

Differentiated Overt Learning Activities for Effective Instruction in Engineering Classrooms

Muhsin Menekse,^a Glenda S. Stump,^b Stephen Krause,^c
and Michelene T. H. Chi^c

^aUniversity of Pittsburgh, ^bMassachusetts Institute of Technology,
^cArizona State University

Abstract

Background Similar to other domains, engineering education lacks a framework to classify active learning methods used in classrooms, which makes it difficult to evaluate when and why they are effective for learning.

Purpose/Hypothesis This study evaluated the effectiveness and applicability of the Differentiated Overt Learning Activities (DOLA) framework, which classifies learning activities as *interactive*, *constructive*, or *active*, for engineering classes. We tested the ICAP hypothesis that student learning is more effective in *interactive* than *constructive* activities, which are more effective than *active* activities, which are more effective than *passive* activities.

Design/Method We conducted two studies to determine how and to what degree differentiated activities affected student learning outcomes; we measured student knowledge and understanding of materials science and engineering concepts.

Results Study 1 showed that students scored higher on all postclass quiz questions after participating in interactive and constructive activities than after the active activities. Student scores on more difficult, inference questions suggested that interactive activities provided significantly deeper learning than constructive or active activities. Study 2 showed that student learning, in terms of gain scores, increased systematically from passive to active to constructive to interactive, as predicted by the ICAP hypothesis. All the increases, from condition to condition, were significant.

Conclusions Our analyses of classroom activities in the engineering domain showed that they fit within the taxonomy of the DOLA framework. The results of the two studies provided evidence to support the predictions of the ICAP hypothesis.

Keywords active learning; engineering learning; instructional methods

Introduction

Active learning refers broadly to innovative student-centered instructional approaches that dynamically involve students in the learning process. The main constructs of active learning are the participation and the engagement of students with concrete learning experiences,

knowledge construction of students via meaningful learning activities, and some degree of student interaction during the process.

Active learning has been studied in many closely related disciplines, including the learning sciences, educational psychology, science education, and recently, engineering education (e.g., Chen, Lattuca, & Hamilton, 2008; Heller, Beil, Dam, & Haerum, 2010; Lin & Tsai, 2009; Prince, 2004; Prince & Felder, 2006). In many studies, active learning has been implemented in the context of problem-based, inquiry-based, discovery, collaborative, cooperative, team-based, and inductive learning methods (e.g., Johnson, Johnson, & Smith, 1991; Prince, 2004; Schroeder, Scott, Tolson, Huang, & Lee, 2007). Examples of active learning from the engineering education literature include examination of student learning from inquiry-based real-life problems (Higley & Marianno, 2001), use of multimedia to facilitate student interaction (Burlison, Ganz, & Harris, 2001), use of a teamwork-based approach to solve complex problems (Pendergrass et al., 2001; Purzer, 2011), use of activity-oriented instruction to increase active engagement (Shooter & McNeill, 2002; Starrett & Morcos, 2001), and comparing collaborative learning methods with traditional instruction (Terenzini, Cabrera, Colbeck, Parente, & Bjorklund, 2001).

In contrast to active learning, *passive learning* usually involves *teacher-centered* methods that favor direct instruction in which students often learn through listening to and observing lectures presented by an instructor. Active and passive learning methods have been contrasted by using pairwise designs in which students in one condition engage in some kind of an active intervention, whereas students in another condition passively receive information from an instructor, expert, or a computer system. These contrasts include comparing inductive versus deductive reasoning (Lott, 1983; Prince & Felder, 2006), inquiry-based instruction versus direct instruction (Minner, Levy, & Century, 2010), discovery learning versus traditional methods (Klahr & Nigam, 2004), and collaborative learning versus learning from lecture (Terenzini et al., 2001).

Although studies have often shown that compared with passive methods, active learning led to superior student learning outcomes (Lambert & McCombs, 1998; National Research Council, 1996; Prince, 2004; Schroeder et al., 2007; Smith, Sheppard, Johnson, & Johnson, 2005), other studies found either the opposite effect (e.g., Kirschner, Sweller, & Clark, 2006; Klahr & Nigam, 2004; Mayer, 2004; Montpas, 2004) or that active is not always better than passive (e.g., Colliver, 2000; Hundley, 2007; Martin, 2009; Osman, 2008a; Osman, 2008b; Sadler, 2002; Stull & Mayer, 2007; Sendag & Odabasi, 2009; Strobel & van Barneveld, 2009; Rittle-Johnson, 2006; Wilson, 1999). These discrepant results make it difficult to reach definitive conclusions about the relative effectiveness of these methods (Lederman, Lederman, & Wickman, 2008; Blanchard, Southerland, Osborne, Sampson, Annetta, & Granger, 2010).

Three general problems with the notion of active learning may shed light on why discrepant results have been obtained. The first problem may be with how the effectiveness of active learning methods is measured (Prince, 2004), since evidence for content validity and difficulty level of individual test items is typically not reported in the literature. Evidence for content validity supports the premise that test items are accurate and cover a representative sample of content from a given domain (Messick, 1995). Knowledge about item difficulty is necessary to understand the depth of student learning as evidenced by their test scores. If test items are easy and measure lower levels of cognitive processing (e.g., recall), test results may not easily favor active learning methods, and the results may not even differ significantly from more passive forms of learning (Chi, 2009). Active

learning methods, however, may have more significant effects on learning in an engineering curriculum in which higher levels of cognitive processing (e.g., knowledge synthesis) are needed to succeed. A second problem that may lead to discrepant results is the lack of shared terminology or definition for active learning methods across various disciplines. For example, some studies use active learning synonymously with inquiry and classify any hands-on activity as an inquiry intervention without stating the important aspects of inquiry, such as to what degree students will be responsible to generate research questions, or who is in charge (i.e., teacher or students) of deciding data collection methods. Another example of the lack of shared terminology appears in team-based learning. Teams and team-based learning are very popular in engineering schools as a way to foster active learning. Some studies classify any group of students working together for any length of time as a team, even though being in a group does not guarantee its members will necessarily collaborate in a productive way.

A third problem is that active learning methods include all sorts of classroom activities that engage students with the learning experience in some manner, and this broad scope may also account for the inconsistent findings. But treating all classroom activities as engaging students in the same way ignores the unique cognitive processes associated with each type of activity.

The lack of a comprehensive framework and taxonomy regarding the components and characteristics of active learning methods make it difficult to compare and contrast the value of active methods in different studies in terms of student learning.

The DOLA Framework

To address the lack of a framework and taxonomy about active learning, Chi (2009) and Chi and Wylie (under review) proposed the Differentiated Overt Learning Activities (DOLA) framework that divided active learning methods into three modes, as being *interactive*, *constructive*, or *active*, depending on what activities students overtly display. The framework differentiates and makes a claim only about overt engagement activities because these are the only behaviors that we (and teachers in classrooms) can observe. Each mode corresponds to a distinct set of cognitive processes.

The motivating question behind development of this framework was whether some modes of overt engagement activities are more effective than others. Based on the cognitive processes corresponding to each mode of activity, the framework can be used to predict the differential effectiveness of these activities on students' learning outcomes. After a review and reinterpretation of experimental studies in the learning sciences literature, Chi (2009) found that all three active, constructive, and interactive modes are better than the passive mode in terms of student learning. There were also differences in the learning of the three nonpassive modes, such that pairwise comparisons of the three nonpassive modes indicated that interactive activities were more likely to be better than constructive activities, which in turn were better than active activities, resulting in the ICAP hypothesis that student learning is more effective in *interactive* activities than *constructive* activities, which are more effective than *active* activities, which are more effective than *passive* activities (Chi & Wylie, under review).

Chi provided an excellent review of the existing learning studies in order to make comparisons across conditions. But since the reviewed studies typically used two conditions, Chi's analysis consisted only of pairwise comparisons. Also, the studies reviewed by Chi used different variables, such as different populations, interventions, concepts, or sample sizes. None of the studies reported by Chi were in an engineering context.

Chi's (2009) DOLA framework differentiates a variety of overt engagement activities that have previously been considered active learning methods. This framework claims that different modes of overt engagement have different learning effectiveness because they have different features and involve different cognitive processes. Many types of activities fit into each mode. The framework assumes that the activities designed as active are expected to involve learners in doing some manipulation with the learning materials; the activities designed as constructive are expected to facilitate the generation of new ideas, beyond those directly presented; and the activities designed as interactive are often expected to generate ideas that build on each other, but only when both students are contributing substantial joint intellectual effort. These overt engagement activities, as defined, predict learning effectiveness such that interactive activities are more likely to be superior to constructive activities, which in turn are almost always better than active activities, and all three active modes are better than a passive mode such as receiving instruction only (referred to as the ICAP hypothesis).

Chi (2009) discusses three main advantages of the DOLA framework: (1) the classification of overt activities helps researchers, instructors, and instructional designers decide what type of activity or intervention would be appropriate for the intended research or instruction; (2) the hypothesized underlying cognitive processes of each mode of activity predict the relative effectiveness of the activities in terms of learning; and (3) the differentiation of activities or interventions based on the underlying cognitive processes allows us to reanalyze or reinterpret the studies in the literature and to clarify the inconsistent findings.

Note that this framework differentiates and makes a claim only about overt or observable engagement activities. Clearly, students may also covertly interact cognitively with information (e.g., construct knowledge while self-explaining silently), but this behavior is difficult to assess reliably in a classroom and may occur with only a small portion of students in any classroom. Similarly, it is possible that overt activities may be provided to students and yet they still do not cognitively interact with the information; their attention may be focused elsewhere at that moment. Despite these caveats, the framework suggests that, overall, different modes of engagement activities differentially affect the amount of learning.

A possible barrier to results as predicted by Chi's hypothesis is proper implementation of activities. In other words, even if researchers properly design and classify activities as active, constructive, or interactive, there still may be obstacles to successful implementation of those activities in the classroom, and learning outcomes may not match expectations. For example, in an interactive activity such as argumentation, if students are not actively challenging each other's claims (Hausmann, 2006) or if only a few of the students participate in the discussion, the activity may not provide the anticipated benefits for those who do not contribute.

The goal of this study is to enrich the engineering education literature by adding experimental learning studies in the engineering context. Our contribution is twofold. First, this is the first study to evaluate the DOLA framework using data both from a controlled experiment and from an *in vivo* classroom setting. Second, this is the first study to explore the value and utility of this framework in an engineering context. Engineering educators are investigating new approaches to curricula and rethinking and developing innovative ways to replace traditional teaching methods; we believe our study will substantively contribute to this effort and discussion. Since there are few controlled studies investigating engineering classroom activities, our study supports engineering educators' use of the DOLA framework when developing classroom activities as a means to increase student

learning. It also supports researchers' use of the framework when designing future studies to advance our understanding of the relative effectiveness of active learning methods.

Three Active Modes of Activities: Active, Constructive, and Interactive

The following sections describe each mode of active learning engagement.

Being Active

In the active mode, students undertake overt activities that activate their own knowledge within the boundaries of the desired content. Students do something or manipulate the instructional information overtly, rather than passively receive information or instruction while learning or studying (Chi, 2009). Active activities emphasize the selected passages or manipulated components of a task, thus allowing students to pay more attention to them. The cognitive processes hypothesized by Chi that correspond with active activities are activating and searching for related knowledge, and encoding, storing, or assimilating new information with activated knowledge. These processes strengthen the existing knowledge and fill the gaps in knowledge, making it more retrievable and more complete. On the basis of these cognitive processes, Chi predicted that students who engage in active learning activities learn better than students who are more passive and not engaging in any observable learning activities, even though these passive students are oriented toward instruction and are receiving the learning materials.

Examples of the active mode include following the procedure of a highly structured experiment, repeating sentences out loud after hearing them, underlining or highlighting some sentences while reading, copying the solution of a problem from the board while the teacher is solving it, selecting from a list of choices as in matching tasks, looking and searching for specific information in a text or problem, or playing a video game without making strategic decisions. For example, an in-class activity demonstrating the relationship of macroscopic properties to the strength of atomic bonding of pure metals could be implemented in an active mode if students underline the text sentences explaining this topic in their class notes, or if students flex three rods of three different metals to feel the stiffness of each rod whose center is heated to a high temperature. Students may be able to link this experience to their prior hands-on "everyday experience" or knowledge of materials when they see and feel the flexing of the rods. Table 1 shows the four modes (passive, active, constructive, interactive) of activities in the context of illustrating the relationship of the stiffness and melting points of rods with the atomic bonding strength of pure metals. This table also shows the associated underlying cognitive processes with each type of activity.

Being Constructive

In the constructive mode, students undertake activities in which they generate knowledge that extends beyond the presented materials. In the active mode, for example, simply repeating a paragraph or underlining text does not extend beyond what was presented. But self-explaining, or explaining aloud to oneself a concept presented in a text, is constructive because it constructs meaning beyond the given content (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, de Leeuw, Chiu, & Lavancher, 1994). The following types of activities can all be considered to be constructive: drawing a concept map, taking notes in one's own words from a lecture, generating self-explanations, comparing and contrasting different

Table 1 Modes of Presenting an In-class Student Activity with the Associated Cognitive Processes

Passive	Active	Constructive	Interactive
<p>Exemplary in-class activity</p> <p>Each student reads a text or watches a video explaining that the higher melting point of pure metals gives them greater bond strength and higher elastic modulus than the metals with lower melting point temperatures.</p>	<p>Each student underlines the text sentences explaining that the higher melting point of pure metals gives them greater bond strength and higher elastic modulus; or each student flexes 3 rods of 3 different metals to feel the stiffness of each rod with the associated melting points.</p>	<p>Each student <i>flexes</i> 3 rods and after finding that the stiffest rod has the highest melting point, the student is then asked to <i>draw</i> a microscopic model of the stiffer metal, showing a matrix array of small spheres (atoms) connected to each other by thick, strong, stiff springs.</p>	<p>Pairs of students can flex rods and discuss to agree on selecting the stiffest rod with the highest melting point, then the two or more students collaboratively construct the microscopic representation of the stiffer metal by building on, or challenging each other's contributions.</p>
<p>Cognitive processes</p> <p>Storing new information directly, without assimilating it with relevant knowledge</p>	<p>Activate/retrieve search existing knowledge</p> <p>Strengthen knowledge</p> <p>Encode or assimilate new information</p>	<p>Create & infer new knowledge</p> <p>Integrate newly created knowledge with old knowledge</p> <p>Re-organize knowledge</p> <p>Repair or accommodate old knowledge</p>	<p>Co-construct new knowledge that is novel to both partners</p> <p>Build on each other's knowledge</p> <p>Resolve own conflicts based on partner's comment</p>

situations, asking comprehension questions, solving a problem that requires constructing knowledge, justifying claims with evidence, designing a study, posing a research question, generating examples from daily lives, using analogy to describe certain cases, monitoring one's comprehension, making strategic decisions in a video game, converting text-based information into symbolic notation, drawing and interpreting graphs, or hypothesizing and testing an idea.

A constructive version of the metal rod activity described earlier could be offered if after flexing the rods of three different metals and finding the stiffest rod with the highest melting point, students then represented that macroscopic property by drawing a microscopic model of the stiffer metal, showing a small matrix array of small spheres (atoms) connected to each other by thick, strong, stiff springs. Thus, from bending the metal rods, students will recognize the stiffest rod and relate that to the normative bonding structure. From that, they can then construct a microscopic representation of the metal. When students provide information beyond what was observed, they create an explanatory model that is constructive.

The cognitive processes hypothesized to accompany constructive activities can generate new ideas, insights, and conclusions in a way that allows learners not only to infer new knowledge but also to repair or improve their existing knowledge. Repairing one's existing knowledge makes it more coherent, more accurate, and better structured, which serves to deepen one's understanding of new information. A variety of constructive activities, such as

self-explaining (Chi, 2000) and explaining-to-others (Roscoe & Chi, 2007) can improve learning (Chi, 2009).

Being Interactive

The interactive mode refers to two or more learners undertaking activities that develop knowledge and understanding extending beyond the materials being studied (similar to the constructive mode), but the interaction of the learners further enables them to build upon one another's understanding. The main (but surface-level) difference between the interactive and constructive mode is that learners in the latter engage in activities alone. However, interaction between learners affords them the benefit of receiving feedback or prompting from each other, with each partner having some complementary knowledge or perspectives. The different knowledge and perspectives further provide the opportunity for co-creation or joint-construction, which is not possible in solo activities.

Examples of interactive activities are studying or working in pairs or groups; reciprocal teaching; interacting with feedback from a teacher, an expert, or a computer agent; or arguing or defending one's position with evidence. Interactive conditions bring about co-construction of knowledge between pairs or group members. Chi (2009) cautions, however, that it is inappropriate to classify any group work as interactive. For example, if one group member dominates the discussion or if one member does not contribute to the discussion or product, then the group is not fully interacting. Therefore, the quality of discourse among members is critical for determining the degree of an activity's interactivity. When group members challenge each other by using normative scientific evidence, high-quality interaction results, thus, leading to enhanced learning for all members (see, e.g., Clark, D'Angelo, & Menekse, 2009; Sampson & Clark, 2009). The effectiveness of interactive activities depends on factors such as the degree of interactivity, the degree of student knowledge construction, and student willingness to challenge or criticize each other (Hausmann, 2006).

An interactive version of the metal rod activity could be two students working together on the activity of flexing rods, questioning each other about the rationale for selecting the stiffest rod with the highest melting point, and discussing how they should draw the microscopic representation of the stiffest metal. Through this give-and-take discussion, students would be building knowledge in a way that would not have occurred if they had been working alone, since they can build on each other's contributions or refine and modify an original idea in ways that can produce novel ideas. Thus, interactive learning has the potential to be more beneficial than constructive learning, in which single individuals can only extend beyond the given information with their own ideas; in interactive learning, two individuals can further enrich the topic of discussion through jointly extending on a given content topic from two different perspectives and sets of ideas.

Research Goals

This study evaluated the effectiveness and applicability of the DOLA framework and tested the ICAP hypothesis in a materials science and engineering context. We conducted two studies to find out how and to what degree differentiated activities affect undergraduate engineering students' learning outcomes. Whereas Study 1 was conducted in an engineering classroom during the actual class sessions, Study 2 was conducted in a more controlled environment. In both studies, we measured students' cognitive learning outcomes rather than

affective and psychomotor aspects of their learning. We used introductory materials science and engineering concepts as content to be taught, since these concepts are rich and difficult, with a fundamental goal of bridging nano-scale structural features (i.e., electronic structure, atomic bonding, lattice parameters, and grain size) to macro-scale properties (i.e., stiffness, strength, deformation, and functional properties).

In particular, this article has two main goals. First, since this is the first research project to specifically evaluate the DOLA framework and test the ICAP hypothesis, we aim to investigate the strengths and possible limitations of this framework by testing it in a complex engineering context. Second, we aim to contribute to the recent curriculum development efforts in engineering education by not only designing and implementing activities in engineering classes but also testing and refining them by using actual empirical data.

In Study 1, we evaluated the DOLA framework in a naturalistic setting, an engineering classroom, by testing the ICAP hypothesis. Because this was a real engineering classroom, we did not include a passive condition. In Study 2, we investigated the DOLA framework in a more controlled environment by recruiting engineering students for a laboratory study. This study tested the ICAP hypothesis by using all four conditions. For both studies, we created or modified classroom activities and developed and tested assessment items.

The rationale behind our decision to conduct both classroom and laboratory studies was to obtain the most complete picture of the extent of the DOLA framework. One can argue that the data from classroom studies are the most reliable sources for educational studies because of the natural interactions and settings in classrooms. However, laboratory studies also provide valuable data since the controlled settings allow us to isolate the variables in order to understand the details of the obtained results. The current research was designed not to make comparisons between the results from instructional settings and experimental settings but, instead, to understand the value and effectiveness of the DOLA framework on students' learning outcomes from the collective data of both classroom and laboratory settings.

Study 1: Classroom Study

Participants

The sample for Study 1 consisted of 42 undergraduate engineering students enrolled in an introductory materials science and engineering course in a large public university located in the southwestern United States. Thirty-five students were male, and seven were female. The mean age was 19 with a range from 18 to 21 years. Each student had already completed a college-level general chemistry class as a prerequisite. Participation in the project was voluntary, and students were assured that their participation would have no effect on their grades. Data collection was completed on five different days during the first three weeks of the semester. In addition to regular class hours, participants were asked to stay for 15 to 20 minutes after the regular class hours during these five days. Student received \$5 per day for their participation.

Development of the In-Class Activities

One of the researchers attended the introductory materials science and engineering course for a semester prior to the study to document all learning activities already used in the classroom. We gathered instructional materials (i.e., slides and handouts) and assessment measures (i.e., concept tests, unit tests, and homework assignments) that were used for

each class period. In preparation for our study, we classified the 19 overt activities that were used as active, constructive, or interactive on the basis of Chi's (2009) framework.

We selected two units, atomic bonding and crystal structures, to be used for this study. After negotiating with the faculty, we agreed on the mode of activities that would be offered within each of these two units. We used only one type of activity per class period, regardless of how many activities were offered, so that we could test for learning that could be attributed to that one particular type of activity. For example, if a constructive activity was implemented on a specific day, no active or interactive modes of activities were used on that day.

Modifications were made to some of the existing class activities to allow testing of the ICAP hypothesis. For example, in previous semesters, students learned about features of a face-centered cubic structure by means of a constructive activity in which they used given indices to draw unit cell directions on a worksheet. They also used a set of directions to determine the indices of unit cells. We modified this activity to be active by having the instructor demonstrate both processes and instructing students to simply copy the instructor's work. This activity then met an active mode criterion of having students manipulate the information in some way, without generating new information from it.

We planned the activity modes so that a contrast could be made between active and interactive activities in the atomic bonding unit, and between active, constructive, and interactive activities in the crystal structures unit. For example, on Day 1 for the atomic bonding unit, students participated in an active activity, consisting of selecting the most likely material, property of that material, type of bonding, and processing method from a given list of motorcycle parts, such as a fender or seat. On Day 2, they participated in an interactive activity, consisting of drawing and completing a partially constructed concept map for bonding and explaining their reasoning for each decision they made to complete the concept map. The final study design included three active, two constructive, and three interactive activities for the two units. There was one activity per day for the atomic bonding unit and two activities per day for the crystal structures unit. Table 2 presents the mode, task name, and order of activities used for both units. (See Appendix A for a detailed description of each activity.)

To promote high-quality productive interaction between students during the interactive activities, we devised written guidelines to help group leaders facilitate discussion. The guidelines included detailed directions for the task, timelines for completion of the activity, and ideas for probing questions that could stimulate knowledge construction by group members.

Measures

Student learning from in-class activities was measured immediately after each class period in which the two topics were studied. We chose to measure student learning at this time in an attempt to differentiate between learning that may have resulted from the in-class activities and learning that may have resulted from homework or alternate learning strategies that students employed outside of class. Daily quiz questions for each activity were generated in order to measure student learning and comprehension of the content covered in the activities. Because the content and activities were different each day, we measured knowledge gained from them on a common metric in order for the ANOVA significance tests to be meaningful. Due to our interest in examining the depth of processing and resultant knowledge associated with each activity, we chose "cognitive level of quiz questions" as the common metric (this metric will be defined below).

Table 2 Names, Type, and Order of Activities Used

	Atomic bonding		Crystal structures		
	Day 1	Day 2	Day 3	Day 4	Day 5
Active	Materials selection			Unit cell directions Unit cell families of directions	
Constructive					Unit cell planes Unit cell worksheet
Interactive		Bonding concept map	Concepts in context Hidden treasurers		

(See Appendix A for the detailed descriptions of each activity)

There were 16 two-tiered quiz questions, two for each activity. The first tier questions were in a multiple-choice format, consisting of one verbatim question and one integration question for each activity. The second tier of each of the 16 questions was in an open-ended format, consisting of two inference questions for each activity. Counting each part as a question, there were in total 32 questions.

The three categories of questions represent different levels of cognitive activity required to respond to the question; these categories were also considered to be indicative of question difficulty (Chi et al., 1994). Here we will present examples to illustrate each type of question. Verbatim questions were generated from ideas and information explicitly stated in the activity. They required students to merely recall the correct responses and thus required the shallowest understanding for the student to answer them correctly. For example, to correctly answer the verbatim question in the “concepts in context” activity, students had to select a disaster or failure that occurred as a result of an incomplete phase transformation; this information was explicitly stated in the activity. The integration questions were also generated from the ideas and information explicitly stated in the activity, but they required students to integrate two or more ideas from the activity; thus they required a slightly deeper level of understanding. For instance, to correctly answer the integration question from the activity mentioned above, students needed to integrate the ideas of the most likely condition for phase change, properties of materials, and unit cell transformation. These three ideas are explicitly covered in the activity, but they need to be integrated in order to answer the question. Finally, the inference questions required students to generate ideas beyond the information presented in the activity; thus they required the deepest understanding. For example, one of the inference questions for the “concepts in context” activity asked students to specify recommendations to prevent disaster or failure on the basis of the relationship between a component material’s macroscopic properties and its atomic-level structure. The activity itself did not include any discussion about recommendations to prevent disasters or failures, so this question required students to think about these recommendations like a consulting engineer giving advice about failure prevention to a company. Accordingly, our question

categories had an ordinal relationship in which inference questions were considered to be more difficult than the integration questions, which, in turn, were considered to be more difficult than the verbatim questions. Table 3 summarizes the characteristics and associated cognitive processes for the question types used in daily quizzes.

Evidence for content validity was obtained by having an expert from the materials science and engineering department as well as an expert in measurement and test development. These experts provided continuous feedback and suggestions for improvement during question development, until they approved the final version of each question. The quiz questions were closely aligned with the content covered in each activity, thus ensuring representative sampling of content in the assessment of student learning. (See Appendix B for an activity sample and Appendix C for a quiz sample.)

Procedure

The data was collected over five days. The class topic was atomic bonding during the first two days and crystal structures during the last three days. Students completed one activity per day during the atomic bonding units and two activities per day during the crystal structures unit. The activities each took approximately 15 minutes of class time. On Day 1, each student was given an activity worksheet and individually completed the active version of the Materials Selection activity. The instructor told students to work alone and not to interact with each other. After the regular class hour, participants stayed in the classroom and completed the daily quiz questions individually, which took 10 minutes. The students were not allowed to use any instructional materials to answer the quiz questions.

On Day 2, students completed the interactive version of the Bonding Concept Map activity in small groups (approximately five students in each group). One activity worksheet was provided for each of the nine groups in the class. The students were encouraged to question each other's reasoning and reach a group consensus for their final answers before recording their responses on their group worksheet. As for Day 1, participating students stayed after class and took the daily quiz questions individually.

On Day 3, students completed the interactive versions of the Concepts in Context and Hidden Treasures activities in small groups. As for the Bonding Concept Map activity, they were encouraged to question each other and reach a consensus as a group before recording their responses on their group worksheets. After the regular class period, participants stayed in the classroom and took two daily quizzes (one for each activity) individually, which took a total of 20 minutes.

On Day 4, students completed the active versions of the Unit Cell Directions and the Unit Cell Families of Directions activities. Each student copied the activity answers given

Table 3 Characteristics and Associated Cognitive Processes for the Question Types Used in Daily Quizzes

	Verbatim	Integration	Inference
Characteristics			
Single idea or concept		Explicit information	Multiple ideas or concepts
Explicit information		Multiple ideas or concepts	Implicit information
Cognitive processes			
Recall		Integrate	Construct

by the faculty onto their worksheet for each activity and selected the Miller indices of unit cells from a given set of unit cell directions. After the regular class period, the participants again took two daily quizzes individually.

On Day 5, students individually completed the constructive version of the Unit Cell Planes and Unit Cell Worksheet activities during the regular class hour. Again, each student received a worksheet for each activity and was asked to draw unit cell planes by using the Miller indices and to determine the Miller indices of unit cells for various unit cell planes. They stayed after class and took two daily quizzes individually. During these five days, the instructor provided no feedback or instructional support to students about the subject matter regarding activities.

Scoring the Daily Quizzes

The multiple-choice questions were dichotomously scored as correct or incorrect. The correct responses received three points, and incorrect ones no points. Unanswered questions were scored as incorrect. We developed a rubric to score student responses to open-ended questions. (See Appendix D for a rubric sample.) The open-ended responses were scored as fully correct, partially correct, or incorrect. Fully correct responses were complete and contained no inaccurate explanations, ideas, or solutions. Partially correct responses contained some correct explanations, ideas, or solutions, but were not complete. Fully correct responses received four points; partially correct responses received one, two, or three points based on the rubric criteria; and incorrect responses received no points. Incorrect responses contained erroneous or inconsistent explanations. Student open-ended responses were not penalized for spelling or grammar errors. The maximum score was 14 for each daily quiz of two questions.

Twenty-five percent of the daily quizzes were scored individually by two raters to calculate inter-rater reliability. The percentage agreement between them was 94% for the open-ended questions. The rest of the quizzes were scored by one of the raters.

Results for Study 1

To evaluate the effectiveness of differentiated overt learning activities on student learning, we conducted a one-way repeated-measure analysis of variance (ANOVA). The within-subject factor was type of activity, and the dependent variable was student total scores on daily quiz questions. The means and standard deviations for students' scores are presented in Table 4, and effect sizes (using Cohen's *d* formula) for pairwise comparisons are presented in Table 5.

Because the topics, atomic bonding and crystal structures, have different characteristics and difficulty levels, it is not meaningful to directly compare the effectiveness of activities across topics. Therefore, we compared student achievement scores within each topic across

Table 4 Means and Standard Deviations for Student Scores for Activities by Topic

Type of Activity	Atomic bonding		Crystal structures	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Active	6.69	3.19	7.74	2.76
Constructive	na	na	9.10	1.52
Interactive	9.67	2.49	9.03	2.67

Table 5 Effect Sizes by Using Cohen's *d* Formula for Pairwise Comparisons by Topic

	Atomic bonding	Crystal structures
Interactive vs. active	1.04	.48
Interactive vs. constructive	na	.05
Constructive vs. active	na	.63

Table 6 Means and Standard Deviations of Student Scores by Question Types for Atomic Bonding

Type of Activity	Verbatim		Integration		Inference	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Active	2.31	1.21	1.92	1.14	2.46	1.58
Interactive	2.85	.66	2.54	1.08	4.13	1.77

different activities. Accordingly, the analysis compared active and interactive activities for the atomic bonding unit and active, constructive, and interactive activities for the crystal structures unit.

Results for atomic bonding A one-way repeated-measures ANOVA was conducted. The factor was the type of activity (active, interactive), and the dependent variable was the student achievement scores on the daily quiz questions corresponding to each atomic bonding activity. The ANOVA results indicated a significant effect of activity type, Wilks' $\Lambda = .57$, $F(1, 38) = 28.69$, $p < .01$, multivariate $\eta^2 = .43$. These results suggested that students learned significantly more from interactive activities than from active ones. Table 4 gives the means and standard deviations across type of activities for the atomic bonding unit.

We were also interested in determining how students performed on the basis of the type of question. Table 6 gives the means and standard deviations for verbatim, integration, and inference questions within each activity. Three separate one-way repeated measures ANOVAs comparing each question type across two atomic bonding activities showed significant effects of activity mode for all questions types. For verbatim questions, Wilks' $\Lambda = .77$, $F(1, 38) = 11.40$, $p < .01$, multivariate $\eta^2 = .23$; for integration questions, Wilks' $\Lambda = .86$, $F(1, 38) = 6.02$, $p < .05$, multivariate $\eta^2 = .14$; and for inference questions, Wilks' $\Lambda = .62$, $F(1, 38) = 23.57$, $p < .01$, multivariate $\eta^2 = .38$. Overall, students performed significantly better on all types of question when undertaking the interactive activity than when undertaking the active activity.

Results for crystal structures To evaluate the effect of the activity types on student daily quiz question scores for the crystal structures unit, we initially conducted a one-way repeated measures ANOVA with the activity type as a factorial variable, and student total scores as dependent variables. The results showed a significant main effect for the activity type on learning, Wilks' $\Lambda = .76$, $F(2, 34) = 5.40$, $p < .01$, multivariate $\eta^2 = .24$. Table 4 gives the means and standard deviations across type of activities for the crystal structures unit.

Next, three unique pairwise comparisons were conducted between the means of student scores for active, constructive, and interactive activities by using post-hoc Tukey's test at

Table 7 Means and Standard Deviations of Student Scores by Question Types for Crystal Structures

Type of Activity	Verbatim		Integration		Inference	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Active	2.47	.94	1.88	1.08	3.40	1.74
Constructive	2.88	.64	2.75	.69	3.47	1.57
Interactive	2.54	1.08	2.13	.76	4.36	1.81

$p < .05$. Two of the three pairwise comparisons were significant. The means for active and constructive activities were significantly different. The difference between the active and interactive means was also significant. The difference between constructive and interactive means was not significant.

As we did for the atomic bonding activities, three separate one-way repeated measures ANOVAs were conducted to compare each question type across three crystal structures activities. The results were significant for all three one-way repeated measures ANOVAs. For verbatim questions, Wilks' $\Lambda = .66$, $F(2, 34) = 8.65$, $p < .01$, multivariate $\eta^2 = .34$; for integration questions, Wilks' $\Lambda = .67$, $F(2, 34) = 8.55$, $p < .01$, multivariate $\eta^2 = .34$; and for inference questions, Wilks' $\Lambda = .76$, $F(2, 34) = 5.28$, $p < .01$, multivariate $\eta^2 = .24$. Table 7 gives the means and standard deviations for verbatim, integration, and inference questions within each activity.

Pairwise comparisons were also conducted to determine how activity type affected student scores on the question types. We conducted three unique pairwise comparisons for each question type using post-hoc Tukey's test at $p < .05$. For the verbatim questions, one of the three pairwise comparisons was significant: constructive was significantly better than active. For the integration questions, two of the three pairwise comparisons were significant: constructive was significantly better than active, and constructive was significantly better than interactive. Finally, for the inference questions, two of the three pairwise comparisons were significant: interactive was significantly better than active, and interactive was significantly better than constructive.

Summary and Discussion for Study 1

In Study 1, we compared the effects of three activity types for two topic areas in an introductory materials science and engineering course. For the atomic bonding unit, the results showed that student scores on postclass quizzes were significantly better after engaging in interactive activities compared with active activities. For the crystal structures unit, students did significantly better after both the interactive and constructive activities than with active activities. Although there were no significant differences between constructive and interactive activities in terms of total scores, when we analyzed student performance by question type, student inference scores (indicative of the deepest learning and understanding) for the interactive activities were significantly higher than their inference scores in both constructive and active activities for both units. Thus, the comparison of scores from inference questions revealed that after engaging in interactive activities, students were better able to respond to more difficult, challenging questions about their engineering course material.

When we examined overall scores, the results of Study 1 provided preliminary evidence to support Chi's (2009) hypothesis that both interactive and constructive activities provide greater returns in terms of student learning than do active activities. Those interactive activities provide

greater returns than either constructive or active activities for inference questions. However, this classroom study had one counter-intuitive result. When we analyzed performance per question type, students performed better on integration questions after engaging in constructive activities than after engaging in interactive activities. This result should not be the case because, overall, interactive activities should involve construction of new knowledge with the enhancement of contributions from one's peers. However, this finding could be influenced by the nature of the integration questions because they ask for information that is explicitly presented in the instructional materials. Thus, additional benefits may not be derived from having complementary knowledge or an alternative perspective from a partner, in order to answer these shallower questions. Additional perspectives, feedback, and elaborations from a partner may, however, be particularly useful for answering the more challenging and deeper inference questions, as our results show. In general, Chi and Wylie (under review) predict that the ICAP hypothesis may not hold for shallow questions, and that differences between instructional conditions are more likely to become apparent when assessing deep understanding. Since deep questions typically require students to integrate new information with prior knowledge, they are more sensitive for detecting differences between instructional interventions.

This classroom study has several limitations. Since this study was conducted in a real classroom, it was difficult to control for confounding factors like the level of student interaction and time spent to complete tests and activities. Moreover, there were no pretests to measure student prior knowledge. There was also no passive condition, which can be thought of as a control condition that often reflects standard instruction in college classrooms. Another limitation was that we used a within-subject design in which the same students were engaged in different activities on different days and completed tests with different questions. This arrangement possibly introduced order effects from the activities, which may have influenced the degree of learning students experienced. Students may acquire more knowledge from the later activities due to the cumulative effect of several days of class activities. For instance, in the crystal structures unit, students participated in the interactive activities on Day 3, the active activities on Day 4, and the constructive activities on Day 5; so if students benefited from the cumulative effect of several days of class activities, they might have more prior knowledge for the constructive activities than for the interactive and active activities. Finally, the different activities may have inherently different requirements, some requiring conceptual understanding and others requiring procedural understanding.

Study 2: Laboratory Study

On the basis of the limitations just cited for Study 1, we designed Study 2 with a more controlled between-subjects design in which the participants were randomly assigned into one of four experimental conditions: passive, active, constructive, or interactive. All students took pre- and posttests in this study. As in Study 1, we used introductory materials science and engineering concepts to create our activities for the four conditions in Study 2. We constructed the Connecting Atomic Bonding and Physical Properties activity that required students to understand the relations between bonding energy, elastic modulus, melting points, and coefficient of thermal expansion concepts.

Participants

The sample for Study 2 was 120 undergraduate engineering students in a large public university located in the southwestern United States. Seventy-two participants were male, and

48 were female. The mean age of the participants was 20 with a range from 18 to 23 years. The participants were recruited through announcements by posters and flyers across campus and e-mails sent to engineering instructors and department secretaries. A prerequisite was that participants had completed a college-level general chemistry class with a grade of B or better, so that they were familiar with the terminology used in the activities.

Materials

Introductory text A two-page introductory text, consisting of definitions and short descriptions for concepts, was created for this study. The text's content was based on materials science and engineering textbooks that are commonly used in U.S. universities and colleges written by William D. Callister (2006) and James Newell (2009). We used definitions of terms such as chemical bonding, bond energy, bond strength, and tensile properties taken directly from these texts. All participants read the introductory text to get familiar with (or as a reminder of) the terminology used before taking the pretest.

Long text The same textbooks cited above were used to create a longer eight-page text to deliver content. This text described bonding energy, elastic modulus, melting points, and coefficient of thermal expansion and included examples of each concept. We selected the relevant passages that provided fundamental definitions and descriptions for each concept, and we created related examples for each concept. This text's content was conceptual; we avoided using complex mathematical representations or statements, and there were no questions embedded in the reading. The text was in 12-point font, double spaced, and formatted in one column.

Graphs, figures, and activity sheet We created two graphs, two figures, and an activity sheet based on the information and examples given in the long text. The graphs and figures illustrated the properties of three metals in terms of elastic modulus, bond energy, thermal expansion, and melting points. For example, the long text defined the elastic modulus of an object as the slope of its stress-strain graph and included an example of three metals with different slopes. So we created the stress-strain graph of the same example by using the three metals that were described in the long text. Figure 1 is the constructed graph of three metals and the corresponding text paragraph that described the elastic modulus concept.

We also constructed an activity sheet with five short-answer questions to scaffold and guide students to interpret specific aspects of the information provided in the graphs and figures. For example, one question asked students to compare values for the three metals in each graph and figure and to explain their findings for each comparison. Students were asked to write their responses on the worksheets. Taken together, the graphs and figures accompanied by the activity sheet provided a guided inquiry-oriented activity in which the data or information embedded within the graphs and figures followed by question prompts supported students toward construction of their own reasoning and conclusions.

Measures

We used a pretest-posttest design to measure student prior knowledge and learning from the activities intervention. The pretest consisted of 15 true-false, seven multiple-choice and two open-ended questions, which totaled 24 questions. The true-false questions were two-tiered in that the first part required students to determine the correctness of a given statement, and the second part required students to explain their selection. The multiple-choice questions had five responses that included one correct answer and four distractors. The open-ended questions required short answers. The posttest consisted of the same 24 questions along

The degree to which a structure deforms or strains depends on the magnitude of an imposed stress. For most metals that are stressed in tension and at relatively low levels, stress and strain are proportional to each other through the relationship $E = \sigma/\epsilon$ where E is the elastic modulus, σ (sigma) represents stress, and ϵ (epsilon) represents strain. For example, assume we have three metals: metal A, metal B and metal C. The metal A has the greatest elastic modulus among all three and the metal B has greater elastic modulus than metal C. This relationship also implies that the metal A has the greatest slope in a stress-strain curve and the metal C has the smallest slope in the same curve.

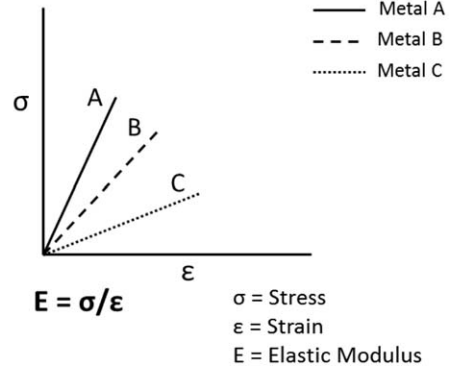


Figure 1 Exemplary text scrap (left) and graph for the elastic modulus concept (right).

with six additional questions. Overall, the posttest consisted of 16 true-false, 11 multiple-choice, and three short-answer open-ended questions.

The questions were closely aligned with the content covered in the learning materials, thus ensuring representative sampling of content in the assessment of student learning. The reliability calculations based on the pretest revealed that Cronbach's alpha was .81, which indicated good internal consistency.

Design and Procedure

The study consisted of four intervention conditions corresponding to the DOLA framework. Content about the materials science concepts was presented in a different manner in each of the conditions. In the passive condition, students read the long text passage out loud; they were not allowed to use highlighters or pens while reading the text. In the active condition, students read the same long text as described above; they were given highlighters and instructed to highlight the most important or critical sentences in the text. In the constructive condition, students completed the graphs and figures interpretation activity but did not read the long text that was used in the passive and active conditions. Instead, they were instructed to study the graphs and figures and provide written responses to the questions on the activity sheet. In the interactive condition, pairs of students completed the same graphs and figures interpretation activity (as in the constructive condition), also without reading the long text. Student pairs in the interactive condition shared one activity sheet and completed it collaboratively. At the beginning of the activity, pairs were told to reach consensus for each question before writing their answers on the activity sheet. The pairs were videotaped to record their dialogues. (These dialogues will be analyzed and reported in another paper.) For all conditions, no content-related feedback was provided during any of the sessions across any condition.

The recruited participants were randomly assigned to a condition and were scheduled to a particular session on the basis of their availability. All participants started by reading the two-page introductory text for 10 minutes. Participants were then given 25 minutes to complete the pretest. Twenty-four students were in the passive, active, and constructive conditions, and 48 students (24 pairs) in the interactive condition. Students were given 25 to

30 minutes to complete their learning activity. Finally, all participants took the posttest individually. Each session lasted approximately 90 minutes. Data collection was completed in one session. Participants received \$15 after they completed the study.

Scoring the Tests

The multiple-choice questions were scored as correct or incorrect. Correct responses received two points; incorrect responses or blank answers received no points. The first-tier true-false questions were scored as correct or incorrect, and the second-tier open-ended questions were scored as fully correct, partially correct, or incorrect. A rubric was developed for scoring these open-ended questions. Fully correct responses were complete and contained no inaccurate explanations, ideas, or solutions; partially correct responses contained some correct explanations, ideas, or solutions, but were incomplete; and incorrect responses contained erroneous or inconsistent explanations. Student received three points for a correct true-false response and a fully correct explanation; two points for a correct true-false response and a partially correct explanation; one point for a correct true-false response and an incorrect explanation or no explanation; and no points for an incorrect true-false response.

The same rubric was used to score the additional open-ended questions as fully correct (three points), partially correct (one or two points), or incorrect (no points). The maximum scores were 65 for the pretest and 79 for the posttest.

Thirty percent of the pre- and posttests were scored individually by two raters to calculate the inter-rater reliability. The percentage agreement between the two raters was 96% for the second tier of the true-false questions, and 93% for the open-ended questions. The rest of the pre- and posttests were scored by one of the raters.

Results

First, we evaluated the randomness of participant assignment into conditions by conducting a one-way ANOVA to assess whether there were differences between student pretest scores across conditions. The results indicated no significant difference. Table 8 gives student pretest and posttest scores for all conditions.

On the basis of the null result from pretest scores, we used student gain scores from pre- to posttest to evaluate the relationship between the conditions and student gain scores. We conducted a one-way ANOVA in which the within-subject factor was condition, or activity type, and the dependent variable was percentage of student gain scores from pre- to posttest. We used percentage of gain scores instead of raw scores because of the six additional questions in the posttest. The results for the ANOVA indicated a significant effect of condition, $F(3, 116) = 25.34, p < .00$. The strength of the relationship

Table 8 Means and Standard Deviations of Student by Raw Scores and Percentages for Pretest and Posttest by Type of Activity

	Pretest Raw (Max = 65)			Posttest Raw (Max = 79)		
	<i>M</i>	<i>SD</i>	<i>M</i> (%)	<i>M</i>	<i>SD</i>	<i>M</i> (%)
Interactive	33.98	9.86	52.28	64.71	11.03	81.91
Constructive	34.54	9.82	53.14	57.63	13.54	72.94
Active	32.86	8.76	50.58	49.33	10.89	62.45
Passive	33.79	8.01	51.99	43.71	11.40	55.33

Table 9 Results for Pairwise Comparisons Using Holm's Sequential Bonferroni Method

Comparison ^a	Mean difference	<i>p</i> -value	Alpha-value
Interactive vs. passive	26.29	.000	.008
Interactive vs. active	17.76	.000	.01
Constructive vs. passive	16.46	.000	.013
Interactive vs. constructive	9.83	.003	.017
Active vs. passive	8.53	.023	.025
Constructive vs. active	7.93	.035	.05

^aAll comparisons are significant (*p*-value < alpha-value)

Table 10 Effect Sizes by Using Cohen's *d* for Pairwise Comparisons

Comparison	Cohen's <i>d</i>
Interactive vs. passive	1.88
Interactive vs. active	1.42
Interactive vs. constructive	.64
Constructive vs. passive	1.04
Constructive vs. active	.54
Active vs. passive	.65

Gain percentage scores were used for effect size calculations.

between the conditions to which students were assigned and their gain scores, as assessed by η^2 , was strong, with the condition factor accounting for 40% of the variance of the dependent variable. The means and standard deviations for student raw scores and percentage of pre- and posttests are presented in Table 8.

Follow-up tests were conducted to evaluate pairwise differences among the means of the four conditions. We used Holm's sequential Bonferroni method to control for Type I error at the .05 level across all six comparisons. All pairwise comparisons were significant. Table 9 gives mean differences, *p*-values, and alpha-values for all comparisons. The effect sizes (using Cohen's *d*) for each pairwise comparison by using gain percentage scores are presented in Table 10.

Summary and Discussion

Using an experimental design with random assignment, we compared four learning activity conditions using introductory materials science concepts to test the ICAP hypothesis. We found that student gain scores increased steadily from the passive to the active to the constructive and to the interactive conditions, as predicted by the ICAP hypothesis.

The study's learning materials were presented in the format of texts with definitions and examples for each concept in the passive and active conditions, and in the format of graphs and figures for the same concepts in the constructive and interactive conditions. Although the materials for the passive and active conditions clearly provided the normative information for each concept, students did significantly better in the constructive and the interactive conditions even without such information being directly stated, by

constructing their own knowledge and understanding from the guided inquiry-oriented activity sheet that contained only question prompts. The students in the interactive condition did significantly better than the students in the constructive condition since the students in the interactive condition had a chance to enrich their understanding socially by jointly constructing knowledge with a partner.

Conclusion

This article describes our work investigating the applicability of Chi's (2009) DOLA framework in an engineering context and to evaluate the ICAP hypothesis. In Study 1, we explored the value of this framework by using it to classify the existing learning activities already developed and used in a materials science and engineering classroom. We then modified the activities as needed in order to examine learning gains that matched with each type of activity. In Study 2, we used the framework to guide the development of activities in order to examine subsequent learning gains from each activity type. In both studies, we evaluated students' cognitive gains following their participation in different types of learning activities.

In Study 1, the analysis based on question type revealed that students' inference scores following interactive activities were significantly higher than their inference scores following constructive and active activities for both units. Since the inference questions were the most challenging questions, these results help us understand where the real differences appear between conditions and how these differences can be detected.

Our results from Study 1 provided support for the ICAP hypothesis when the DOLA framework was used in a natural setting, despite confounding factors that were present such as differences in level of student interaction in the interactive activities, differences in time on task when completing the various types of learning activities, and possible order effects. In Study 2, we reduced these confounds significantly, which allowed us to compare all four conditions in a more controlled environment with a larger sample size. The results for Study 2 provided strong support for the ICAP hypothesis in which interactive activities are expected to enhance learning better than do constructive activities. In joint dialoguing and co-construction, not only is each student generative but each student can further benefit from feedback, scaffolding, and contributions from the partner. Constructive activities are expected to enhance learning better than do active activities because constructive activities allow students to generate new knowledge and repair old knowledge. Finally, active activities are expected to enhance learning better than do passive activities because actively emphasizing a part of the learning materials allows the learner to attend to and activate relevant knowledge, thereby allowing the learner to assimilate novel information to fill knowledge gaps, whereas passive activities may only store novel information infrequently.

In both studies, the pivotal concepts in materials science and engineering require students to understand a material's macroscale properties on the basis of knowledge and understanding of a material's structures from levels of nano- to micro- to macroscale. Acquiring this understanding is a significant intellectual challenge and requires students to undertake complex cognitive processes such as decision making, spatial reasoning, knowledge construction, and integration. We argue that the principles of the DOLA framework and the results that support the ICAP hypothesis provide a comprehensive methodology to create and design materials and activities that will promote effective learning in engineering classrooms.

Although our study supports the utility of the DOLA framework and ICAP hypothesis, it must be noted that our work evaluated only short-term gains. Both Study 1 and Study 2 lack long-term retention data (i.e., a delayed posttest), which could provide further evidence to support the benefit of constructive and interactive activities as a means to increase student cognitive gains. Future work should investigate long-term gains or retention of material following different types of learning activities.

In conclusion, our studies support Chi's ICAP hypothesis, whose classification of overt learning activities can help researchers, instructional designers, and instructors determine activities appropriate for their intended research or instruction. Our results suggest that when implemented properly, interactive modes are most effective, constructive modes are better than active and passive modes, and active modes are better than passive ones for student learning.

Acknowledgments

We are grateful for support provided by the National Science Foundation grant 0935235 to Michelene T. H. Chi and Stephen Krause, the Institute of Education Sciences (grant 943360412) to Michelene T. H. Chi, and the Arizona State University Graduate and Professional Student Association to Muhsin Menekse. We also appreciate the comments and suggestions provided by Professors Dale Rose Baker and James Middleton.

References

- Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A., & Granger, E. M. (2010). Is inquiry possible in light of accountability?: A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science Education, 94*, 577–616.
- Burleson, W., Ganz, A., & Harris, I. (2001). Educational innovations in multimedia systems. *Journal of Engineering Education, 90*(1), 21–31.
- Callister, W. D. (2006). *Materials science and engineering: An introduction* (7th ed.) New York, NY: Wiley.
- Chen, H., Lattuca, L., & Hamilton, E. (2008). Conceptualizing engagement: Contributions of faculty to student engagement in engineering. *Journal of Engineering Education, 97*(3), 339–353.
- Chi, M. T. H. (2000). Self-explaining expository texts: The dual process of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah, NJ: Erlbaum.
- Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73–105.
- Chi, M. T. H., & Wylie, R. (under review). Differentiating four levels of engagement in learning: The ICAP hypothesis.
- Chi, M. T. H., de Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439–477.
- Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145–182.
- Clark, D. B., D'Angelo, C. M., & Menekse, M. (2009). Initial structuring of online discussions to improve learning and argumentation: Incorporating students' own

- explanations as seed comments versus an augmented-preset approach to seeding discussions. *Journal of Science Education and Technology*, 18(4), 321–333.
- Colliver, J. A. (2000). Effectiveness of problem-based learning curricula: Research and theory. *Academic Medicine*, 75(3), 259–266.
- Hausmann, R. G. M. (2006). Why do elaborative dialogs lead to effective problem solving and deep learning? In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Cognitive Science Society* (pp. 1465–1469). Alpha, NJ: Sheridan.
- Heller, R. S., Beil, C., Dam, K., & Haerum, B. (2010). Student and faculty perceptions of engagement in engineering. *Journal of Engineering Education*, 99(3), 253–261.
- Higley, K., & Marianno, C. (2001). Making engineering education fun. *Journal of Engineering Education*, 90(1), 105–107.
- Hundley, S. A. (2007). *A comparative study of traditional lecture methods and interactive lecture methods in introductory geology courses for non-science majors at the college level* (Doctoral dissertation). The Ohio State University, Columbus, OH. Retrieved from ProQuest. (UMI No. 3286820)
- Johnson, D. W., Johnson, R. T., & Smith, K. (1991) *Active learning: Cooperation in the college classroom*, Edina, MN: Interaction.
- Kirschner, P., Sweller, J., & Clark, R. (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(2), 75–86.
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction. *Psychological Science*, 15(10), 661–667.
- Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, 20(2), 403–437.
- Lambert, N., & McCombs, B. (1998). *How students learn: Reforming schools through learner-centered education*. Washington, DC: American Psychological Association.
- Lederman, J. S., Lederman, N. G., & Wickman, P.-O. (2008). *An international, systematic investigation of the relative effects of inquiry and direct instruction: A replication study*. Paper presented at the annual conference of the National Association for Research in Science Teaching, Baltimore, MD.
- Lederman, N., Lederman, J., Wickman, P. O., & Lager-Nyqvist, L. (2007). *Development of a valid and reliable protocol for the assessment of early childhood students' conceptions of nature of science and scientific inquiry*. Paper presented at the annual conference of the National Association for Research in Science Teaching, New Orleans, LA.
- Lin, C., & Tsai, C. (2009). The relationship between students' conceptions of learning engineering and their preferences for classroom and laboratory learning environments. *Journal of Engineering Education*, 98, 193–204.
- Lott, G. (1983). The effect of inquiry teaching and advance organizers upon student outcomes in science education. *Journal of Research in Science Teaching*, 20(5), 437–451.
- Martin, M. (2009). *The effect of active techniques combined with didactic lecture on student achievement* (Doctoral dissertation). Arkansas State University, Jonesboro, AR. Retrieved from ProQuest. (UMI No. 1467105)
- Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, 59(1), 14–19.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.

- Minner, D., Levy, A., & Century, J. (2010). Inquiry based science instruction – What is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47(4), 474–496.
- Montpas, M. M. (2004). *Comparison of “jeopardy” game versus lecture on associate degree nursing students’ achievement and retention of geriatric nursing concepts* (Doctoral dissertation). Wayne State University, Detroit, MI. Retrieved from ProQuest. (UMI No. 3151328)
- National Research Council. (1996). *National Science Education Standards*. Washington, DC: National Academy Press.
- Newell, J. (2009). *Essentials of modern materials science and engineering*. Hoboken, NJ: Wiley.
- Osman, M. (2008a). Observation can be as effective as action in problem solving. *Cognitive Science*, 32, 162–183.
- Osman, M. (2008b). Positive transfer and negative transfer/antilearning of problem-solving skills. *Journal of Experimental Psychology: General*, 137, 97–115.
- Pendergrass, N., Kowalczyk, R., Dowd, J., Laoulache, R., Nelles, W., Golen, J., & Fowler, E. (2001). Improving first-year engineering education. *Journal of Engineering Education*, 90(1), 33–41.
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, 93, 223–232.
- Prince, M., & Felder, R. (2006). Inductive teaching and learning methods: Definitions, comparisons, and research bases. *Journal of Engineering Education*, 95(2), 123–137.
- Purzer, S. (2011). The relationship between team discourse, self-efficacy, and individual achievement: A sequential mixed-methods study. *Journal of Engineering Education*, 100(4), 655–679.
- Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, 77, 1–15.
- Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors’ explanations and questions. *Review of Educational Research*, 77, 534–574.
- Sadler, K. C. (2002). *The effectiveness of cooperative learning as an instructional strategy to increase biological literacy and academic achievement in a large, nonmajors biology class* (Doctoral dissertation). Tennessee State University, Nashville, TN. Retrieved from ProQuest. (UMI No. 3061781)
- Sampson, V., & Clark, D. B. (2009). The impact of collaboration on the outcomes of argumentation. *Science Education*, 93(3), 448–484.
- Schroeder, C., Scott, T., Tolson, H., Huang, T., & Lee, Y. (2007). A meta analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching*, 44(10), 1436–1460.
- Shooter, S., & McNeill, M. (2002). Interdisciplinary collaborative learning in mechatronics at Bucknell University. *Journal of Engineering Education*, 91(3), 339–344.
- Smith, K., Sheppard, S., Johnson, D., & Johnson, R. (2005). Pedagogies of engagement: Classroom-based practices. *Journal of Engineering Education*, 94(1), 87–101.
- Starrett, S., & Morcos, M. (2001). Hands-on, minds-on electric power education. *Journal of Engineering Education*, 90(1), 93–100.
- Stull, A. T., & Mayer, R. E. (2007). Learning by doing versus learning by viewing: Three experimental comparisons of learner-generated versus author-provided graphic organizers. *Journal of Educational Psychology*, 99(4), 808–820.

- Terenzini, P., Cabrera, A., Colbeck, C., Parente, J., & Bjorklund, S. (2001). Collaborative learning vs. lecture/discussion: Students' reported learning gains. *Journal of Engineering Education*, 90(1), 123–130.
- Wilson, P. N. (1999). Active exploration of a virtual environment does not promote orientation or memory for objects. *Environment and Behavior*, 31(6), 752–763.

Authors

Muhsin Menekse is a research scientist at the Learning Research and Development Center at the University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA, 15260; muhsin@pitt.edu.

Glenda S. Stump is Associate Director for Assessment and Evaluation at the Massachusetts Institute of Technology (MIT) Teaching and Learning Laboratory, 77 Massachusetts Avenue Cambridge, MA 02139-4307; gsstump@mit.edu.

Stephen Krause is a professor of materials science and engineering at the Ira A. Fulton School of Engineering at Arizona State University; ECG 221, School of Materials, Tempe, AZ 85287-6106; skrause@asu.edu.

Micheline T. H. Chi is a professor of learning sciences and cognitive psychology at the Mary Lou Fulton Teachers College at Arizona State University; P.O. Box 871908, Tempe, AZ 85287-1908; michelene.chi@asu.edu.

Appendix A

Description of Activities Used in Study 1

Table A1 Complete Description of In-Class Activities Used in the Introductory Materials Science and Engineering Class

Topic	Name	Type	Description
Atomic Bonding	Materials selection	Active	Selecting the most likely material, property of that material, type of bonding, and processing method from a given list for the motorcycle parts such as motorcycle fender or seat.
Atomic Bonding	Bonding concept map	Interactive	Drawing to complete the partially constructed concept map about atomic bonding. Also, students are asked to discuss and agree on what to draw and explain and write down their reasoning for every single decision they make to complete this concept map.
Crystal Structures	Concepts in context	Interactive	Overall goal is matching the five different historical events (disasters involving failure of materials) with the scientific reasons for the occurrence. Students are asked to discuss and agree on their matching decisions. They are asked to write down their reasoning for their final answers as well. Task 1: Matching with the possible reason for the change of materials. Task 2: Matching with the type of transformation. Task 3: Matching for the condition for change. Task 4: Matching with the processing method.

Table A1 (continued)

Topic	Name	Type	Description
Crystal Structures	Hidden treasures	Interactive	<p>Overall goal is to discover the properties of a face-centered cubic unit cell. Students are asked to discuss and agree on their decisions. They also write down their rational on the activity sheet.</p> <p>Task 1: Calculating the number of atoms on faces, edges and corners of a FCC unit cell.</p> <p>Task 2: Calculating the length of the cube edge, face diagonal and body diagonal in terms of atomic radius.</p> <p>Task 3: Calculating the coordination number and atomic packing factor of a FCC unit cell.</p>
Crystal Structures	Unit cell directions	Active	<p>Overall goal is to reproduce the specified unit cell directions by copying; and selecting indices from a given set of unit cell directions.</p> <p>Task 1: Copying unit cell directions from a cubic unit cell diagram for specific Miller indices to a blank piece of paper.</p> <p>Task 2: Selecting the Miller indices of unit cells from a given set of unit cell directions.</p>
Crystal Structures	Unit cell families of directions	Active	<p>Overall goal is to reproduce the families of unit cell directions by copying; and selecting a family of directions.</p> <p>Task 1: Copying all unit cell families of directions from a cubic unit cell diagram to a blank piece of paper.</p> <p>Task 2: Selecting a family of directions that are equivalent in terms of properties and packing density.</p>
Crystal Structures	Unit cell planes	Constructive	<p>Overall goal is to construct unit cell planes and determine indices of unit cell planes.</p> <p>Task 1: Drawing the planes in the unit cell by using given Miller indices.</p> <p>Task 2: Determining the Miller indices of unit cells from a given positions of planes.</p>
Crystal Structures	Unit cell worksheet	Constructive	<p>Overall goal is to construct the locations of atoms in a unit cell; and calculate total number of atoms for three different unit cells.</p> <p>Task 1: Drawing atom locations in two-dimensions based on the given indices of planes and atomic packing factor.</p> <p>Task 2: Drawing and calculating the total number of atoms per area for various planes.</p>

Concept Learning in Context: Materials Science of Unit Cell Disasters



Occurrence (Object)	Property & Change	Unit Cell Transformation	Condition for Change	Original Processing Method
Napoleon's failed winter invasion of Russia 1812 (tin button)	_____	_____	_____	_____
The World Trade Center 9/11 (steel girders)	_____	_____	_____	_____
The Titanic sank (steel rivets)	_____	_____	_____	_____
Helicopter Crash (steel gear)	_____	_____	_____	_____
Grandma's hip joint failed (ceramic ball cracked)	_____	_____	_____	_____

Appendix B

Active Version of Concepts in Context Activity

(Color figure can be viewed in the online issue)

PROPERTIES & CHANGE	UNIT CELL TRANSFORMATION	CONDITION FOR CHANGE	PROCESSING
<p>I. steel BCC to FCC higher temp</p> <p>II. BCC loses ductility at low temp</p> <p>III. ductile BCT (body center tetragonal) metal transforms to brittle powder with diamond cubic unit cell</p> <p>IV. cracks form in BCT ceramic when it transforms to body center monoclinic</p> <p>V. soft FCC phase is retained in hardened steel phase (BCT)</p>	<p>1. _____</p> <p>2. _____</p> <p>3. _____</p> <p>4. _____</p> <p>5. _____</p>	<p>A. 120°C sterilization phase change</p> <p>B. incomplete phase transform</p> <p>C. loses strength above 730C</p> <p>D. iceberg cold environment</p> <p>E. temp falls below 13°C</p>	<p>a. sintered</p> <p>b. forged</p> <p>c. cast</p> <p>d. machine & heat treat</p> <p>e. hot rolled in steel mill</p>

Appendix C

Sample Quiz for Concepts in Context Activity

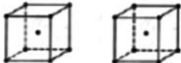
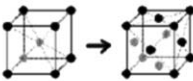
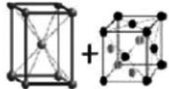
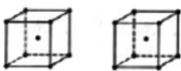
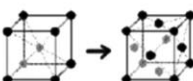
Concepts in Context: Materials Science of Unit Cells in Disasters

1. A. Which of the following disasters/failures has occurred as a result of an incomplete phase transformation?
 - a) Helicopter crash (steel gear)
 - b) Napoleon's failed winter invasion of Russia 1812 (tin button)
 - c) The World Trade Center 9/11 (steel girders)
 - d) Grandma's hip joint failed (ceramic ball cracked)
 - e) The titanic sank (steel rivets)
- B. Using your understanding of macroscopic properties and atomic level structure, explain what could have been done to avoid the disaster that you choose above?

Appendix C (continued)

2. A. A steel skeleton chemical processing plant collapses due to a steel beam failing prematurely a short time after a chemical explosion and a fire.

Choose the most likely condition for change, properties and change, and unit cell transformation for this disaster.

<u>Condition for change</u>	<u>Properties and Change</u>	<u>Unit cell transformation</u>
a) 140 ^o C sterilization phase change	Ductile metal to brittle powder	
b) Loses strength above 730 ^o C	Steel BCC transforms to FCC at higher temperatures	
c) Incomplete phase transformation	Ductile metal to brittle powder	
d) 140 ^o C sterilization phase change	Steel BCC transforms to FCC at higher temperatures	
e) Loses strength above 730 ^o C	BCC loses ductility at low temperature	

B. As a consulting engineer giving advice to the company, specify your recommendation to prevent this failure and justify it based on your understanding of the relationship between macroscopic properties and atomic level structure.

Appendix D

Sample Rubric of the Quiz for the Materials Selection Activity

1. A. What type of processing would be used to create a spark plug insulator?
[Verbatim type question]

- a) Vacuum warm forming
- b) Calendaring
- c) Wire drawing
- d) Metal stamping
- e) Sintering

[Sintering is correct answer]

- B. Based on your selection above, explain why this processing method is optimal to process the spark plug insulators. [Inference type question]

Scoring Rubric for Question 1.B.:

4 points	3 points	2 points	1 point	0 point
Response includes at least four of the five correct statements. It may also include minor irrelevant information but no incorrect explanation.	Response includes three of the five correct statements. It may also include minor irrelevant information but no incorrect explanation.	Response includes two of the five correct statements. It may also include minor irrelevant information and/or minor incorrect explanation.	Response includes one of the five correct statements. It may also include major irrelevant information and/or minor incorrect explanation.	No response. Or, responses without any of the correct statements.

Correct Statements:

1. *A spark plug insulator must resist the flow of electric current and it needs to have a high melting point.*
2. *The metallic materials commonly have high melting points but they are not good insulator therefore, using a material with metallic bonds is not good to manufacture spark plug insulators.*
3. *Ceramics are good insulators and have high melting point due to ionic and covalent bonds. Therefore, using ceramics to manufacture spark plug insulators is ideal.*
4. *The processing method for ceramics should involve consolidation of ceramic powder particles by heating the part to a high temperature below the melting point.*
5. *Among the options above sintering is the only method that fires ceramic powders at high temperatures.*