups were eliminated as too highly stratified. In the Guided condition, the standard deviation of students' standardized test scores was 3.4, 6.2, 17.7, 17.7, and 18.1, which was also somewhat bimodal. For Guided, the lowest group was eliminated for homogeneity alongside another ($SD = 17.7$) because of poor audio quality for the level of microanalysis presented here.

DATA ANALYSIS

Scoring Assessments

Two independent researchers scored students' assessments. Free-response questions were scored 0, 1, or 2 based on clarity and mathematical accuracy. Scorers reached 87.5% interrater reliability; an average score was awarded for each item (for a total of 12 points possible on the exam).

Coding and Analyzing Surveys

The first two survey questions asked for students' overall impressions of the teaching treatment: What did you like about this week's lessons? What did you dislike about this week's lessons? The questions were coded to determine the extent to

which students' responses related to the design principles of each treatment articulated in Table 2. Three additional codes developed. The first code characterized student responses referencing liking (*Positive Affect*) the lessons without specific detail ("It was fun!"). The second code characterized student responses referencing disliking (*Negative Affect*) the lessons without specific detail ("It was kinda boring!"). The third code characterized issues related to social interaction or group work (*Interpersonal*). Two independent researchers coded the surveys with 93% interrater reliability. We then used Fisher's exact test on these categorical data to determine whether there were statistically significant differences in how codes were ascribed by treatment. Results of survey coding show the frequency of each code as a percentage of the total responses.

Developing Codes for Video

We then conducted a focused analysis of video to investigate why student preferences for one or the other condition might have emerged. Codes were iteratively and inductively developed using Studiocode software. We began with a relatively wide net of possibilities that addressed affective features of behavior (e.g., joking around, boasting, expressing frustration) and cognitive features associated with mathematical proficiency (e.g., procedural fluency or adaptive reasoning). In the affective domain, not giving in to high inference, cultural interpretations of affect was a concern; in the cognitive domain, unduly parsing students' talk as demonstrating conceptual understanding versus adaptive reasoning, for example, was a concern. In short, attempting to disentangle what were clearly interrelated affective and cognitive dimensions of learning proved to be analytically ineffective.

Engle and Conant's (2002) principles for fostering productive disciplinary engagement greatly informed our subsequent iterations of coding. For example, coding for students' problem-solving strategies, revisions to strategies over time, and other related practices (e.g., questioning, justifying) fit what Engle and Conant referred to as *Authority*. Specifically, authority was operationalized as interactions or events reflecting students' agency to define, address, and resolve problems. Therefore, in our coding we selected instances when students *first* offered or developed a unique strategy for determining the best website; occurrences of this code (*Strategy Talk*) represent the frequency of *new* mathematical ideas being raised, which relates to students' expressions of authority in each treatment.

Accountability, according to Engle and Conant (2002), refers to the ways in which students' ideas are held up to norms of the discipline and those established by the classroom community (the latter being like what Yackel & Cobb, 1996, referred to as *sociomathematical norms*). This was prolific throughout our initial coding. For example, students asking one another questions about strategies or procedures, revising one another's understandings, or explaining particular

concepts were seen as potential moves of accountability. We revisited instances that related to accountability, specifically, instances initially coded *Questioning*, *Commenting*, *Revising*, and *Explaining*. The first code stood apart because it clearly reflected a move of accountability; these moves began with a clearly stated question directed at peers and related to what was occurring in that moment, what Engle and Conant referred to as students expecting to “consult others in constructing their understandings . . . an expression of the value that each member of learning community is not an authority unto himself” (p. 405). Therefore, occurrences of this code (Questioning) represent the frequency of students holding one another accountable to disciplinary or classroom-based norms through talk.

In their discussion of *problematizing* content as a principle of productive disciplinary engagement, Engle and Conant (2002) drew specifically on problem solving in mathematics to explain the principle as allowing for problems to “open from the perspective of students interpreting them, using their available knowledge and resources” (p. 404) rather than relying on experts in the discipline to define them. In our initial coding, we marked instances that reflected when students drew on prior knowledge to contextualize, rationalize, or otherwise make sense of their problem-solving strategies or solutions. This idea—what we coded as *Meaning Making*—is central to the National Research Council’s (Kilpatrick et al., 2001) definition of conceptual understanding, studies of mathematics teaching and learning, and examinations of the ties between formal mathematics and students’ lived cultural experiences (e.g., Civil, 2006; Enyedy & Mukhopadhyay, 2007; Gutstein, 2012; Hiebert, 1997). Therefore, this code defines instances when students made explicit reference to data and connected this reference to what was known beyond the given task or materials.

Building on these primary descriptive codes (Strategy Talk, Questioning, and Meaning Making), we also considered the extent to which these markers of productive disciplinary engagement were grounded in students’ use of mathematical representations and data. We distinguished, for example, the more general question “What are you doing?” from the more specific “How are you using the maximum value in this graph?” Both questions are moves of accountability, but the latter instance illustrates specific attention to data. To this end, we developed two associated subcodes: *Using Data* and *Not Using Data*. When codes co-occur with subcodes, results are reported as a percentage of the total codes subcoded.

Finally, we added two additional codes that we believe help contextualize productive disciplinary engagement that are not specifically addressed in Engle and Conant’s (2002) framework. First, we considered instances when an adult (teacher or teaching assistant) appeared in the video and engaged a group directly. Such instances were coded *Adult Intervention* and subcoded *Content* or *Management* based on what the adult was primarily focused on during the interaction. Instances when an adult addressed both students’ behavior and their mathematical work were double coded and are reported in the results as a separate category. Second,

we considered the possibility that students reported liking the Open treatment because self-guided inquiry was allowed to devolve into opportunities of task avoidance or goofing off.

In the next section, video coding results are reported by treatment in a number of ways: an average time across groups (which is the same as averaging across cameras); an average time per group (camera); an average number of instances of a code per group (camera); or, for subcodes, a percentage of total codes.

RESULTS AND DISCUSSION

Recall the three claims at the center of our work: (a) that different designs for learning the same mathematics can have different affordances for disciplinary engagement, (b) that the affordances are made visible during instruction and through students' engagement with mathematical practices (e.g., questioning, making meaning), and (c) that overall differences in students' affective response to the designs are more plausibly linked to the differences in mathematical practices than to competing hypotheses (e.g., instructional time on task, amount of off-task behavior or teacher support). In what follows, we detail the results of scoring students' written assessments and analyses of student surveys, followed by analyses of classroom video data.

Assessments Reveal Similar Knowledge Gains for Both Groups

Students in both the Guided and Open treatments showed statistically significant gains from pre- to postassessment: Guided, $t(24) = 4.8235, p < .00001$; Open, $t(26) = 4.7382, p < .00001$ (see Table 4). Neither treatment, however, yielded significantly larger gains over the other, $t(51) = 0.9162, p = .364$.

These results were in keeping with the design's intention that students in both treatments gain proficiency in the same mathematical content.

Surveys Show Students Liked the Open Condition More

For Guided ($n = 24$), 47 responses were coded (one question was left blank); for Open ($n = 27$), 51 responses were coded (three questions were left blank). The

TABLE 4
Statistically Significant Learning Gains for Each Treatment From Pretest to Posttest

<i>Treatment</i>	<i>Pretest M (SD)</i>	<i>Posttest M (SD)</i>	<i>t test</i>
Guided	4.2 (2.4)	7.4 (2.6)	$t(25) = 4.8235, p < .00001$
Open	3.4 (2.6)	6.8 (2.5)	$t(26) = 4.7382, p < .00001$

TABLE 5
Results of Survey Coding Distribution Using Fisher's Exact Test by Treatment

<i>Code</i>	<i>Guided (n = 47)</i>	<i>Open (n = 51)</i>	<i>p</i>
Invention	0 (0)	3 (5.9)	.24
Exploration and Discussion	2 (4.3)	7 (13.7)	.16
Best Inference	1 (2.1)	5 (9.8)	.21
Practice and Application	3 (6.4)	0 (0)	.11
Formalization	4 (8.5)	0 (0)	.049*
Formal Assessment	1 (2.1)	0 (0)	.48
Interpersonal	3 (6.4)	10 (19.6)	.029*
Positive Affect	5 (10.6)	14 (27.4)	.043*
Negative Affect	5 (10.6)	0 (0)	.023*

Note. Data are number of codes (%).

* $p < .05$, two-tailed.

range of answers students provided to the open response survey questions resulted in relatively few specific mentions of the design principles. Using Fisher's exact test, we compared the responses by treatment (see Table 5).

As Table 5 reflects, ascription of the codes Formalization, Interpersonal, and Positive and Negative Affect emerged as significantly different across treatments. Of all of the design principles (see Table 2), only Formalization emerged as significantly different in student responses. This implies that using formal statistical language and procedures had a more prominent position for students in Guided, which reflected our intentional design. The Interpersonal code emerged significantly more often in Open responses, which could be a result of that treatment being designed for greater discussion and exploration of peers' problem solving. Results for the remaining two statistically significant codes—Negative Affect and Positive Affect—together suggest that students in Open were significantly more positive in their reflections on the experiment than students in Guided. Not only were students in Open more often positive overall, but *none* of the Open responses were coded as conveying negative affect.

Video: What Observable Interactions Might Lead to Students Liking Learning?

Videos were coded to better investigate why students might like learning in one condition better than the other. We begin by turning our attention to the results that pertain to productive disciplinary engagement (Strategy Talk as authority, Questioning as accountability, and Meaning Making as problematizing). We then reject some obvious conjectures for why some students liked the Open treatment better that do not have to do with productive engagement (i.e., Off-Task Behavior and Adult Intervention).

Strategy talk using data is higher with Open approach. Recall that the code Strategy Talk represents the frequency of *new* mathematical ideas being raised in small-group discussions, which suggests greater engagement with content and the agency to define, address, and resolve problems as part of productive disciplinary engagement. For Guided, Strategy Talk often arose when students were explicitly asked by the teacher to describe a statistical measure to the entire class (e.g., “What does it mean to take an average?”) or when working in small groups as prompted by a worksheet. For Open, Strategy Talk arose spontaneously during small-group discussions with students brainstorming, refining, or commenting on possible ways of analyzing data. Table 6 reports the average number of instances of strategy talk and the percentage that involved specific reference to data.

As might be expected because of a potentially higher number of opportunities to generate strategies, Table 6 shows students in Open discussing mathematical strategies more frequently than students in Guided. It does not follow, however, that the percentage of talk referencing data would be higher in Open. This latter outcome implies that strategy talk was a more mathematically rich endeavor in Open, whereas students in Guided discussed mathematical strategies less often and overwhelmingly did so without reference to the data.

To illustrate strategy talk in Open, we offer excerpts of a group’s conversation taken from the first day of lessons. Ms. Parker provided groups with all of the data and asked them to begin exploring. Initially, the group members comment that bing seems to have more traffic than Google until Luca notices that the distribution is more erratic than it may seem at first glance.

The students seem surprised that bing would appear more popular than Google (Lines 1–4), going so far as to question the validity of the data themselves (Lines 7, 8). The important moment in this, however, is when Luca notices that despite the higher peak values in the bing distribution, there are also days in the distribution when website traffic on bing is quite low (Line 6), which Jared then affirms by looking at a different representation (Line 10). Comparing maximum and minimum values represents two possible strategies for determining which website is best. As Luca begins to work out a strategy on his own (Line 11), the others continue discussing how they know that Google is better than bing based on personal

TABLE 6
Average Number of Instances Coded Strategy Talk by Treatment With Percentage
Subcoded Using Data

<i>Treatment</i>	<i>Average Instances of Strategy Talk</i>	<i>Percentage Subcoded Using Data</i>
Guided	12	3
Open	29	59

Note: Averages reported from five lessons and three cameras per treatment.

- 1 Adriana What?! Oh my God. There's more people on Google—on bing, than Google.
- 2 Luca I know.
- 3 Adriana This is sad.
- 4 Luca Wow.
- 5 Jared A lot of people use Google.
- 6 Luca Actually, that's not true, you know what?
- 7 Jared This is fake!
- 8 Luca Yeah, it is. But bing also has days when there's like no one on it. Google always has people on it.
- ... Parker [Ms. Parker interrupts to remind the class to look at all of the data]
- 10 Jared Well, the lowest for bing traffic ordered by low to high is 17 and the lowest for Google is 22. But they beat it in highest—it's 75. ← *Strategy 1: Compare minimum and maximum values*
- 11 Luca Yeah, one thing we need to do is—wait, hold on.

experience. The conversation is interrupted when Adriana leans over to see what Luca is working on.

Luca's strategy is to determine the total website traffic for bing, which he does by calculating a running total of data from the table of values (Line 24). Aside from Jared's previous comment in Line 10, this moves the group toward strategies that numerically quantify a comparison rather than on relative visual comparisons alone (e.g., Lines 1, 8, 10, 21, 22). Whereas Jared commends the idea (Line 30), Adriana seems taken aback by Luca's strategy (Lines 27, 29), but also understanding that the next logical step would be to determine the total for Google (Line 32).

- 20 Luca Hold on; hold on. 77, 78, 79 [counting on fingers]
- 21 ...
- 21 Adriana The highest days are—
- 22 Jared Okay, a lot of people are on, more people are on bing.
- 23 Eduardo I know.
- 24 Luca Hold on, I'm adding them together. ← *Strategy 2: Determine total website traffic*
- 25 Adriana You're adding all of them?
- 26 Luca ... 96; hold on ...
- 27 Adriana You're adding all of those numbers together?!
- 28 Luca ... 102, yeah, it's easy.
- 29 Adriana [Makes exaggerated face] Woah-oh-oh!
- 30 Jared Not a bad idea.
- 31 Eduardo Yeah, so—
- 32 Adriana And then we just have to add all of the Google ones. ← *Strategy 2+: Compare total website traffic*

Jared, Luca, and Eduardo go on to discuss the attributes of Google that make it more appealing than bing (e.g., Google Earth, images, its popularity in schools), conceding only that bing is newer, which may account for momentary spikes in its distribution. Ms. Parker then interrupts the class to determine whether groups need additional time, which she provides. As they reconvene, Luca suggests that they move on to a different representation when Adriana interrupts him to say, “Wait, wait, wait. Let me just count this.” What appears at first like Adriana applying Luca’s strategy actually becomes a new instance of Strategy Talk.

67	Luca	I would say bing.	
68	Adriana	Sh, sh. I’m almost done. Okay, so 1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16. Okay, there’s 16 times where bing has got more than Google.	← Strategy 3: Compare total number of days a site has more traffic than the other
69	Luca	And then how many times—	
70	Adriana	Wait. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14.	
71	Luca	bing!	
72	Jared	Wait, I’m scoring it. Look, I’m making it like a game. It’s whoever scores!	
73	Adriana	3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16. bing has more— [overlapping you hear Jared counting]	
74	Luca	Days—	
75	Adriana	—days with more people coming in. [Jared continues to count quietly]	

Adriana’s strategy (Lines 68, 70) is to count how many of the 30 days a website has more traffic than the other. This is a different strategy than the ones that previously emerged. Adriana’s strategy remains a comparative analysis (i.e., bing has more traffic on 16 of the 30 days, which is two more days than Google’s 14 days) but does not necessarily mean bing “wins,” as Luca suggests it does (Line 71). At this point, however, that is not apparent to the group. Instead, Adriana’s suggestion sparks a new but related strategy from Jared (Lines 73, 75), which becomes apparent shortly.

After a whole-class discussion clarifying what information is available in each representation, Ms. Parker asks groups to discuss how they are going to compare the websites. Eduardo begins by suggesting that they subtract the totals they found to see how much more traffic one website had over the other. Jared interrupts, saying, “I’m comparing right now.” Adriana suggests Jared use a calculator when Eduardo jokes he thinks Jared is avoiding the calculator to test his math. Jared replies emphatically, “I’m not doing math. I’m not doing math. I’m doing the scoring game!” Jared’s “game” is actually a new strategy built on Adriana’s previous one.

Using Adriana’s strategy of determining which website has more days of higher traffic, Jared narrows his focus on what he later calls “winning streaks”—intervals of time when one website’s values are comparatively higher (akin to a histogram)

93	Jared	I know, I know, I know. We should choose Google and I know what week to choose: the second week of the month!	← <i>Strategy 4: Determine which website has the greater “winning streak” in 30 days</i>
94	Luca	We don’t know what week it will show up.	
95	Jared	Cause look, look, look. That week, that whole week, Google is winning. That whole second week it’s winning, and more.	
96	Adriana	No, but it lost by 2.	
97	Jared	It lost by 2 because it [bing] caught up over here [gestures further down the table]. But look, but look right here. It’s winning that whole time [gestures to Week 2 in the table].	

than the other (see Figure 4). Days 6 through 16 favor Google (Lines 93, 95)—days Jared interprets as “Week 2”—although the overall comparison shows Bing winning more days overall (Lines 1, 97).

In Figure 4 Jared has marked dark circles next to winning days, small boxes next to losing days, and notes in the left margin “# of the highest hits in a row = 6” for Bing and “# of highest hits in a row = 10” for Google.¹ Also, in the margins Jared provides a comprehensive list of the strategies his group has discussed during the experiment, even commenting on the fourth day “I have it all on this one page!” as he holds this up to the teacher. Although Jared overlooks the constraints of the task (i.e., Line 94 when Luca reminds him they cannot deliberately pick the week), his winning streak strategy is a potentially viable new way to interpret the data in favor of Google. Jared’s strategy demonstrates how strategy talk leading to one outcome was often what inspired more strategy talk seeking alternative outcomes. From the initial visual comparisons of maximum and minimum to Luca’s focus on total traffic and Adriana’s focus on daily comparisons, Jared’s strategy became an intellectual descendent of them all.

Unlike Guided, in which students were given various statistical measures to characterize the difference in data distributions, Open offered no prescription for comparison. This group generated at least four different mathematically viable and related strategies in just over 15 min. The flexibility in Open offered opportunities for students to express their agency in working with the math as well as with one another. Analytically speaking, strategy talk and the opportunities for agency it illustrates were taken as possibly accounting for students’ more positive response to Open. The richer mathematical character of strategy talk in Open was taken as reflecting a noteworthy difference in student engagement from flexibility in working with the task to agency in collaboratively defining, addressing, refining, and resolving their thinking.

¹ Although Jared notes the winning streak as 10 days from Days 6 to 16, the actual count is 11 days.

FIGURE 4 Jared's classwork revealing a "winning streak" of 11 days where Google traffic exceeded bing traffic; in margins, his comprehensive list of strategies discussed by the group.

Student questioning about data is higher in the Open approach. Recall that the code Questioning refers to instances when students questioned one another and is an index of productive engagement because it shows how students are *holding one another accountable to the group's ideas and to disciplinary norms*. Table 7 reports the average number of instances of Questioning and the percentage of instances subcoded as specifically referencing data or not. Table 7 shows that although students in Guided were more frequently asking questions of their peers, the questions were less often based in data. The video data suggest instead that students in Guided were asking questions to orient logistically—what

TABLE 7
Average Number of Instances Coded Questioning by Treatment With Percentage Subcoded Using Data

<i>Treatment</i>	<i>Average Instances of Questioning</i>	<i>Percentage Subcoded Using Data</i>
Guided	97	16
Open	41	44

Note. Averages reported from five lessons and three cameras per treatment.

were they supposed to be doing, what were they supposed to be filling out, and so on. In contrast, nearly half of the questions posed by students in Open made direct reference to the data. We argue that this suggests questioning was a more mathematically rich practice in Open.

To illustrate this in Open, consider a different group of students looking at all of the representations to determine a strategy to pursue. One student, Anne, begins by determining the total number of hits for each website based on the table of values. Dominic, the student who poses the questions in this excerpt, asks Marco about his (rather illogical) strategy to sum the highest values shown on each graph while also asking Anne about the strategy she is pursuing through hand calculations.

- 1 Marco Let's take the highest bar. That's what I did.
- 2 Dominic Wait, from which graph?
- 3 Marco From every graph.
- 4 Dominic Anne, what are you doing? Are you adding these together?
- 5 Anne No.
- 6 Dominic How are you adding them?
- 7 Anne This thirty four thousand and thirty four thousand is sixty eight thousand, plus twenty seven thousand is ninety five thousand. Get it?
- 8 Dominic Oh! I get it, okay. Wait, why don't you just use a calculator?

Dominic's first question clarifies Marco's strategy (Line 2), and his next questions (Lines 4 and 6) work to clarify Anne's strategy and the specific way she is adding the column of numbers. For Open, keeping track of the strategies as they asynchronously developed was unlike the Guided approach, in which each strategy was introduced, formalized, and practiced synchronously. Once Dominic understands Anne's approach ("Oh! I get it, okay."), he suggests a way for her to accomplish her strategy more efficiently, which he also poses as a question ("Wait, why don't you just use a calculator?"). Though most of these questions are without specific reference to the data (except Lines 2 and 4), they demonstrate how asking questions serves to clarify and elaborate emerging mathematical strategies.

In this next excerpt, taken a little later in the group's conversation, Dominic thinks Anne has finished comparing the websites' total traffic values.

- 30   Dominic   I think you got it. How did you get it? Where did you get it?
 31   Anne   It's one million three hundred sixteen thousand. I'm still doing bing, though.

As Anne says in Line 31, she hasn't yet "got it" because the total for bing is unknown. Previously, Anne asked Natalie to calculate bing's total while she completed Google's total, but Natalie is not yet done. As she awaits Natalie's calculations, she and Dominic look at the values left in the table and each makes a conjecture about which website will "beat" which one.

- 35   Anne   [bing is] probably going to beat [Google] before it gets to the bottom.
 36   Dominic   Well, not in this area [gestures toward bottom of table]. Google is beating bing so it's playing catch up.
 37   Anne   No, it's not.
 38   Dominic   Uh huh. Three hundred sixty three thousand. This only has—
 39   Anne   Yeah, but see? Forty five thousand is less than fifty seven thousand.
 40   Dominic   Does this tell you this one will be better?
 41   Anne   Yes, but eighteen thousand is less than seventy thousand.
 42   Dominic   Oh, yeah. Okay, whatever.

By disagreeing with Dominic (Line 37), Anne challenges his conjecture. He responds by questioning her logic (Line 40), and she holds her ground, pointing out that the next value in the Google table is less than the next one in bing's table (Line 41). Just after Dominic concedes somewhat dismissively ("Okay, whatever.") Marco suggests a different strategy. He suggests they do a daily comparison for each of the 30 days to see which website wins the most days (what Adriana did in the previous group example). Initially, Dominic thinks this is a more efficient strategy, but soon he begins questioning it, realizing that it offers a relative but not absolute comparison.

- 47   Marco   Okay, I put "G" for [unintelligible] but at least I'll get the number.
 48   Dominic   But you still have to add them, because you don't know how much, you don't know how much you're beating with that.
 49   Marco   Yeah, duh.
 50   Dominic   That won't work, Marco.
 . . .
 55   Marco   Okay, bing won by two.
 56   Natalie   By two what? Thousand?
 57   Anne   Ugh, Marco! No, they didn't!
 58   Dominic   Marco, your strategy will not work, I'm telling you right now. Wait, how are you adding them so quickly?

Dominic contends that Marco's strategy still requires a running total because "you don't know how much you're beating with that [strategy]" (Line 48); that is, a relative daily comparison suggesting that Google is best does not ensure that the

absolute total will indicate the same. Nonetheless, Marco completes the strategy and announces that Bing has greater website traffic 2 days more than Google in Line 55 (i.e., Bing's traffic is higher 16 days to Google's 14 days, out of 30 days). Natalie asks a clarifying question (Line 56), thinking that Marco may have found Bing's total before she did. Finally, Dominic says explicitly that Marco's "strategy will not work" before asking a new question that draws attention back to Anne's strategy of comparing totals (Line 58).

In this example we see students asking a range of questions grounded in data. Some questions work to clarify what is being discussed (e.g., Lines 2, 30); others clarify strategies (e.g., Lines 6, 31), offer alternatives (e.g., Line 8), or even challenge the validity of strategy results (e.g., Line 56). This practice of asking questions grounded in data, as opposed to questions regarding the logistics of learning, distinguished Open from Guided. We argue that the mathematically oriented nature of questioning in Open not only supports the prior finding regarding strategy talk but also suggests a complementary accounting of why students responded more positively to Open. These results bring us to the final point of comparison: making meaning of mathematical work.

Meaning making is higher in the Open approach. Recall that the code Meaning Making refers to instances when students draw on prior knowledge to make sense of mathematical context, their strategies, and the answers that their strategies produce. This reflects disciplinary engagement because it shows that the students are oriented and committed to the meaning of the mathematics and not just producing an answer or finishing the assignment. Table 8 shows the average number of instances the Meaning Making code was ascribed by treatment. By definition, instances of Meaning Making involved explicit reference to the data (which are then connected to prior knowledge), and thus the percentage of instances subcoded Using Data was 100% regardless of treatment.

Students in Open made connections between the task and their prior knowledge or experiences 5 times as often as students in Guided. Opportunities to make meaning of the problem by connecting activity to personal experience or knowledge may further account for why students responded more positively to Open.

TABLE 8
Average Number of Instances Coded Meaning Making by Treatment

<i>Treatment</i>	<i>Average Instances of Meaning Making</i>
Guided	6
Open	31

Note. Averages reported from five lessons and three cameras per treatment.

FIGURE 5 Image of Anton’s class work showing day-to-day differences in Google traffic.

To illustrate, we present another group considering a strategy together: that a more consistent distribution of values (Google in this case) indicates the better choice. When Ms. Parker approaches the group and asks what they are working on, Anton remarks “consistency.” She then presses him to explain how he’s “calculating that.” Anton shows her his work (see Figure 5) calculating the day-to-day differences in traffic for each website.

Anton’s intuitive approach to examining variability is a mathematical concept (standard deviation) that we had not explicitly addressed in our lesson plans. In this group, Rachel, who is working with Anton on his strategy, explains that Bing has greater daily differences than Google and is therefore less consistent. When Ms. Parker pulls Clarice into the conversation, Clarice responds with support from what she knows about the websites beyond the data given. Clarice draws on knowledge beyond the data to make sense of the outcome: Google is better known than Bing (Line 6), which explains why Rachel (and Anton before her) found Google’s distribution to be more consistent and Bing’s more erratic (Line 2). Ms. Parker recognizes Clarice’s meaning making as part of her “intellectual processing” but presses her and the group to be explicit about how the data might be considered in relation to that sense making (Line 7).

- 1 Parker So you found a cluster here [gestures at Anton's work], I'm not sure what these are. Oh, you're adding and subtracting the difference! So you're finding the daily differences.
- 2 Rachel Yeah, so I did all the differences together and compared the two differences and Bing had a lot more differences—by a lot. And Google—
- 3 Parker So it came back to consistency.
- 4 Rachel Yeah, so Google is a lot more consistent.
- 5 Parker [To Clarice] What are you sharing with your group?
- 6 Clarice I think people know what Google is so people know what to expect from Google and are kind of ushered into it. [Unintelligible] because I don't think everyone has figured [Bing] out and it's kind of slowed down.
- 7 Parker Okay, so what strategy are you guys using? That's your intellectual processing, but using this set of data, how are you going to convey that?

Later in their deliberations, Anton reiterates his math work by showing that Google drops and rises only incrementally (he uses the word *increments*) over time. Clarice then responds by challenging his analytic approach. Drawing on prior knowledge to give a contextual explanation for what Anton is describing, she uses that context to extrapolate and argue that the strategy will not hold over time. In short, she argues that eventually Bing will be as well known as Google and therefore become more consistent in its distribution.

- 9 Anton So Google has been dropping by increments. This is Google and this is Bing.
- 10 Clarice My prediction is that it's going to flatten out. I don't really know [unintelligible].
- 11 Parker With the data you have, how can you show us? Remember you are making a decision based on the data that you have—we're going to look at the most. What is that? The most hits in a week? What do you think when they say "the most"?

When Clarice suggests that the graph will flatten out (Line 10), she is arguing that Bing will cease to be as erratic as it is currently depicted. She is basing this "prediction" on what she knows of the websites now: Bing is relatively new and Google is better known. Ms. Parker presses for the group to stay true to the data before them (Line 11: "With the data you have, how can you show us?") despite Clarice's assertions. In subsequent group and whole-class discussions, Clarice continues arguing against "consistency" and, later, range as a way to decide between the websites based on her sense that the popularity of the sites should matter.

In this example we see how making meaning was part of the group's engagement in interpreting data and reasoning about strategies. Given the contrast in meaning making between teaching approaches, we argue that supporting students' sense making contributed to students responding positively to Open. For Clarice, it makes no sense to limit the meaning of what they are doing to 30 days

TABLE 9
Average Number of Off-Task Instances and Total Length of Time Off Task by Treatment

Variable	Guided	Open
Average number of off-task instances	18	14
Average total length of time off task (hours:minutes:seconds)	00:11:13	00:13:49

Note. Averages reported from five lessons and three cameras per treatment.

of data and ignore what seem like obvious contextual factors. By allowing students opportunities to develop strategies, and for the class to make sense of them collaboratively, Open made room for the use of prior knowledge and assertions of agency to challenge more literal (and perhaps limited) mathematical arguments.

Off-task behavior does not differ between approaches. We examined whether instances of students being off task were notably greater in the Open treatment, in which self-guided inquiry might be seen as straining students’ abilities to stay engaged. As Table 9 shows, however, neither the total number of off-task instances for groups nor the average total time off task was significantly different, despite a comparable amount of time spent in small groups overall (see Table 3).

In Table 9, the difference in average total length of time off task is just under 3 min out of approximately 300 min (5 hr) of lessons, which represents about a 1% difference. The near parity reflected in this quantitative outcome is somewhat misleading when one considers the amount of adult intervention in each treatment, as we do next.

Adult intervention is more content oriented for the Open approach. Figures 6 and 7 show the results of coding the intervention of adults in videos of the Guided and Open treatments, respectively. Each adult interaction was subcoded to reflect the adult’s focus as related to content, behavior management, or both.

Figures 6 and 7 illustrate adults intervening around behavior management *twice as often* in Guided than Open (18% for Open vs. 37% for Guided). This suggests that students in Guided required greater adult attention to remain on task than students in Open. In contrast, students in Open were more often addressed regarding matters of content rather than behavior.

EVIDENCE FOR THE CLAIMS

Our argument is that particular pedagogical features afford students opportunities to develop proficiency in mathematics while also cultivating disciplinary interest.

FIGURE 6 Percentage of instances in Guided coded Adult Intervention and subcoded for attention to content, behavior management, or both.

FIGURE 7 Percentage of instances in Open coded Adult Intervention and subcoded Content, (behavior) Management, or both.

We collected data to investigate three claims: (a) that different designs for learning the same mathematics can have different affordances for disciplinary engagement; (b) that the affordances are realized during instruction as differences in mathematical practices such as questioning, meaning making, and strategy talk; and (c) that overall differences in students' affective response to the designs are more

plausibly linked to the differences in mathematical practices than to competing hypotheses (e.g., off-task behavior, teacher support).

Our first claim is that the Open instructional design afforded students opportunities to cultivate disciplinary interest. We speculate that the Open approach had several key elements that uniquely created space for students to take up important intellectual challenges and act with authority to solve those challenges in creative and personally meaningful ways. These elements were problem articulation, strategy invention, and exploration. Found only in the Open approach, these elements likely led to greater problematizing and strategizing of solution pathways with peers (principles of productive disciplinary engagement), which may have led students to feel greater ownership over the purpose and nature of their mathematical engagement. In contrast, students in Guided were given the problem and, through structured inquiry, were led through a variety of pathways to possible solutions. Despite having an equal amount of time in small groups as the Open approach, we suspect the Guided design shifted the balance in inquiry away from one's own reasoning to that of understanding the formalized processes presented. This latter issue of formalization, common to Open and Guided but appearing at different stages of inquiry, draws attention to the role of instructional sequencing.

The two learning designs had several common elements, including *best inference*, *formalization*, and *practice in procedural fluency*. What differed was the sequencing of those elements, which framed the activities differently for students. The designed sequence of Open had students reinvent canonical mathematical procedures before asking them to map their methods onto formal disciplinary procedures. In contrast, students in Guided were given formal definitions and procedures near the onset of inquiry. We speculate that transitioning from a student-driven understanding to a formal disciplinary conception of mode carried different weight and meaning based on when it occurred in the instructional sequence. Delaying formalization in Open meant that students could engage in mathematics in a way that Guided students did not. Thus, although both instructional sequences led to similar proficiency in written assessments, Open seemed to cultivate students' disciplinary interest better. Tracing students' mathematical and affective responses to features of the design is important because it reaffirms that pedagogical approaches structure more than cognitive processes or proficiency; they also structure opportunities for students to enjoy learning.

Our second claim is that the affordances of the Open design were realized during instruction as differences in students' engagement with mathematical practices (e.g., questioning, meaning making, and discussing strategies). Using Engle and Conant's (2002) principles of productive disciplinary engagement as potentially having both positive cognitive and affective outcomes, we explored students' engagement as they discussed strategies, asked questions, and made meaning of mathematical ideas in terms of concrete experiences. Our quantitative comparisons of peer interactions showed that students in the Open approach engaged in

these practices more often and in richer ways than did the students in the Guided approach. Students in the Open approach also reported liking learning at a statistically higher frequency than those in Guided. Therefore, although these three practices are among those valued for deepening students' conceptual and practical understandings of mathematics (Common Core Standards Initiative, 2011), we show that they may also have affective value. Our claim, then, is that disciplinary engagement and interest are interrelated with a sense of liking learning.

Our third claim is more of a conjecture about how and why liking learning is causally associated with engaging in disciplinary practices. Proving this is beyond the scope of this article and our design, but we see our evidence as consistent with such logic and as refuting other commonplace possibilities. For instance, we show that students in Open are no less on task than those in Guided; that is, isolating goofing off as the source of enjoyment is not a viable argument. Likewise, the National Research Council recommends using tasks that students find enjoyable (Kilpatrick et al., 2001) in order to promote student engagement, which we certainly did by capitalizing on a popular cultural debate between Bing and Google. In our case, however, students in Open and Guided had the same task and spent similar amounts of time addressing the task. Therefore, task alone cannot account for the difference in affective outcomes. We therefore find it probable that richer engagement in mathematical practices was the source of enjoyment for students in the Open design. If, for example, we understand the creation of one's own problem-solving pathway as a creative act (Schoenfeld, 1992), it may follow that creating strategies, discussing one's creations with peers, aligning that creation to what one knows, and realizing that one's own creation has a formal mathematical counterpart are *emotional* acts inasmuch as they evidence deep engagement in mathematical practices. Take Jared, the student who summarized his group's many strategies on one paper (see Figure 4) and who jumped out of his seat waving the mathematical worksheet in the air shouting "I have it all here!" as an example. He and other cases in our qualitative data remind us that there is a great amount of emotion in meaningful learning. Thus, designing learning opportunities to support deep engagement in mathematical practices involves designing for students to like learning.

LIMITATIONS

One limit of our teaching experiment is that it only explored one kind of instructional task. If the topic and task were changed to something of social, cultural, community, or political relevance (e.g., Enyedy, Danish, & Fields, 2011; Enyedy & Mukhopadhyay, 2007; Gutstein, 2003), would the achievement and affective outcomes seen here be enhanced further? We can imagine that the answer is likely yes, because, as we have attempted to show here, the pedagogy and context is not just a matter of motivation—engagement and affect are tied to meaningful

mathematical practices. A more relevant context for inquiry might interact with our instructional design and afford new and different resources for articulating the problem, developing strategies and discussing them with peers, aligning those ideas with what one knows (presumably more than one might of search engines), and connecting those practices with formal mathematics.

A second limit of our instructional design is the assessment. We developed an assessment of data and statistics knowledge based on existing standardized assessments that center on procedural or conceptual understanding but not mathematical practices. If, for example, we had had an assessment that better reflected proficiencies in mathematical practices (e.g., questioning, use of multiple strategies), we may have seen a difference in the written assessment outcomes by design condition. The development and use of formative and summative assessments to evaluate students' proficiencies in mathematical practices is an important consideration for future studies.

A third limitation of the work is the role of teaching assistants during the study. Although the teacher was well informed on the research and design of the instructional contrast, we allowed the teacher to inform her instructional team on the design. Although field notes reflect that the teacher told the assistants of the design, we did not systematically characterize those exchanges or their possible implications for the design's implementation.

Fourth and finally, we do not yet know enough about how mathematical practices might relate to equity. In a sense, our work suggests that mathematical practices may be a promising way to cultivate students' positive domain identities as competent learners (Nasir & Hand, 2008). Hand (2003, 2012) conceptualized equitable instruction in mathematics in terms of students "taking up space" to participate broadly and feel comfortable in sharing ideas, questioning others, and, most important, working in ways that align with who they see themselves becoming. Our design principles in Open align with that articulation and others focusing on the importance of developing students' mathematical dispositions as part of the great project of equity (e.g., Gresalfi & Cobb, 2006). However, our study was not designed to deeply investigate equity considerations; this remains to be done in future work.

Overall, although the data here are not sufficient to reach strong instructional recommendations, teachers may be inspired to see mathematical practices as a way to nurture and sustain students' enjoyment of mathematics.

CONCLUSIONS AND IMPLICATIONS

The *Common Core Standards of Mathematics* (Common Core Standards Initiative, 2011) introduce eight practice standards that encourage teachers to think

deeply about developing proficiency in mathematical practices and not just content mastery. A potentially important conclusion of our study is that mathematical practices may mediate the relationship between varied designs for learning (such as Guided vs. Open) and the important outcome of cultivating and sustaining students' disciplinary interest.

Our focus on the mediating role of mathematical practices was investigated by a matched instructional design for elementary school students to develop an understanding of central tendency, which appeared in the fifth-grade curriculum standards for data and statistics. At the center of this teaching experiment was the design of teacher-led (Guided) and student-driven (Open) conditions, building on the established distinction and strengths of each in prior research. By requiring that each teaching approach occur in the same amount of time (five lessons), including time spent in small-group and whole-class discussions, we controlled for time on task as an explanation for different outcomes. Both approaches resulted in learning gains, such that students performed significantly better on posttests. Where the results differed was in students' written responses to how well they liked the lessons: Students in Open were significantly more positive than students in Guided.

As reported in our prior research (Sengupta-Irving et al., 2013), the teacher in this study came to prefer the Open approach for two reasons: First, Open was just as efficient as Guided in terms of time and resources and yielded similar assessment outcomes; and second, the teacher was moved by the experiences of a mathematically anxious student who developed a strategy she later shared in the class. For the teacher, this student embodied how the Open design could afford disaffected students new opportunities to engage mathematically with their peers.

Our study goes beyond this teacher's opinion to investigate some of the empirical consequences and mediating practices afforded by the different designs. Using video analysis, we were able to empirically trace the association between students' positive responses to learning and richer engagement in mathematical practices to specific features of the student-driven (Open) design. Our findings suggest that student engagement in developing strategies, questioning, and making meaning of mathematical work can mediate enjoyment and feeling positive about learning.

For designers and researchers, it is important to recognize that our design grew from the principles that Engle and Conant (2002) initially articulated. Our study explores the extent to which learning opportunities that afford productive disciplinary engagement may also foster *positive affective* engagement. This is an important confirmation of Engle and Conant's disciplinary engagement framework, which was based on a very different task. In their task, students engaged in an academic debate about the classification of orcas; at times this debate became heated and personal. Our findings help establish that Engle and Conant's framing of engagement as mediated by practices is not limited to scientific argumentation but may lead to students' enjoyment of a discipline in a wide variety of

tasks. Furthermore, our focus on specific practices helps explain why Engle and Conant's principles apply to a wide variety of tasks and contexts, as it links affect to the process rather than the content.

For teachers, the implication of this study is that a focus on mathematical practices may be important both to achieving the new definition of proficiency in curricular standards and assessments and also to sustaining students' interest in mathematics. We see this work as taking an important step, especially at a time of significant mathematics reforms, to support teachers in seeing the importance of nurturing students' interests alongside the practices and proficiencies of mathematical expertise. Our instructional designs demonstrate that supporting cognitive and affective gains in mathematics learning are not separate endeavors with separate time allocations. For example, if a teacher presumes that he or she must make extra room in the pacing guide for fun (as though apart from math learning), then fun will always give way to learning. We argue that learning can be fun when, as we speculate here, students are afforded opportunities to engage richly in mathematical practices.

We do not yet know how connecting cognitive and affective outcomes in this way might increase and diversify the pipeline of students who pursue mathematics over years and not simply during mathematics activities. We know that some portion of students stop taking mathematics courses or pursuing a science, technology, engineering, or mathematics career not because they are not proficient but because they do not actively *like* math. We speculate that the Open design principles could be building blocks for designing ways of broadening and diversifying the pipeline of students pursuing mathematics. And yet there is work to be done to unify theories of learning and emotion with the development of instructional design principles that cultivate the practices and proficiencies of students inasmuch as they elevate the interests and passions of all students for the discipline.

REFERENCES

- Ainley, M. (2012). Students' interest and engagement in classroom activities. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 283–302). New York, NY: Springer.
- Ainley, M., & Ainley, J. (2011). Student engagement with science in early adolescence: The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychology*, 36(1), 4–12.
- Ainley, M., Corrigan, M., & Richardson, N. (2005). Students, tasks and emotions: Identifying the contribution of emotions to students' reading of popular culture and popular science texts. *Learning and Instruction*, 15, 433–447.
- Ainley, M., & Hidi, S. (2014). Interest and enjoyment. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 205–227). New York, NY: Routledge.
- Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, 94, 545–561.

- Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *The Elementary School Journal*, 93(4), 373–397.
- Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, 12, 307–359.
- Beaton, A. E., Mullis, I. V., Martin, M. O., Gonzales, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College, TIMSS International Study Center.
- Boaler, J. (2002). Learning from teaching: Exploring the relationship between reform curriculum and equity. *Journal for Research in Mathematics Education*, 33, 239–258.
- Boekaerts, M. (1999). Motivated learning: Studying student-situation transactional units. *European Journal of Psychology in Education*, 14, 41–55.
- Bright, G. W., & Friel, S. N. (1998). Graphical representations: Helping students interpret data. In S. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessments in grades K-12* (pp. 63–88). Mahwah, NJ: Erlbaum.
- Burris, C. C., Heubert, J. P., & Levin, H. M. (2006). Accelerating mathematics achievement using heterogeneous grouping. *American Educational Research Journal*, 43(1), 137–154.
- California Department of Education. (1999). *Mathematics framework for California public schools, kindergarten thru grade 12*. Sacramento, CA: Author.
- Case, R. W. (2005). The Dutch revolution in secondary school mathematics. *Mathematics Teacher*, 98(6), 374–384.
- Civil, M. (2006). Building on community knowledge: An avenue to equity in mathematics education. In N. Nasir & P. Cobb (Eds.), *Improving access to mathematics: Diversity and equity in the classroom* (pp. 105–117). New York, NY: Teachers College Press.
- Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning*, 1(1), 5–43.
- Cobb, P., Gravemeijer, K., Yackel, E., McClain, K., & Whitenack, J. (1997). Mathematizing and symbolizing: The emergence of chains of signification in one first-grade classroom. In D. I. Kirschner & J. A. Whitson (Eds.), *Situated cognition: Social, semiotic, and psychological perspectives* (pp. 151–233). Mahwah, NJ: Erlbaum.
- Cobb, P., & McClain, K. (2001). Supporting students' ability to reason about data. *Educational Studies in Mathematics*, 45, 103–129.
- Cobb, P., & McClain, K. (2004). Principles of instructional design for supporting the development of students' statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 375–396). Dordrecht, The Netherlands: Kluwer Academic.
- Cohen, E. G., & Lotan, R. A. (1997). *Working for equity in heterogeneous classrooms: Sociological theory in practice*. New York, NY: Teachers College Press.
- Common Core Standards Initiative. (2011). *Standards for mathematical practice*. Retrieved from <http://www.corestandards.org/read-the-standards/>
- Dewey, J. (1913). *Interest and effort in education*. New York, NY: Houghton Mifflin.
- diSessa, A. A., Hammer, C., Sherin, B., & Kolpakowski, T. (1991). Inventing graphing: Meta-representational expertise in children. *Journal of Mathematical Behavior*, 10, 117–160.
- Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction*, 20, 399–483.
- Enyedy, N. (2003). Knowledge construction and collective practice: At the intersection of learning, talk, and social configurations in a computer-mediated mathematics classroom. *Journal of the Learning Sciences*, 12, 361–407.
- Enyedy, N., Danish, J. A., & Fields, D. (2011). Negotiating the “relevant” in culturally relevant mathematics. *Canadian Journal for Science, Mathematics, and Technology Education*, 11(3), 273–291.

- Enyedy, N., & Mukhopadhyay, S. (2007). They don't show nothing I didn't know: Emergent tensions between culturally relevant pedagogy and mathematics pedagogy. *Journal of the Learning Sciences*, 16, 139–174.
- Eppstein, S. (1990). Cognitive-experiential self-theory: Implications for developmental psychology. In M. Gunnary & L. A. Sroufe (Eds.), *Minnesota symposia on child psychology: Self processes and development* (pp. 79–123). Mahwah, NJ: Erlbaum.
- Esmonde, I. (2009). Ideas and identities: Supporting equity in cooperative mathematics learning. *Review of Educational Research*, 79, 1008–1043.
- Eysink, T. H. S., de Jong, T., Bertold, K., Kolloffel, B., Operfmann, M., & Wouters, P. (2009). Learner performance in multimedia learning arrangements: An analysis across instructional approaches. *American Educational Research Journal*, 46, 1107–1149.
- Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two "male-typed" domains. *Developmental Psychology*, 38, 519–534.
- Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, 20, 507–537.
- Freudenthal, H. (1979). New math or new education? *Prospects: Quarterly Review of Education*, 9(3), 321–331.
- Ge, X., & Land, S. M. (2003). Scaffolding students' problem-solving processes in an ill-structured task using question prompts and peer interactions. *Educational Technology Research and Development*, 51(1), 21–38.
- Gravemeijer, K., & Doorman, M. (1999). Context problems in Realistic Mathematics Education: A calculus course as an example. *Educational Studies in Mathematics*, 39, 111–129.
- Greeno, J. C. (2006). Authoritative, accountable positioning and connected, general knowing: Progressive themes in understanding transfer. *Journal of the Learning Sciences*, 15, 537–547.
- Gresalfi, M. S. (2009). Taking up opportunities to learn: Constructing dispositions in mathematics classrooms. *Journal of the Learning Sciences*, 18, 327–369.
- Gresalfi, M. S., & Cobb, P. (2006). Cultivating students' discipline-specific dispositions as a critical goal for pedagogy and equity. *Pedagogies*, 1(1), 49–57.
- Gutstein, E. (2003). Teaching and learning mathematics for social justice in an urban, Latino school. *Journal for Research in Mathematics Education*, 34, 37–73.
- Gutstein, E. (2012). Connecting community, critical, and classical knowledge in teaching mathematics for social justice. In S. Mukhopadhyay & W. M. Roth (Eds.), *Alternative forms of knowing (in) mathematics* (pp. 300–311). Rotterdam, The Netherlands: Sense.
- Hadi, S. (2012). *Mathematics education reform movement in Indonesia*. Retrieved from http://www.icme12.org/upload/submission/1897_F.pdf
- Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, 27(3), 337–364.
- Hand, V. (2003). *Reframing participation: Meaningful mathematical activity in diverse classrooms* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.
- Hand, V. (2012). Seeing power and culture in mathematics learning: Teacher noticing for equitable mathematics instruction. *Educational Studies in Mathematics*, 80, 233–247.
- Hidi, S. (2006). Interest: A unique motivational variable. *Educational Research Review*, 1(2), 69–82.
- Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research*, 70(2), 151–179.
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111–127.
- Hidi, S., Renninger, K. A., & Krapp, A. (2004). Interest, a motivational variable that combines affective and cognitive functioning. In D. Dai & R. Sternberg (Eds.), *Motivation, emotion, and cognition:*

- Integrative perspectives on intellectual functioning and development* (pp. 89–115). Mahwah, NJ: Erlbaum.
- Hiebert, J. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.
- Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., . . . Wearne, D. (1996). Problem solving as a basis for reform curriculum and instruction. The case of mathematics. *Educational Researcher*, 25(4), 12–21.
- Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), 99–107.
- Horn, I. S. (2006). Lessons learned from detracked mathematics departments. *Theory Into Practice*, 45, 72–81.
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, 73, 509–527.
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26, 379–424.
- Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science*, 38, 523–550.
- Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academies Press.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41, 75–86.
- Klahr, D. (2009). "To every thing there is a season, and a time to every purpose under the heavens": What about direct instruction? In S. Tobias & T. M. Duffy (Eds.), *Constructivist theory applied to instruction: Success or failure?* (pp. 291–310). New York, NY: Taylor & Francis.
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction effects of direct instruction and discovery learning. *Psychological Science*, 15, 661–667.
- Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education*, 32, 448–470.
- Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, 3(1). Retrieved from <http://www.amstat.org/publications/jse/v3n1/konold.html>
- Konold, C., & Higgins, T. L. (2003). Reasoning about data. In J. Kilpatrick, G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193–215). Reston, VA: NCTM.
- Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A. D., Wing, R., & Mayr, S. (2002). *Students' use of modal clumps to summarize data*. Retrieved from http://iase-web.org/documents/papers/icots6/8b2_kono.pdf
- Krapp, A. (1999). Interest, motivation and learning: An educational-psychological perspective. *European Journal of Psychology in Education*, 14, 23–40.
- Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist*, 42(2), 109–113.
- Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal*, 27(1), 29–63.
- Langer-Osuna, J. M. (2011). How Brianna became bossy and Kofi came out smart: Understanding the trajectories of identity and engagement for two group leaders in a project-based mathematics classroom. *Canadian Journal of Science, Mathematics and Technology Education*, 11(3), 207–225.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1), 65–100.

- Lehrer, R., & Pritchard, C. (2002). Symbolizing space into being. In K. Gravemeijer, R. Lehrer, B. van Oers, & L. Verschaffel (Eds.), *Symbolization, modeling and tool use in mathematics education* (pp. 58–86). Dordrecht, The Netherlands: Kluwer.
- Lehrer, R., Strom, D., & Confrey, J. (2002). Grounding metaphors and inscriptional resonance: Children's emerging understanding of mathematical similarity. *Cognition and Instruction*, 20, 359–398.
- Lin, X., Hmelo, C., Kinzer, C. K., & Secules, T. J. (1999). Designing technology to support reflection. *Educational Technology Research and Development*, 47(3), 43–62.
- Linchevski, L., & Kutscher, B. (1998). Tell me with whom you're learning, and I'll tell you how much you've learned: Mixed-ability versus same-ability grouping in mathematics. *Journal for Research in Mathematics Education*, 29, 533–554.
- Meyer, D. K., & Turner, J. C. (2002). Discovering emotion in classroom motivation research. *Educational Psychologist*, 37(2), 107–114.
- Mullis, I. V., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., . . . Smith, T. A. (2000). *TIMSS 1999 international mathematics report*. Chestnut Hill, MA: TIMSS International Study Center.
- Nasir, N. S., & Hand, V. (2008). From the court to the classroom: Opportunities for engagement, learning and identity in basketball and classroom mathematics. *Journal of the Learning Sciences*, 17, 143–180.
- National Council of Teachers of Mathematics. (2000). *Principles for school mathematics*. Reston, VA: Author.
- Panksepp, J. (2003). At the interface of the affective, behavioral, and cognitive neurosciences: Decoding the emotional feelings of the brain. *Brain and Cognition*, 52(1), 4–14.
- Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education*. Columbus, OH: Merrill.
- Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, 13, 273–304.
- Renninger, K. A., & Hidi, S. (2011). Revisiting the conceptualization, measurement, and generation of interest. *Educational Psychologist*, 46(3), 168–184.
- Samarapungavan, A., Patrick, H., & Mantzicopolous, P. (2011). What kindergarten students learn in inquiry-based science classrooms. *Cognition and Instruction*, 29, 416–470.
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem-solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334–370). New York, NY: Macmillan.
- Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, 16, 475–523.
- 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, 129–184.
- Sengupta-Irving, T., Redman, E., & Enyedy, N. (2013). Re-storying practice: Using stories about students to advance mathematics education reform. *Teaching and Teacher Education*, 31, 1–12.
- Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, 27(2), 4–13.
- Sherin, M. G., & Drake, C. (2009). Curriculum strategy framework: Investigating patterns in teachers' use of a reform-based elementary mathematics curriculum. *Journal of Curriculum Studies*, 41, 467–500.
- Smith, J. P., III. (1996). Efficacy and teaching mathematics by telling: A challenge for reform. *Journal for Research in Mathematics Education*, 27, 387–402.
- Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science*, 19(1), 97–140.
- Suthers, D., & Hundhausen, C. (2003). An empirical study of the effects of representational guidance on collaborative learning. *Journal of the Learning Sciences*, 12, 183–219.

- Trautwein, U., Ludtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, 98, 788–806.
- Turner, S. A., Jr., & Silvia, P. J. (2006). Must interesting things be pleasant? A test of competing appraisal structures. *Emotion*, 6, 670–674.
- Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, 27, 458–477.
- Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge-building communities. *Journal of the Learning Sciences*, 18, 7–44.