THE JOURNAL OF THE LEARNING SCIENCES, 1–25, 2011

ISSN: 1050-8406 print / 1532-7809 online
DOI: 10.1080/10508406.2011.611775

A Comparison of Lecture-Based and Challenge-Based Learning in a Workplace Setting: Course Designs, Patterns of Interactivity, and Learning Outcomes

Timothy K. O'Mahony, Nancy J. Vye, John D. Bransford, and Elizabeth A. Sanders
College of Education
University of Washington

Reed Stevens
Learning Sciences
Northwestern University

Richard D. Stephens and Michael C. Richey
The Boeing Company

Kuen Y. Lin
Aeronautical Engineering
University of Washington

Moe K. Soleiman
The Boeing Company

We describe findings from a research partnership involving a global airline manufacturing company (The Boeing Company), and learning scientists and aeronautical engineers from the University of Washington. Our starting point for the partnership focused on an 8-hour introductory composites course that was designed

Correspondence should be addressed to Timothy K. O'Mahony, College of Education, University of Washington, 1100 Plaza Building, Suite 200, Seattle, WA 98103. E-mail: tko2@u.washington.edu

for company employees. In phase one, learning scientists observed the company's course development activities and the course as taught by company experts. In phase two, we collaboratively designed and implemented a quasi-experimental study comparing two approaches to teaching. One involved lectures with PowerPoint slides. The second, a "challenge-based" learning approach, combined a set of composites-relevant challenges with individual, small-group, and large-group collaborative inquiry. Comparisons between these methods showed greater interaction among participants in the challenge-based group. In addition, the challenge-based group performed significantly better on posttest items requiring integration and synthesis of concepts. Increased interactivity in the challenge course provided opportunities for participants to articulate connections among concepts and may have contributed to the challenge participants' better synthesis of learned concepts. This work highlighted the benefits for learning scientists of collaborating with industry partners to explore learning in workplace settings, as these settings provide illuminating contrasts to the structures of teaching, learning, and assessment found in schools.

Researchers in the learning sciences are increasingly exploring the nature of learning arrangements in a variety of settings (e.g., schools, homes, community centers) and seeking ways to enhance learning by bridging formal and informal learning experiences (Bell, Lewenstein, & Shouse, 2009; Bransford et al., 2006; Shutt, Phillips, Vye, Van Horne, & Bransford, 2010; Tzou & Bell, 2010). In this broader research landscape, workplace learning, which is the focus of the present research, is also receiving new notice (e.g., Eraut, 2004; Stevens, 2000). Many workplace settings include formal courses plus opportunities for informal learning among colleagues either in a course or after a course is completed or both.

According to Stokes (1997), one can take different approaches to studying workplace learning, for example by working (a) in "Edison's quadrant" and asking whether teaching and learning techniques that work in other settings (e.g., K–16) also work in workplace settings; or (b) in "Pasteur's quadrant," with its emphasis on use-inspired theory building. The research we describe fits Edison's quadrant, but important findings also came from focusing on Pasteur's quadrant and the wide range of new questions about teaching, learning, and assessment that emerged as the work proceeded.

The research involved a partnership between members of the University of Washington's LIFE (Learning in Informal and Formal Environments) Center and The Boeing Company (hereafter, Boeing), which offers more than 6 million hours of instruction (face-to-face, online, and blended courses) each year to more than 150,000 employees across 45 countries. We report on two phases of research that explore different instructional methods: (a) a pilot study of a course offered by the company; and (b) a subsequent comparison study of different instructional approaches to an Introduction to Composites, a topic of great importance to the company and all of its employees. In Phase 2, we conducted a quasi-experimental study comparing Boeing's lecture-based approach to teaching with a "challenge-based" design (Schwartz, Lin, Brophy, & Bransford, 2000). We briefly describe these two phases next.

PHASE 1 PILOT STUDY

A group of five learning scientists from the LIFE Center (www.life-slc.org) worked with subject matter experts (SMEs) in materials science and others involved in training at Boeing to observe the development and implementation of a new introductory course on composite materials. The introduction of composites is a critical new element to aircraft manufacturing, as technologies make this medium more suitable and more desirable than metals (because of weight, strength, durability, fuel efficiency, and much more) for building modern aircraft. Multiple offerings of this course were to be delivered in face-to-face sessions that each lasted about 8 hr. After observing the course development process, two learning scientists audited one of the course offerings to get first-hand experience with the teaching practices. In addition, a media group from Boeing video-captured the interactions during the course so that they could be shared with others later on.

To develop the new composites course, the instructors (two SMEs in materials science) divided the course content according to their particular knowledge strengths and engaged in careful analyses of what was to be taught. Course content was placed on PowerPoint slides, and the instructors, plus several other experts, vetted the importance and accuracy of each slide's content and resolved any disagreements. The course instructors later prepared printouts of the PowerPoint slides to distribute to students for note taking during the course and for reference afterward.

The two learning scientist auditors (and those who later watched the video recordings) were struck by the carefully crafted presentation of the content and by the motivation of the engineers to sit through the roughly 8 hr of presentations. However, they noticed the virtual absence of both student-to-student interaction and student-generated questions in the course. Overall, the pattern of interactions fit the Initiate, Respond, Evaluate pattern that researchers such as Mehan (1985) have explored.

Because the composites course was introducing students (i.e., incumbent engineers) to new ways of doing their familiar jobs—all of them had experience building parts of aluminum airplanes—it seemed plausible that many had questions and comments that would have been valuable for all students, and even the instructors, to hear and discuss. However, as noted, these kinds of questions rarely appeared given the lecture-driven format of the course. In light of this, members of Boeing and the LIFE Center asked whether it might be possible to present relevant content while also supporting student-to-student and student-to-instructor questions and comments. Simply asking people to work in collaborative groups seemed unwieldy—a more focused design was needed that invited interactivity but was efficient as well. We settled on a challenge-based learning (CBL) design called STAR Legacy (STAR stands for Software Technology for Action and Reflection). This design format had been used successfully with different types of learners, including undergraduate students in bioengineering (Martin et al., 2006; Roselli & Brophy, 2006). The CBL format has shown greater performance gains over a lecture format on test items that require a synthesis of information across a number of different information sources (Martin et al., 2006). However, the design has not been used in the workplace with incumbent engineers (much of the prior work with the Legacy cycle has been focused on formal settings, including schools, but not on industrial settings). Nor has prior research focused on the patterns of interactions that occur during CBL among engineers (or other learners). Boeing's willingness to allow video and audio recording provided an opportunity to study these interactions. Based on prior research, we speculated that challenge-based discussions could offer learners greater opportunity than lectures to connect sources of information that otherwise might remain disconnected. However, it was also possible that incumbent engineers are sufficiently expert, from the perspective of both their metacognition and their engineering knowledge, to make these connections and elaborations without a need for high levels of interactivity during the course.

PHASE 2 COMPARISON OF TEACHING METHODS

In Phase 2, we conducted a quasi-experimental study comparing the lecture approach to teaching with the challenge-based design. Given that Boeing needed high-quality learning plus efficiency, we selected CBL as our comparison to lectures. The latter consists of a set of problems or challenges, each with its own cycle of inquiry activities. Each cycle begins with a content-relevant challenge (e.g., "What are some major advantages and challenges of building airplanes out of composites materials rather than aluminum?") and is followed by a request for learners to generate their initial thoughts about the challenge, access to student-controlled audio and video resources (essentially mini-lectures) designed to deepen learners' initial thoughts, a chance for small-group discussions about the challenge and resources, and finally a large-group discussion that includes key ideas students have learned and further questions for the instructor. Because the incumbent engineers came to this study from different work areas (with regard to geography and domain specificity), prior knowledge was varied with regard to experience with composites. However, because the course was advertised as an introduction to composites, engineers who chose to take this course were novices for the most part in their use of composites for manufacturing airplanes.

The challenge-based approach had its roots in *anchored instruction*, in which learners focused on an initial problem situation (that was often presented on videodisc) and then were able to experience how new perspectives from peers and experts changed their thinking (Bransford, Sherwood, Hasselbring, Kinzer, & Williams, 1986; Cognition and Technology Group at Vanderbilt [CTGV], 1997; Michael, Klee, Bransford, & Warren, 1993). As video storage requirements shrank from videodiscs to the hard drives of laptops, and as software became easier to program, work on anchored instruction evolved into the STAR Legacy program (e.g., Schwartz et al., 2000).

A review by Williams (1992) explored similarities between anchored instruction and other approaches, such as case-based learning (e.g., Gragg, 1940), problem-based learning (e.g., Hmelo-Silver & Barrows, 2006), and design-based instruction (e.g., Kolodner et al., 2003). (The challenge-based STAR Legacy program was not available at the time of Williams's review.) Similar to anchored instruction, CBL aligns with these other learning approaches by virtue of its focus on learning in the context of problem solving (rather than on teaching followed by "application problems").

In contrast to many other approaches, CBL within the STAR Legacy framework represents a more condensed (in time) and structured (via technology) sequence of problems with related inquiry and collaborative activities (i.e., initial generation of problem solutions followed by access to resources and discussions; see Martin et al., 2007, for studies of these features). For these reasons, we deemed the Legacy framework a good fit for the Boeing context, where training needs to be time efficient, needs to be focused on specific concepts and skills that are prime for near-term use in the workplace, and at the same time promotes conceptual learning and learning for application.

Our research study posed two major questions about the two instructional groups:

1. How do they differ on curriculum-based learning outcomes?
2. How do they differ in their social interactions during the learning process?

To address these questions, we used a mixed-methods research design that (a) tested group differences on pretests and posttests and (b) examined discourse during (video-recorded) large- and small-group interactions.

METHODS

Participants

A total of 64 engineers initially signed up to attend one of three seemingly identical course offerings (announced as an 8-hr "Introduction to Composites" listed on the company website's professional development webpage[1]). Unbeknownst to the engineers who registered for the courses, the research team arranged for traditional learning (TL) instruction to be used in the first two course offerings and the newer CBL instruction in the third offering of the course. (We realize that two vs. one was not ideal, but this is one of the latent challenges that emerges when one is working in authentic industrial settings. Sometimes one has to make do with what one gets.) Once registered, engineers in their respective courses were invited to participate in the study. Those who had signed up for the third course (CBL) were informed that they would be participating in a new course format and were offered the option of switching to the traditional format if they desired. None chose to switch to one of the TL courses. Of those recruited, all aerospace company engineers consented to participate in the study ($n = 44$ across the two TL courses, and $n = 20$ in the CBL course). Although this research study design was quasi-experimental in nature (comparing convenience groups of engineers who received two different instructional approaches), we note that (a) the engineers were blind to which group they were initially registering for, (b) none switched courses after they were recruited for the study, and, (c) as shown in the Results section, there were no significant differences between the groups on the pretest.

The sample included both male ($n = 56$), and female ($n = 8$) engineers from various functional fields of expertise within the aerospace company (e.g., structures, payload). Participants ranged in age from 22 to 63 years ($M = 37.50$ years, $SD = 11.12$). Although there were a number of new employees ($n = 26$ with 5 years or less of experience working at the company), most participants had been employed there longer: 13 engineers had between 6 and 10 years of experience at the company, and 24 had worked there between 11 and 15 years. One engineer had more than 15 years with the company.

Procedures

TL course. The TL course was an 8-hr introductory course that consisted of lectures with accompanying PowerPoint slides and was taught by company SMEs and university professors who were specialists in composite materials. The course

[1]The company provides thousands of hours of training in areas such as Multiple Structures, Composites Design and Analysis (D&A), Manufacturing, Structures, and Product Lifecycle Management (CATIA-ENOVIA-DELMIA) and offers certificate and master's programs that integrate the rigors of university courses with the real-world application knowledge of company SMEs.

was organized into four major topics related to composites: composite materials, structural analysis and design of composites, manufacturing and tooling with composites, and maintenance and repair of composite materials.

CBL course. The CBL course used in the current study covered the same content related to composites as the TL course; however, the content was reorganized around four challenges with inquiry activities that built on one another over time. Engineers were first asked to individually write their "initial thoughts" about each challenge. As an illustration, the challenge for the first topic was as follows:

> The aerospace company has made a strategic decision to build new aircraft out of composite materials instead of aluminum. In doing so, the new aircraft will be lighter, stronger and more comfortable than if it were built of aluminum. What are composites and what is it about them that make this claim plausible?

They then had approximately 30 min to view resources that showed challenge-relevant, expert content in video format. Then the engineers met for 30 min in small-group discussions that began with an opportunity to write down "later thoughts," compare them to their earlier responses (initial thoughts), and discuss "insights" and "burning questions" in their small groups (see PT3 Group at Vanderbilt, 2003, for the value of comparing initial thoughts with later thoughts).

Finally, participants moved into a large group where members of the smaller groups presented their synthesized ideas about each challenge question, asked further questions that had occurred to them in their small-group discussion, and received feedback from the course experts and other peers from the large group. The large-group session lasted approximately 30 min. Instructors encouraged students to try to answer one another's questions but eventually provided correct information and also elaborated on information when it was relevant. The challenge cycle approach (with different challenges each time) was repeated three more times during the day.

Measures

As noted earlier, we evaluated the differences between the two instructional groups using two primary sources: (a) a paper-and-pencil test used to assess participants' learning and (b) video recordings that captured the interactions of participants in various learning sequences during instruction.

Learning assessment. Learning scientists worked with the company SMEs to design four clusters of items aligned with course content: (a) items requiring integration and application of information to solve problems (Adaptive Knowledge), (b) items requiring explanations of composite concepts

(Comprehension—Structural), (c) items requiring explanations of concepts related to the manufacturing of composite materials (Comprehension—Manufacturing), and (d) items requiring recall of simple factual information (Recall). Members of the research team (aerospace SMEs and University of Washington learning scientists) collaboratively designed a rubric for scoring item responses. The number of points assigned to each test question ranged from 2 to 15, depending on the complexity of the answer required (1 point was given for each key idea as articulated by the rubric). As noted previously, the reason for varying item complexity is that previous research has shown that CBL formats show greater performance gains over lecture formats on test items that require a synthesis of information across a number of different information sources (Martin et al., 2006; Roselli & Brophy, 2006).

Interrater reliability. All participants were individually pretested and posttested on the same set of items that were initially developed (different random orders were used at pretest and posttest). Four scorers, each blind to treatment group and time of test, independently used the rubric to score a common set of 20 randomly selected tests. Total test scores from each pair of scorers correlated at $\geq$.80, and interrater reliability among all scorers was .95 (interrater reliability was computed using Cronbach's alpha, which may be thought of as the average correlation among the four scorers; cf. Nunnally & Bernstein, 1994, pp. 232, 251–252). Disagreements were resolved through discussion with SMEs. Scorers then divided the remaining tests equally among themselves for independent scoring.

Construct validity. Although face validity for the aforementioned content areas was established through the process of creating each test item in collaboration with SMEs, we chose to empirically test construct validity using exploratory factor analysis (EFA) on the intercorrelations among the 20 pretest items. This type of analysis reveals patterns of correlation among variables (in this case, test items) that are thought to reflect distinct underlying processes (factors; in this case, content areas) affecting item response patterns (cf. Stevens, 2002, pp. 385–453; Tabachnick & Fidell, 2007, pp. 607–675). Rules of thumb for EFA sample size ratios vary from 20 subjects per variable to 2 subjects per variable (i.e., Stevens, 2002, p. 395, recommends a ratio of 5:1). Although our data were limited to 64 engineers and 20 items (yielding a ratio of approximately 3:1), the impact of having fewer subjects than would be desired for psychometric test development simply limits generalization of these EFA results (our EFA results may be somewhat sample specific).

There are choices among EFA estimation and rotation algorithms. Our EFA used maximum likelihood estimation (i.e., maximizing the probability that the observed item correlations are sampled from the model-implied parameters; cf. Tabachnick & Fidell, 2007, p. 636) and a Varimax orthogonal rotation. All orthogonal rotations for EFAs have the advantage of preserving estimated item–item relationships while aiding in the interpretation of results by geometrically shifting axes simultaneously in space so that factor–item relationships are as close as possible to respective factor axes; however, Varimax is the recommended orthogonal rotation because it minimizes the complexity of factors by maximizing the variance of the item–factor relationships for each factor (Tabachnick & Fidell, 2007, pp. 637–639).

Results of the EFA (provided in Table 1) showed that a four-factor solution fit the data well, $\chi^2(116, N = 64) = 134.79$, $p > .05$, and furthermore that most of the estimated item–factor correlations, or loadings (correlations between each item and each factor or content area, after the intercorrelations among items are

TABLE 1
Exploratory Factor Analysis Results for Evaluating Leaning Assessment Items

		Loading (Factor–Item Relationships)			
Item	*Communalities*	*Factor 1: AK*	*Factor 2: CS*	*Factor 3: CM*	*Factor 4: R*
1	0.26	0.21	0.13	0.07	0.01
2	0.58	**0.6**	−0.09	0.16	0.31
3	0.46	**0.59**	0.27	0.26	−0.22
4	0.42	0.26	0.03	0.1	0.04
5	0.32	**0.46**	0.06	0.14	0.06
6	0.59	0.22	0.12	**0.56**	0.17
7	0.39	0.09	0.2	**0.44**	0.06
8	0.56	−0.01	−0.11	**0.56**	−0.09
9	0.26	−0.1	0.11	0.03	**0.44**
10	0.45	0.09	0.36	−0.22	**0.41**
11	0.47	0.31	0.01	0.01	**0.49**
12	0.48	**0.59**	−0.08	−0.29	−0.07
13	0.53	**0.7**	−0.07	−0.15	0.14
14	0.38	−0.04	−0.09	−0.03	0.27
15	0.34	0.08	−0.08	0.09	**0.43**
16	0.55	−0.01	**0.43**	0.21	−0.08
17	0.58	0.92	**0.99**	0.05	−0.07
18	0.42	0.19	−0.39	−0.25	**0.48**
19	0.49	0.34	−0.23	−0.25	**0.37**
20	0.34	0.13	−0.11	−**0.41**	0.2
Unrotated variance accounted for		8%	13%	5%	7%
Rotated variance accounted for		11%	8%	7%	8%

Note. $N = 64$. Siginificant loadings are in boldface. AK = Adaptive Knowledge; CS = Comprehension—Structural; CM = Comprehension—Manufacturing; R = Recall.

taken into account), corresponded well with the four aforementioned factors, thus providing empirical support for the theoretically derived test content areas. Only three of the items (1, 4, and 14) had low item communalities (a communality is the percentage of variance in an item explained by the set of factors) and no meaningful relationships with any of the factors. Furthermore, one item (20) exhibited a negative relationship with its factor. Henceforth, these four problematic items were dropped from analyses.

Content reliabilities. To obtain a reliability estimate for each content area of the assessment, we computed internal consistencies (Cronbach's alpha). The typical rule of thumb for modest reliability is .70 (Nunnally & Bernstein, 1994, p. 265), because the squared value of Cronbach's alpha provides the percentage of variance that is shared among the items (and squaring .70 yields 49%—or nearly half—shared variance). However, all reliability estimates are affected by the number of items on which they are based. In this study, the number of items per factor ranged from two to six. Thus, it was not surprising that the estimated internal consistencies were somewhat lower than the preferred .70 value: .63 for Adaptive Knowledge content ($n = 5$ items), .52 for Comprehension—Structural content ($n = 2$ items), .59 for Comprehension—Manufacturing content ($n = 3$ items), and .59 for Recall content ($n = 6$ items). Despite these lower than preferred values, shared variance among items was not insubstantial (i.e., 40%, 27%, 35%, and 35% for each of the factors, respectively), and they corresponded as a whole to each of the content areas developed by the research team.

Content scores. Given our validity and reliability results, we created composite content area scores for each individual at pretest and posttest by averaging the scores of each item related to the content area/factor. For example, an Adaptive Knowledge score was computed by averaging each engineer's score on Items 2, 3, 5, 12, and 13. However, because content area items had a different number of points possible (some items were awarded more points than others based on the rubric), we standardized the items to have a common metric prior to averaging them. Mathematically speaking, standardization transforms scores from their original units, such as test points, to the relative number of standard deviations each score is from its mean (a standard deviation is the average distance of scores to their mean). In this manner, each item's scores, once standardized, had a mean of 0 and standard deviation of 1, with 68% of most scores falling between ± 1 *SD* (cf. Pagano, 2006, p. 82). It is important to note that standardization does not change the distribution of the scores—only their metric.

Learning interactions. Our data sources for capturing discourse patterns included field notes from observations and video and audio recordings of the engineers participating in various learning activities. More specifically, we recorded the engineers talking, listening, and interacting with their peers throughout the 8-hr Introduction to Composites course. Using multiple cameras and audio devices, we captured each of the following:

1. The whole-class portions of both the TL and CBL courses.
2. The small-group discussions related to each challenge (CBL only). These discussions were focused on three broad questions designed to elicit conversations about what engineers were learning (or not learning) about the challenge problem and content. The questions were as follows:
   - In relation to what you have just seen in the Resources, what was surprising?
   - What was not new but you now see it in a different light?
   - And what do you still need help understanding?
3. The small-group "report-out" session at the end of each CBL challenge cycle when each small group shared its ideas with the whole class.

Content logs (Jordan & Henderson, 1995), which captured key moments of activity and discourse, were created from the videotapes to aid analysis. Two researchers used these logs and the video recordings to independently identify significant interactional episodes. Using standardized transcription conventions, content logs, and field notes, we reconstructed in writing what the engineers said and did in relation to one another, preserving the temporal sequence of the interactions. Participant verbal interactions were transcribed and coded. Emergent categories and themes in relation to course content and participant engagement (through questions that stemmed from discussions and interactions) were documented. Verbal interactions were analyzed for sequences that captured participant meaning making.

RESULTS

To evaluate differences between the TL and CBL instructional groups, we consider first the results from the learning assessment pretests and posttests and then the results of discourse analyses.

Learning Assessment Pretest and Posttest

Table 2 displays group pretest and posttest descriptive statistics for each of the assessment content areas (as described in the Methods section, we used standardized scores for equating item metrics prior to creating the four learning assessment content area composites). A positive value indicates that the score is above the grand mean (average across groups), a negative score indicates that the score is

TABLE 2
Learning Assessment Descriptive Statistics

	TL (n = 44)				*CBL (n = 20)*			
	Pretest		*Posttest*		*Pretest*		*Posttest*	
Content Area	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
AK	0.01	0.72	−0.10	0.57	−0.03	0.59	0.24	0.73
CS	−0.08	0.53	0.05	0.79	0.17	1.31	−0.12	0.75
CM	−0.06	0.63	0.08	0.56	0.13	0.92	−0.19	0.71
Recall	0.01	0.53	−0.02	0.53	−0.03	0.67	0.04	0.57

Note. Because of the attrition of two participants, CBL posttest $n = 18$. TL = traditional learning; CBL = challenge-based learning; AK = Adaptive Knowledge; CS = Comprehension—Structural; CM = Comprehension—Manufacturing.

below the grand mean. We tested the differences between groups at pretest on the four content areas using two-group analyses of variance and found no significant differences (F test ps $> .05$).

We next tested whether groups differed on posttests. Of special interest was the Adaptive Knowledge content area, because prior research with novice engineers has shown the benefits of CBL over TL on more cognitively complex knowledge in formal education settings (Martin, 2005; Martin et al., 2006; Roselli & Brophy, 2006) and because the kinds of interactions supported by CBL should support greater connectivity of learning. To explore these ideas we conducted analyses of covariance comparing the groups on posttests using respective pretests as the covariate. This analysis is often recommended for quasi-experimental designs because it allows one to test differences between groups after adjusting for mean pretest level (cf. Tabachnick & Fidell, 2007, pp. 195–237). We note that we tested the assumption of homogeneity of regression slopes for analysis of covariance and found no evidence to the contrary.

Results (see Table 2) showed that the CBL group significantly outperformed the TL group on the Adaptive Knowledge content area, $F(1, 59) = 4.501$, $p < .05$; however, no significant differences were detected for the other content areas (F test ps $> .05$). The effect size for Adaptive Knowledge, calculated as the difference in adjusted posttest means divided by the square root of the mean square error term (cf. Rosenthal, Rosnow, & Rubin, 2000, p. 11), was $d = .59$, which indicates that, after pretest variation is controlled, the CBL group scored approximately 0.6 *SD* higher than the TL group on this assessment content area. By Cohen's (1988) standards, this effect is moderate to large.

Interaction Patterns

We focused on two kinds of discourse analysis: (a) Using classroom talk as a measure of interactivity in the large-group sessions, we coded and counted utterances to describe the reciprocity and types of interactions from a classic Initiate, Respond, Evaluate viewpoint (Mehan, 1985); and (b) using discourse analysis from a grounded theory perspective (Becker, 1995; Strauss & Corbin, 1997), we traced the emergence of patterns of interactions over the small-group discussions. We discuss these findings next.

Large-Group Discussions

Instructor and learner utterances in the large-group sessions were coded as follows:

- *Teacher Instructional Monolog.* Teacher instructs/gives information to students, usually accompanied by PowerPoint slides and whiteboard drawings.
- *Teacher Question.* Teacher asks questions of students during the course of the instruction/presentation.
- *Teacher Response.* Teacher responds to questions that were asked by students during the course of the presentation.
- *Student Topic-Related Comments.* Student offers comments on what has been said during the presentation. Comments are addressed either to an instructor, to another student, or to the entire group.
- *Student Question.* Student asks questions of the instructor or another student during the presentation.
- *Student Response.* Student responds to a question posed by a teacher during the course of the instruction/presentation.

We first consider the interactions in the TL (traditional lecture) group. The content in the TL instruction was divided into four topics, and each topic was discussed for about 2 hr. Instructor A taught the first and third topics; Instructor B taught the second and fourth topics. Figure 1 shows types of instructor and student talk typical of each topic (in this case Manufacturing & Tooling), represented as the percentage of total talk during the (approximately) 2-hr lesson on each topic. Percentages were derived from word counts based on transcripts of the video for each topic.

By far the largest percentage (consistent across all four topics) in the TL condition was Teacher Instructional Monolog. For instance, Teacher Instructional Monolog represented 99.22% of the talk during the Manufacturing & Tooling topic, 100% of the Introduction & Materials topic, 98.99% of the Test & Repair topic, and 88.27% of the Analysis & Design topic. Much lower percentages were found for Teacher Questions (asked of students) and, consequentially, Student Questions (asked of teachers). The low percentage of Student Response reflects the fact that most of the Teacher Questions were rhetorical or required a very short ("yes" or "no") answer. Most of the instructors' questions were recall questions

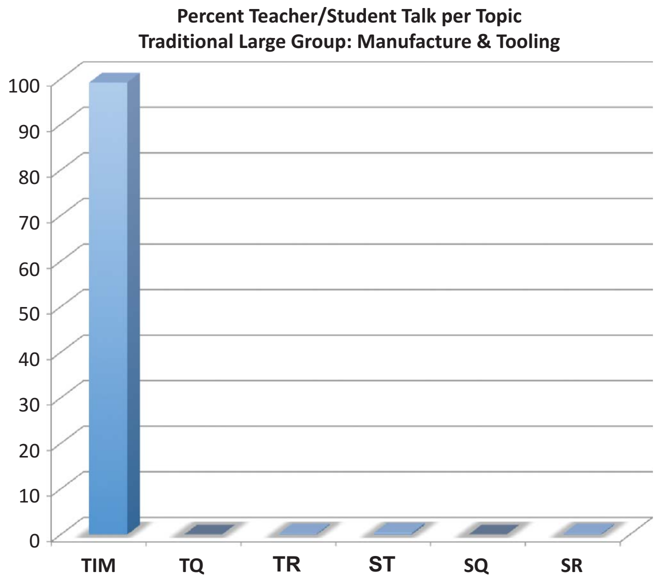

FIGURE 1 Percentage of teacher and student topic-related comments in the Manufacturing & Tooling topic of the traditional learning instruction. TIM = Teacher Instructional Monolog; TQ = Teacher Question; TR = Teacher Response; ST = Student Topic-Related Comments; SQ = Student Question; SR = Student Response (color figure available online).

that required minimal thinking. The following is an example of a typical recall question that came from the TL course (Cazden & Beck, 2003; Mehan, 1979):

> Instructor: The stress concentration for an open hole is factor three. Say, if you see an open hole like this, what's the stress concentration for that?
> Engineer Joe: Three?
> Instructor: Three, right.

The data in Figure 1 also illustrate that student-initiated questions to the instructor or the class were extremely low. Overall, the findings from the TL course show that a great deal of information was presented by the instructors, with relatively little participation by students.

Engineers in the CBL course received the same instructional topics, content, instructors, and learning time as those in the TL course. However, as discussed previously, the topics presented in the CBL course were organized around challenges that were worked on individually (e.g., initial thoughts and resources), in small groups, and then with the whole class. As in the analysis of the TL condition, six categories of talk were coded. Figure 2 displays the percentages for

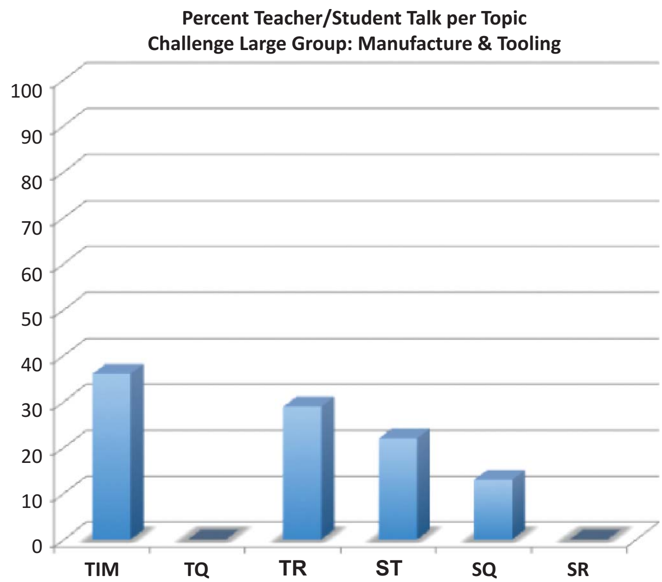

FIGURE 2 Percentage of teacher and student topic-related comments in the Manufacturing & Tooling topic of the challenge-based learning instruction. TIM = Teacher Instructional Monolog; TQ = Teacher Question; TR = Teacher Response; ST = Student Topic-Related Comments; SQ = Student Question; SR = Student Response (color figure available online).

the Manufacturing & Tooling topic in the CBL condition. Teacher Instructional Monolog represented only 36.08% of the talk during the Manufacturing & Tooling topic, 22.62% of the Introduction & Materials topic, and 41.81% of the Test & Repair topic. (The interactions during the fourth topic have special characteristics that we discuss below.) The pattern in the whole-class sessions of the CBL group was very different from that of the TL group (see Figure 2) in that much of the large-group teacher "talk" in the CBL condition was in response to student questions. The following is an example of the interactions that occurred during the whole-group CBL sessions. In this example, an engineer (John) is probing the nuances of something that has specific significance for him in his line of work at the company.

Engineer John: How are we to handle the electrical system in this change from metal to composites?

Instructor: On the wing we have copper strips. On the fuselage we have interwoven copper and bronze wire with the fabric . . . [and so forth, in considerable detail for ≥2 min; sensitive information omitted because of the proprietary nature of the content]

The overall electrical system, which is intimately connected with lightning strike, is more complex for engineers working with composites than for engineers working with metals because composites are nonconductive, whereas aluminum is highly conductive. In the preceding example, Engineer John focuses on the question "how" and includes his teammates as "we," even though many of the individuals in the room would not be as knowledgeable as he is in this area. Furthermore, he asks about an issue of vital importance for aircraft manufacturing—lightning strike.

A feature of engaged, attentive questioning (Dillon & Wittrock, 1984) is illustrated in John's follow-up to his initial question, when clarification and depth are achieved not only for him but also for the other participants. As shown in the next excerpt, another participant, Engineer Bob, who was initially silent, widens the discussion around an issue that was not on the syllabus in the first place:

Engineer John: And so repair for that would be . . . ?
Instructor: The same material. [Directly responding to John]
Engineer John: Oh! The same. Thank you.
Engineer Bob: So then you have to test the resistance? That would be a part of the repair as well? [Thinking out loud]
Instructor: Yes. [With more explanation in considerable detail—more than 2 min. Sensitive information omitted because of the proprietary nature of the content]

Although the topics of conductivity and repair had been included in the resources that participants watched prior to the whole-group sessions, the engineers' discussions of their different work settings and particular concerns (e.g., ensuring that repairs do not damage the electrical pathways) seemed to provide a context that gave these topics special meaning and import. Analysis of the videotapes suggested that the richness of the exchanges was engaging and relevant not only for the actors (John and Bob, who asked the questions, and the instructor) but also for the other participants of the course who were actively sitting forward and acknowledging the information.

The CBL approach provides focused challenges and resources that set the stage for and guide small-group discussions in which people can learn not only the content but also about and from one another. Because participants had deep knowledge of their jobs, they could inform one another in ways that went beyond the fixed curriculum. Analysis of the videotapes confirmed that many issues were raised and that many questions were asked and answered—both in the small groups and when the small groups returned to the large group with the instructor. Many of these issues and questions were not on the syllabus for the course, nor were they measured in the pretest or posttest, yet they touched on a number of important concerns for working with composites. Two SMEs independently rated the importance of issues that were identified from the transcripts. Ten of the issues (which are not "export compliant" and therefore cannot be listed because of the sensitivity of the content) were rated as "highly important" in the context of composite airplane design and manufacturing. These emergent discussions could result in learning from the course that is not documented by the posttest; at a minimum, these discussions represent a platform for continual course improvement, where issues and ideas that emerge could be incorporated in subsequent courses.

As noted earlier, both the TL and CBL courses covered the same four topics related to composite materials. The only difference between courses stemmed from the organization of the material around interactive challenges as opposed to PowerPoint-aided lectures. For the CBL group, we have discussed only three of these topics. For the fourth topic, the instructor reverted at the last minute to lecturing. He reported a worry that there was not enough time to cover the content and prepare the students for the test.

Despite receiving a lecture for the final challenge cycle, previous experiences with challenges seemed to carry over to the participants' behavior during the fourth topic. Figure 3 shows the discourse patterns for the fourth topic of the TL class and compares them to the patterns of the CBL class, who had experienced the small-group interactions for the first three topics and who listened to a lecture for the fourth topic.

Both graphs represent the same teacher teaching the same topic (i.e., Analysis & Design). The graph on the right—from the fourth topic of the CBL class when the instructor lectured—shows changes toward more discussion by engineers. The instructor talked less and was more likely to ask questions of the students. Similarly, students were more likely to add information to the lecture and ask questions of the instructor.

Knowledge Establishing vs. Knowledge Sharing

As noted earlier, students who were in the CBL group first saw and responded to a challenge, studied relevant resources on their own for approximately 30 min, and then met in small groups to prepare a set of "new insights" and "burning questions" that they would then present to the whole class. To study the engineers' small-group interactions, we analyzed the amount and type of student-to-student discourse for the first three challenges. The total amount of time for small-group discussion was approximately 90 min—30 min for each of the first three challenges.

We categorized the engineers' discourse during these sessions as either knowledge establishing or knowledge sharing. An *exchange* was defined as a set of statements or questions that related to a common topic. *Knowledge-establishing*

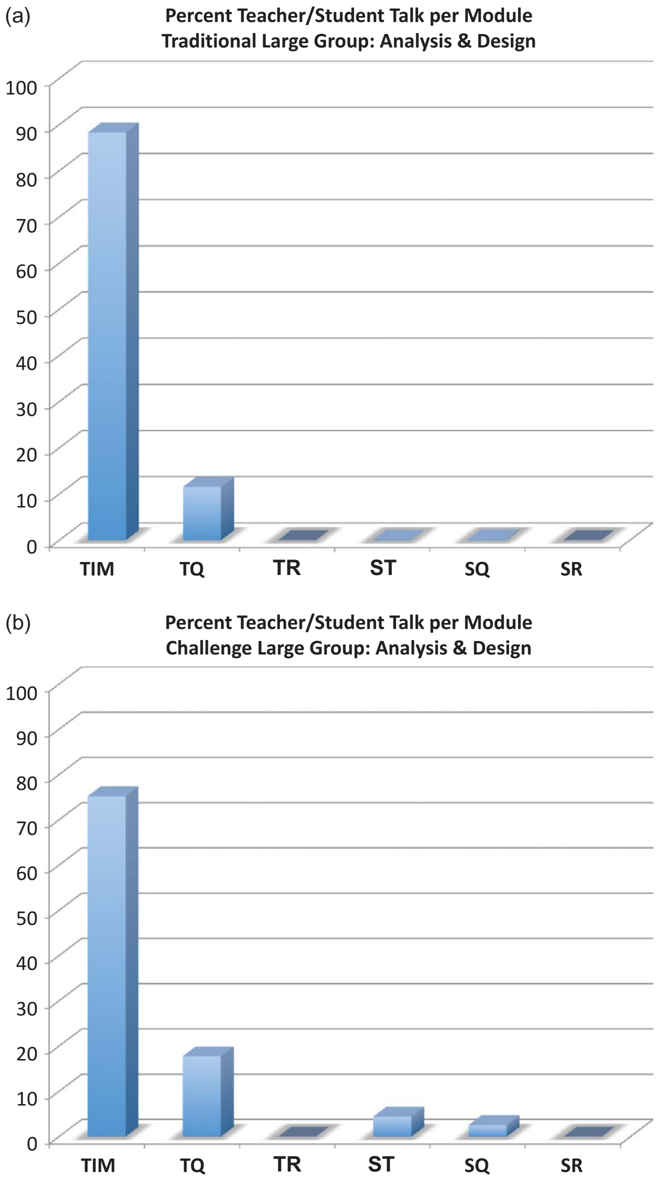

FIGURE 3 Percentage of teacher and student topic-related comments in the Analysis & Design topic of the (a) traditional and (b) challenge-based learning instruction. TIM = Teacher Instructional Monolog; TQ = Teacher Question; TR = Teacher Response; ST = Student Topic-Related Comments; SQ = Student Question; SR = Student Response (color figure available online).

exchanges were statements that individuals made about their training and knowledge; for example, the following exchange occurred early in the first breakout session:

Engineer Norm:	I actually got my master's in physics a couple of years ago. It was in optics not in composites.

Knowledge-sharing utterances were ones in which individuals were asking questions of one another and jointly working to find answers. For example, we counted the next utterance as a knowledge-sharing utterance—it was the first instance in which an individual was willing to acknowledge something that he did not know and ask for help from the group.

Engineer Leo:	Yeah. I've done repairs and understand the basic program [but] I was not aware that these materials were as strong as they are.
Engineer Sue:	Yeah.
Engineer Norm:	What is the difference in effect between, um, graphite and fiber?
Engineer Sue:	Six times.
Engineer Pat:	Huge strength!
Engineer Norm:	It's lighter at the same time, yeah.
Engineer Leo:	And ten times more capable. That's impressive. I wasn't aware that there was that much difference in strength.

We conducted a developmental analysis of small-group exchanges over sessions. We were able to do a complete analysis for only one of the three groups because we did not have video of the first small-group session for the other two groups. Nevertheless, data from the groups we were able to follow from beginning to end, and later data from the other two groups, are informative.

As shown in Figure 4, early exchanges among participants in the small-group sessions focused primarily on knowledge sharing. During the first 10 min of their small-group interactions, we counted 4 knowledge-sharing utterances and 14 knowledge-establishing utterances. All of the engineers in the group made statements to establish their expertise. For example, one of the new engineers explained, "I got my master's in physics . . ." Another member stated, "I was an advisor for years . . ." A third mentioned a number of years at school and in different geographic locations with well-known, reputable engineering firms: ". . . seven years you know Company X down in San Diego. Also down in Tucson and also down in LA."

During the last 10 min of the small group's session, there were no knowledge-establishing comments and many more knowledge-sharing comments (8 in total) than in the first part of the session.

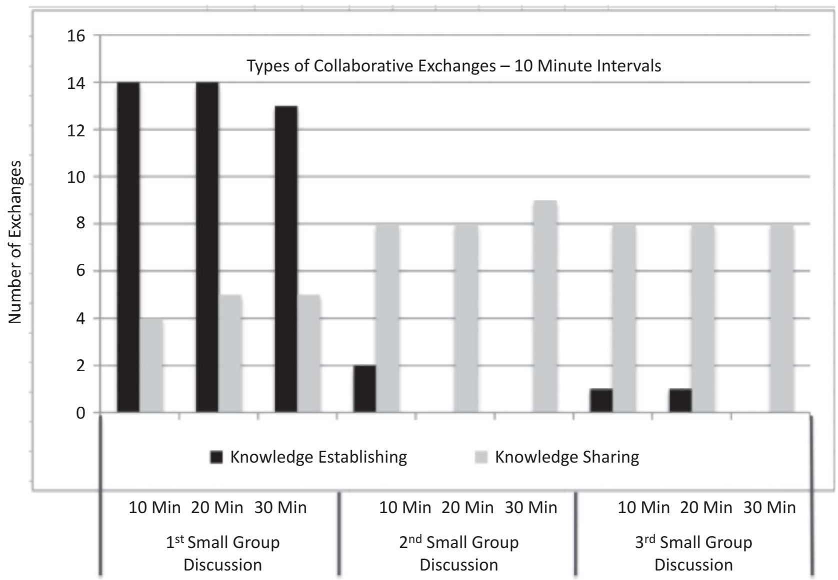

FIGURE 4 Number of knowledge-establishing and knowledge-sharing exchanges per 10-min interval.

The following is an example of a knowledge-sharing exchange:

Engineer Ted: I guess there was one thing that Bud and I were trying to figure out about the matrix. [Looks at Bud]
Engineer Bud: Glass Transition Temperature, glass temperature [Looks at Ted, shifts gaze to the rest of group]
Engineer Ted: Yeah, yeah, they were saying where the matrix, um, the matrix starts to lose its stiffness. There was a curve where they show as you increase temperature, it falls off gradually until there is a drop-off point. [Elaborate hand swoops to show the drop off]
Engineer Bud: I knew about surface temperature, I didn't know what was the word for it.
Engineer Ted: . . . I'm just trying to think, well, glass unit. It doesn't go solid. Right? I'm not sure, maybe we could ask [the instructor] about that. [Looks toward SME]

Another example of a knowledge-sharing exchange is the following:

Engineer Jan: [reads from challenge cycle document] What concepts would you like to have clarified? [Turns to other members of the group, eyes each one separately] I was just saying I'm . . . unclear about, compression loading. I'm curious to understand it. You know. I assume that the fibers take the compression loading. And also, um, although, if you think about the . . . most common composite material, like reinforced concrete. Right!

Engineer Bob: Good idea building airplanes with reinforced concrete! [Laughter]

Engineer Jan: [Smiling—then turns serious] The concrete takes the compressive loading there. Now, is the matrix taking the compressive loading or is it the fiber?

Engineer Bob: [Serious now] The fiber is taking it, and then the matrix stabilizes it.

Engineer Jan: Stabilizes it!

Engineer Bob: Stabilizes it, so that it won't buckle.

Engineer Jan: Oh, OK. [Nodding] That makes sense.

The engineers also discussed issues that went beyond airplane design; for example, one engineer asked, "Why are we driving metal (automobiles) instead of composites—the vehicles we are driving are not composite?"

As noted earlier, there were two other groups in the CBL condition, but for these we were only able to videotape their interactions during later parts of the training after they had already had the chance to get to know one another. During these later sessions, both of these groups showed almost all knowledge sharing and no knowledge establishing.

DISCUSSION

As noted earlier, the present study involved collaboration between people trained in the learning sciences and instructional designers, coaches, and SMEs from Boeing. When we began our work, Boeing was going through a transition from building planes primarily out of aluminum to building them from composite materials, which were much stronger and lighter than aluminum. This has major implications for building airplanes, including changes in materials procurement and testing, airplane design, tooling for putting the planes together, ways to ensure safety (e.g., from lightning strikes) and repair, and many other issues. The incumbent workers at Boeing needed to learn about composites, and an initial 1-day course that became the focus of the present study was designed to give them an introduction to the topic and its implications for their future work.

The results of our comparative research involving lecture- and challenge-based instruction showed greater interaction—more sharing of knowledge related to and beyond the course content—among participants in the challenge-based group. In addition, the challenge-based group performed significantly better on posttest items requiring integration and synthesis of concepts. The increased interactivity in the challenge course provided opportunities for the course instructor and participants to articulate connections among concepts and build upon expressed ideas. In turn, working on connections may have contributed to the challenge participants' better synthesis of learned concepts.

As with any study, these findings have limitations. From a methodological perspective, the quasi-experimental nature of the research limits our inferences to the pattern of relationships we observed in the data rather than proving that CBL causes improved learning. Nevertheless, data from our pretests demonstrated no significant differences in the groups prior to training and thus bolsters our confidence that the group differences we observed in learning and patterns of interaction may be replicated in a randomized experiment. Also, our small sample size limits generalization of the EFA results of the learning assessment items to other samples. However, the pattern of results did correspond to the theoretically derived content areas.

An important outcome of the present work involves new questions about learning that arose in the context of work with Boeing. When the collaborative study began, it fit closely with what Stokes (1997) defined as Edison's quadrant. The challenge-based legacy cycles had been used in a number of K–16 settings and had shown stronger learning (compared to lecture-based methods) for posttest items that required a synthesis and application of information (e.g., Martin, Rayne, Kemp, Hart, & Diller, 2005; Roselli & Brophy, 2006). However, the challenge-based legacy cycles had not been used to study learning in the workplace or with domain experts. Nor had prior research focused on the patterns of interactions that occur during CBL among engineers (or other learners).

As the work progressed and the learning scientists learned much more about Boeing, the work moved from an Edison-like question of "How does the Legacy pedagogy affect learning in a workplace setting?" to more of use-inspired "theory building" (Pasteur's quadrant) that raised a number of questions that go beyond the study presented here. The surfacing of these questions suggests the value for the learning sciences of working in settings that are quite different from those such as schools and universities.

One question that arose—perhaps obvious in retrospect but not when we began this work—involved the fact that scores on the final test of a course are far from sufficient to ensure the kinds of performance criteria needed for successful performance by the engineers. Boeing is accountable for building airplanes (each has more than 1 million parts) in a manner that is safe and has almost no room for error. Getting a perfect score on the posttest of the course discussed in this study does not guarantee this kind of expertise by the engineers. One answer to this problem, of course, is to note that we focused on the introductory course to composites and the company has many other more advanced courses. Still, would perfect scores in these courses ensure the kinds of expertise needed for Boeing to be successful? The answer is no.

The accountability for effective learning from training courses comes from how people thereafter improve at their jobs. Employees are never given a summative standardized test and put to work with no follow-up. Teams of coaches monitor all work, as do work foremen, and they reach out to individuals and to groups when they see performances that are below par. High standards are expected. Support and incentives are readily available for achieving these standards. Furthermore, unlike many school tests, the performances expected of learners are not "hidden." By using strong and *ongoing* performance measures, and seeing how much remediation is needed to help people achieve at increasingly higher levels, Boeing is able to create return on investment indices that allow them to evaluate the impact of local innovations on the organization as a whole. This was a totally new concept for the learning scientists involved in this project, and we are still learning how to think about this in productive ways.

The role of coaches at Boeing is also different from the role of coaches in many schools. Coaches in K–12 settings tend to focus on helping instructors teach better. They do this at Boeing, too, and often even create new courses when they identify particular knowledge or skill gaps among groups of employees. But coaches at Boeing are also available to help individuals who either request help or may want to try some new innovation that can help the company. And they keep an eye out for individuals who need help but do not realize it.

A related issue involves choice and agency. Some courses are mandatory, and people have to take them until they reach a specific level of certification. In many other cases Boeing employees are free to select from a variety of online and face-to-face courses and choose those that they feel are most helpful for helping them do better work. This sense of agency is often missing in schools and can have implications for students' levels of motivation and involvement in the learning process (Shutt et al., 2010). Of course, for both schools and the workplace, safeguards need to be put in place to help people choose the best options for their future. But we could imagine schools, or at least project structures in schools, in which the goals of performance are clear and coaches and others (e.g., older students) are tasked with helping everyone succeed.

Another issue involves much more complete study of the informal and formal learning opportunities that exist in a company such as Boeing. Ideally, opportunities to work collaboratively in courses will allow people to learn more about one another's expertise and hence be able to tap it once they leave the course, potentially amplifying the impacts of courses on learning and job performance. Testing this idea was beyond what was possible in the present study, but we think that it is an important idea for future research. A second set of collaborative studies exploring issues of online learning and social interaction at Boeing is discussed by Lawton, Vye, Bransford, French, and Richey (in press).

Overall, our findings suggest that exploring learning science–based designs from the K–12 environment in companies such as Boeing could broaden thinking in the learning sciences. Furthermore, theory building grounded in workplace settings may add value to learning science research in K–12 settings.

REFERENCES

Becker, H. S. (1995). The epistemology of qualitative research. In R. Jessor, A. Colby, & R. A. Shweder (Eds.), *Ethnography and human development: Context and meaning in social inquiry* (pp. 53–72). Chicago, IL: University of Chicago Press.

Bell, P., Lewenstein, B., & Shouse, A. (Eds.). (2009). *Learning science in informal environments: People, places and pursuits*. Washington, DC: The National Academies Press.

Bransford, J. D., Sherwood, R. S., Hasselbring, T. S., Kinzer, C. K., & Williams, S. M. (1986). Anchored instruction: Why we need it and how technology can help. In D. Nix & R. Spiro (Eds.), *Advances in computer video technology* (pp. 115–141). Hillsdale, NJ: Erlbaum.

Bransford, J. D., Vye, N. J., Stevens, R., Kuhl, P., Schwartz, D., Bell, P., . . . Sabelli, N. (2006). Learning theories and education: Toward a decade of synergy. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (Vol. 2, pp. 209–244). Mahwah, NJ: Erlbaum.

Cazden, C. B., & Beck, S. W. (2003). Classroom discourse. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 165–199). Mahwah, NJ: Erlbaum.

Cognition and Technology Group at Vanderbilt (CTGV). (1997). *The Jasper Project: Lessons in curriculum, instruction, assessment, and professional development*. Mahwah, NJ: Erlbaum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Dillon, J. T., & Wittrock, M. C. (1984). Research on questioning and discussion. *Educational Leadership*, *42*(3), 50–56.

Eraut, M. (2004). Transfer of knowledge between education and workplace settings. In H. Rainbird, A. Fuller, & A. Munro (Eds.), *Workplace learning in context* (pp. 201–221). London, England: Routledge.

Gragg, C. I. (1940, October 19). Because wisdom can't be told. *Harvard Alumni Bulletin*, pp. 78–85.

Hmelo-Silver, C. E., & Barrows, H. S. (2006). Goals and strategies of a problem-based learning facilitator. *Interdisciplinary Journal of Problem-based Learning*, *1*, 21–39.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences*, *4*(1), 39–103.

Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., & Holbrook, J. (2003). Problem-based learning meets case-based reasoning in the middle-school classroom: Putting learning by design into practice. *Journal of the Learning Sciences*, *12*, 495–547.

Lawton, D., Bransford, J. D., Vye, N. J., French, D., & Richey, M. (in press). Structuring feedback for online learning in the workplace. *Journal of Engineering Education*.

Martin, T. (2005, April). Measuring preparation for future learning in children's mathematics: Instructional implications. Paper presented at the American Educational Research Association, Montreal, Canada.

Martin, T., Petrosino, A. J., Rivale, S., Rayne, K., Pierson, J., Svihla, V., & Diller, K. (2006, April). *The Legacy cycle as a resource for curriculum development*. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

Martin, T., Pierson, J., Rivale, S. R., Vye, N. J., Bransford, J. D., & Diller, K. R. (2007). *The function of generating ideas in the Legacy cycle*. Arlington, VA: International Network for Engineering Education and Research.

Martin, T., Rayne, K., Kemp, N. J., Hart, J., & Diller, K. R. (2005). Teaching for adaptive expertise in biomedical engineering ethics. *Science and Engineering Ethics*, *11*(2), 257–276.

Mehan, H. (1979). *Learning lessons*. Cambridge, MA: Harvard University Press.

Mehan, H. (1985). The structure of classroom discourse. In T. van-Dijk (Ed.), *Handbook of discourse analysis* (Vol. 3, pp. 119–131). London, England: Academic Press.

Michael, A., Klee, T., Bransford, J., & Warren, S. (1993). The transition from theory to therapy: Test of two instructional methods. *Applied Cognitive Psychology*, *7*, 139–154.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Pagano, R. (2006). *Understanding statistics in the behavioral sciences* (8th ed.). Belmont, CA: Thomson Wadsworth.

PT3 Group at Vanderbilt. (2003). Three AMIGO3s: Using "anchored modular inquiry" to help prepare future teachers. *Educational Technology, Research and Development*, *51*(1), 105–123.

Roselli, R. J., & Brophy, S. P. (2006). Effectiveness of challenge-based instruction in biomechanics. *Journal of Engineering Education*, *95*(4), 311–324.

Rosenthal, R., Rosnow, R., & Rubin, D. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge, UK: The Press Syndicate of the University of Cambridge.

Schwartz, D. L., Lin, X., Brophy, S., & Bransford, J. D. (2000). Toward the development of flexibly adaptive instructional designs. In C. M. Reigeluth (Ed.), *Instructional design theories and models*: *A new paradigm of instructional theory* (Vol. 2, pp. 183–213). Hillsdale, NJ: Erlbaum.

Shutt, K., Phillips, R. S., Vye, N. J., Van Horne, K., & Bransford, J. D. (2010, April). *Developing science inquiry skills with challenge-based student-directed learning*. Paper presented at the meeting of the American Educational Research Association, Denver, CO.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Erlbaum.

Stevens, R. (2000). Divisions of labor in school and in the workplace: Comparing computer and paper-supported activities across settings. *Journal of the Learning Sciences*, *9*, 373–401.

Stokes, D. E. (1997). *Pasteur's quadrant—Basic science and technological innovation*. Washington, DC: Brookings Institution Press.

Strauss, A., & Corbin, J. (Eds.). (1997). *Grounded theory in practice*. Thousand Oaks, CA: Sage.

Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics* (5th ed.). New York, NY: Allyn and Bacon.

Tzou, C., & Bell, P. (2010, June). *Micros and me: Leveraging home and community practices in formal science instruction*. Paper presented at the International Conference of the Learning Sciences, Chicago, IL.

Williams, S. M. (1992). Putting case-based instruction into context: Examples from legal and medical education. *Journal of the Learning Sciences*, 2, 367–427.