

Beyond and Within Classroom Walls: Designing Principled Pedagogical Tools for Student and Faculty Uptake

Elizabeth S. Charles, Dawson College, echarles@dawsoncollege.qc.ca
Nathaniel Lasry, John Abbot College, lasry@johnabbott.qc.ca
Chris Whittaker, Dawson College, cwhittaker@dawsoncollege.ca
Michael Dugdale, John Abbot College, michael.dugdale@johnabbott.qc.ca
Kevin Lenton, Vanier College, lentonk@vaniercollege.qc.ca
Sameer Bhatnagar, Dawson College, sbhatnagar@dawsoncollege.qc.ca
Jonathan Guillemette, John Abbot College, jonathan.guillemette@johnabbott.qc.ca

Abstract: This paper tells the story of DALITE, a design-based experiment, involving a co-design project, spanning three iterations and nearly four years. DALITE was designed around the principles of peer instruction, in particular, self-explanation and intentional reflection. DALITE consists of two databases: a database of college-level physics questions and another of student-generated rationales. We report on the iterative design process that has shaped this web-based asynchronous learning environment and on its use as part of an active learning pedagogy. Using mixed methods we have tested DALITE's effectiveness to support conceptual change, examined the content of its databases, observed its use in classrooms, and conducted interviews from users – students as well as instructors. These data help to reveal the conditions of DALITE's effectiveness. We discuss our findings and provide design guidelines.

Keywords: active learning; peer instruction; design-based experiment; conceptual change.

Introduction

Student-centered or *active learning* privileges the role of collaboration, reflection and discourse as principle methods and goals. The effectiveness of active learning on improving learning outcomes is well documented in the literature (see Freeman *et al.*, 2014). Research in the learning sciences and CSCL has devoted much attention to the design of tools and activities that promote active learning. What is surprising, however, are the barriers that faculty members perceive in the use and adoption of research-based active learning instructional strategies (Henderson & Dancy, 2007).

One approach that has managed to overcome this barrier is *peer instruction* (PI; Mazur, 1997). It is reported to be one of the most widely adopted active learning pedagogies at the post-secondary level (Meltzer & Thornton, 2012; Henderson, 2008). Researchers have documented the benefits of PI (Crouch & Mazur, 2001; Lasry, Mazur & Watkins, 2008), and the impact of variations in how it is scripted (Turpen & Finkelstein, 2009). But there is still much to learn about this pedagogy, how its components might be effectively used differently, and how it might be designed for use outside the classroom. This current research explored the impact of several key features of PI and took up the challenge of bringing it online with the design of a tool called DALITE – the Distributed Active Learning Interactive Technology Environment.

DALITE is a web-based learning platform designed to provide students with an asynchronous version of peer instruction. It is designed to engage students in written self-explanation (rationales), comparison of these rationales to those of peers, and reflection on the quality of rationales (thumbs-up) along with comparison of rationales to that of an expert. DALITE is intended to be part of an active learning pedagogy and allows the course content to follow students outside of class. Through a unique database of student-generated answers and rationales, students share their understandings with peers asynchronously. Activity in DALITE starts with the class being provided with a DALITE assignment, usually a series of 3-5 conceptual questions. Individual students can then log into the system using a computer or mobile device. Each time a student answers a question they are asked to provide a written rationale. The student-generated rationales become part of the DALITE database, and eventually are used in the DALITE *script* for future students. DALITE assignments provide teachers with the opportunity to see what their students are thinking, either before or after classroom instruction. We elaborate on this script later.

In this paper we report on the results of this design-based research project and what was learned from the experience. In particular, we focus on the efforts to design DALITE to promote reflection and discourse, and promote construction of conceptual knowledge – i.e., conceptual change. Additionally, we add our findings to the growing body of principles for design of CSCL-inspired tools, particularly those related to asynchronous methods for engaging students in learning activities.

Theory and background

Findings on active learning pedagogies show strong improvements in students' conceptual learning (Freeman et al., 2014; Meltzer & Thornton, 2012) and a better adoption of discipline-based thinking (Van Heuvelen, 1991). A recent meta-analysis by Freeman, et al., (2014) provides a comprehensive and up-to-date review of the results of active learning.

While there are different models of conceptual change (Chi, Slotta & deLeeuw, 1994; diSessa, 1993; Vosniadou, 1994) evidence suggests mechanisms that promote the process, these include: intentional reflection (Sinatra & Pintrich 2003), self-explanation (Chi, et al., 1994), compare and contrast tasks (Bransford, Franks, Vye, & Sherwood, 1989). The common features of these methods is the reflective and discursive practices that each tap into, in their own way. Equally, the value of such socio-cognitive and socio-cultural practices has been examined in other areas of research in the field of CSCL (Stahl, 2006).

In the past ten years there has been a proliferation of active learning techniques that tap into this literature and that have emerged at the elementary and middle school levels – e.g., inquiry based science, learning by design. However, active learning approaches have been less structured at the post-secondary level and have relied on the general notion of collaboration.

Peer instruction

A major feature of PI is the potential for discussion and debate between students with opposing viewpoints. Implementations of PI generally follow a script similar to think-pair-share (for details of the PI script see Mazur, 1997). Some of the most successful implementations of PI have been those found in university settings that involve large lecture halls, with hundreds of students, supported by teams of teaching assistants (TAs). Rich discussions can arise in such settings because of the increased probability of students with diverse views. In addition, many of the successful examples have the benefit of TAs to monitor conversations and to create more effective student pairings – i.e., more diverse viewpoints. If the success of PI is based on this diversity then what happens in smaller classrooms, or when budgets limit the number of teaching assistants? Equally, if PI depends on all students having equal opportunity to express their view or join the discussion what happens when some students choose not to participate or are silenced by social factors involved in group dynamics? These are some of the questions that the design of DALITE attempts to address.

Summing up what we wanted to learn from DALITE

The current iteration of DALITE was conceived as a way to harness the benefits of PI and address some of its missed opportunities. In traditional enactments of PI, students' reasoning is seldom recorded, save the chance overheard conversation or if the instructor intentionally calls for explanations. As such, it is difficult to know what types of arguments and reasoning are convincing to students, and which elements of an argument might move students towards or away from the correct answer. In addition, until now, there has been no systematic documentation of the rationales used by students to arrive at an answer to conceptual questions. Therefore, there is no way of knowing what could be learned from how students' responses might help to shape more effective conceptual questions. Additionally, might there be different effects regarding the timing of assigning questions, might there be an issue of the context of delivery or the wording? Can we promote better forms of *intercontextualization* (Engle, 2006) with the sequencing and design of questions?

In this study we attempted to look at these questions along with the bigger issues of whether or not DALITE's design promotes reflection and discourse, and conceptual change. In the next sections we describe the three-year project, briefly, then some of the results that have implications for the design of future learning environments.

Methods

DALITE used a *design-based research* (DBR) methodology. DBR is a pragmatic and iterative approach to educational research that allows researchers and pedagogical designers to examine the conditions and context that surround the implementation and use of an innovation. Anderson and Shattuck (2012) describe DBR as consisting of six main features: (1) situated in real educational contexts; (2) focuses on the design and testing of

significant innovations (e.g., pedagogical approaches, instructional tools and systems) that are theoretically based; (3) generally spans a period of several iterations of the innovation; (4) a collaboration between researcher and practitioners; (5) uses mixed methods for data collection; and (6) involves evolution of design principles, thereby adding to the understanding of educational theory. This current research adheres to these six features, in particular, it used a mixed methods approach to data collection.

Context

DALITE is a project conducted by a research team at three English-speaking colleges in Quebec, Canada. In all iterations, it was implemented in a calculus-based mechanics physics course, equivalent to a typical undergraduate freshman physics course at university (sometimes called *gateway*), though the pacing is somewhat slower because of a 15-week semester. Students were of diverse cultural backgrounds with a majority having a language other than English as a first language. The majority of students were first year science majors between the ages of 17 – 19, enrolled in one of two science profiles – Pure and Applied or Health Sciences. In all cases, participants were from the regular stream, which means their high school grades in the science and math prerequisites are between 70-90%. Typically, the gender distribution in these programs is between 1:1 and 1:2 ratio (male:female).

Data collected

The data collected for this project used mixed-methods. Qualitative sources included the DALITE databases, classroom observations, and both student and teacher interviews. At each development phase, these data were used to help describe under what circumstances DALITE worked best, and when it did not. At each point, these data informed future design decisions. In addition to the qualitative data, we also collected pretest-posttest data from the Force Concept Inventory (FCI; Hestenes *et al.*, 1992). It is a 30-item multiple-choice questionnaire and the most widely used and researched assessment of Newtonian concepts (McDermott & Redish, 1999). All three iterations of DALITE used the FCI as a measure to assess students' learning, and for the purpose of comparison (using a quasi-experimental design).

DALITE's databases

To date, the DALITE Curriculum database contains over 120 questions spread across the three main topics that are generally covered in an introductory physics course – i.e., kinematics, dynamics, and conservation principles (energy and momentum). These questions are designed to be roughly at the first-year university level. Influenced by the Ohio State concept test questions (Lee, Ding, Reay, & Bao, 2011), many questions are organized into sets of three to four questions on a single concept that progressively increase in difficulty. These sets of increasingly difficult multiple-choice problems are built on similar deep structures with different surface features, or similar surface features with different deep structures. While the individual instructor has control over the selection of questions to be assigned to his/her students, the corpus of questions was selected by the team of instructors (co-researchers).

The database of student-generated rationales was first developed through a “seeding” process. That is, about 20 students were asked to use a streamlined DALITE script that had them answer the questions and write rationales. This process was necessary so that the first real students using the system could see rationales. This requirement to “seed” the database places constraints on the development of new questions entering the system but was deemed necessary for collecting authentic data. The rationale database was cleaned and nonsense rationales were eliminated (e.g., unreadable text, meaningless strings of symbols). To date there are over 7000 rationales available for the 120 questions with the most populated questions having over 120 rationales each.

History of designing DALITE

First iteration 2010-11

The first iteration began in 2010 as a co-design project with a team of researcher/designers from another institution in Eastern Canada. That instantiation used the open source platform SAIL Smart Space along with tagging capabilities, and built upon the success of studies conducted by Tissenbaum and Slotta (2009). The curriculum was developed using a co-design approach, working in close collaboration with physics instructors at the college-level, where DALITE was enacted regularly throughout a semester in one section of Mechanics with 32 science students. All classes were held in a technology-rich smart classroom environment, an active learning space with a seating arrangement designed to facilitate collaboration.

Design features

Two main ideas steered the design and implementation of the first iteration – adaptive scripting and orchestration (Dillenbourg & Fischer, 2007). To start, there were activities at three levels of social organization - the individual, the dyad and the small group (supergroup). At each level the individual or group was asked to perform four tasks: (1) categorize the type of question by tagging the major elements (element list provided), (2) write a short rationale to explain the choice, (3) answer a multiple-choice question, and (4) write a short rationale to explain the categorization. Each step was scripted in such a way as to encourage students to think about the underlying principles involved in the problem *prior* to solving it (step 1), and then to reflect on why/how this helped them solve the problem (step 3). The first step began as homework (at the individual level), subsequent steps were completed in class (in dyads and supergroups). At the dyad and supergroup levels students were able to view the collective work of the rest of the class (from the homework phase of the activity) in the form of an aggregated histogram of tags and individual reflections. As students moved from one level to the next the questions become progressively more challenging, requiring the thinking of many minds. For a more complete reporting see (Charles, Tissenbaum, Whittaker, Lui, Dugdale & Slotta, 2011).

Second iteration

The second iteration of DALITE began with an effort to modify the original infrastructure to accommodate new constraints and a new programmer. Eventually, the new infrastructure consisted of the following components: (1) a student registration and software application management system; (2) a framework for data mining and tracking of student interactions in real time including the instructional scripts; (3) a locally stored central database or repository; and (4) data displays for instructors. The platform used “Agile” development practices with the aim of ensuring future availability, scalability, and performance. The database repository is composed of two parts: (1) the curriculum content – conceptual multiple-choice questions (sometimes referred to as *concept test* questions); and, (2) the student-generated answers and rationales for these answers.

The important difference between iteration one and two was the changing of the orchestration and the script. In particular, while there was substantial benefit from the in-class work (first iteration) it was too difficult to sustain over the length of the course. Additionally, only one of the two institutions involved had the requisite facilities to adequately implement such an involved active learning curriculum. As such, the in-class portion was removed and the homework script was elaborated. The script continued to put a strong emphasis on the tagging activity and the research team focused on the creation of appropriate tags for the growing question database. Results of this second iteration showed that while students used the tags, they were not sufficiently discriminating between concepts or between categories of questions, which was the intention. Furthermore, when interviewed, students had not recognized how to use the tags or their purpose.

Third iteration

The third iteration, which is the current version of DALITE, continues to use the second-generation infrastructure. Based on the results of the second phase, the tags were removed from the system, but not from the study. Instead, we designed a test of a new form of tag that would be orchestrated in the classroom, which could be more closely scaffolded by the instructor. These tags showed some improvements but are still a work in progress. We do not report on them further in this paper.

Scripting the third iteration of DALITE

In the current iteration of DALITE, students log into the system and are directed to a prepared assignment that consists of sets of questions similar to those used by in-class PI. They are asked to follow the sequence of six steps: Step 1, students select an answer for a multiple-choice question, similar to the first step in PI. Step 2, students select which answer is correct and write a rationale that explains their thinking (Figure 1).

In Step 3, students are asked to reconsider their original answer in the context of rationales for their own answer, and a similar selection of rationales for an alternative answer. This purposeful comparison is designed to provide the variety that is sometimes missing in face-to-face PI. Step 4, students re-select their answer, choosing to stay with their original selection or change, based on the rationales (Figure 2). Step 5, students are asked to vote on most convincing rationale presented (optional step). Step 6, students are presented with the rationale from an expert (teacher) and are encouraged to review their rationale relative to the expert's rationale (Figure 3). Note that the correct answer is not shown to the students, only a correct rationale. It is left up to the student to infer the correct answer and the instructor to follow-up in-class to present the correct answer. There was also an instructor's display interface that allowed DALITE to be brought into the classroom and provides real time review.

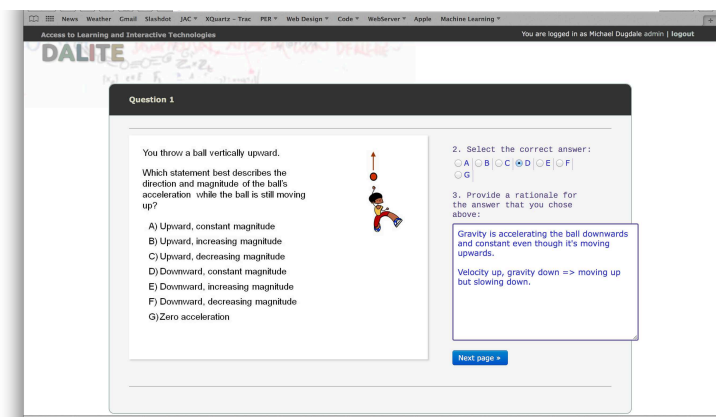


Figure 1. DALITE screens for steps 1 & 2: select, vote and write a rationale.

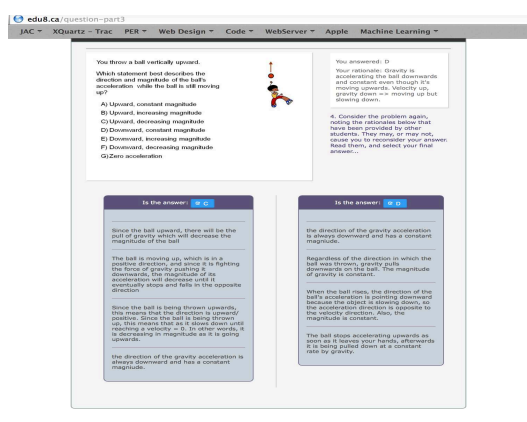


Figure 2. DALITE screens for steps 4.

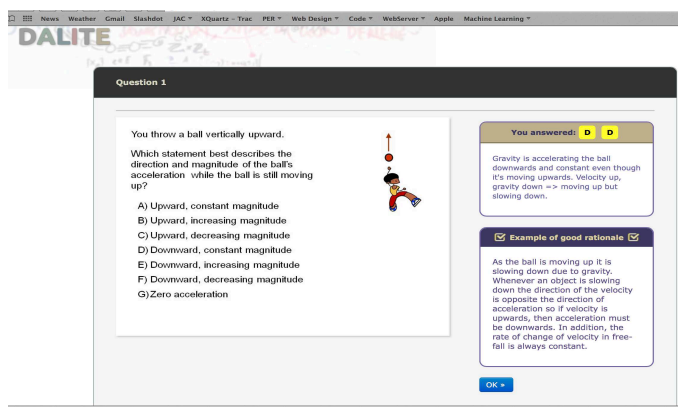


Figure 3. DALITE screens for steps 6, expert rationale.

Context and participants

In the third and final iteration, DALITE was again implemented as part of a first-year introductory calculus-based mechanics (physics) course. DALITE content questions reflect the three units of the Mechanics course – Kinematics (1DKin and 2DKin), Dynamics (LinDyn) and Conservation principles (Momentum and Energy) and arranged in groupings of three to five questions per assignment.

Observations and data were collected from a culturally diverse student population in five sections taught at three urban colleges by four instructors: (1) College A, one instructor teaching two sections – groups T09 & T10; (2) College B two instructors teaching 1 section each – T07 & T08; (3) College C, one instructor teaching one section – T06. All instructors used an active learning pedagogy but had varying degrees of experience with this practice. The instructor for groups T09 and T10 being the most experienced, having over six years developing the skills required to orchestrate such pedagogy; the instructor for group T08 had the least (less than two years). These differences are taken into account in the interpretation of the results and discussion.

Confirming DALITE's effectiveness

In this third iteration, DALITE students (N=137) were compared to two control cohorts. The first comparison cohort came from a database of 13,422 students who had taken the FCI of which a subset of 2,913 had the same incoming score range as our DALITE sample. This control data was comprised of students taught with a variety of pedagogical approaches and thereby provide an unbiased comparison. A *purposeful sampling* method was used to select cohort 2, the “peer instruction no-DALITE” group. These students were part of two sections, one taught by a teacher in College A and another taught by a teacher at a larger institution of higher education

(N=188). Both instructors had used PI in their classes for several years. Comparison to such a sample is critical and ensures our results are authentic and meaningful.

The results show that DALITE students outperformed this regular control cohort (0.47 ± 0.02 vs 0.350 ± 0.006 ; $p < 0.000$). However, no difference was found in conceptual gain between DALITE students and those who had in-class Peer Instruction (0.47 ± 0.02 vs 0.48 ± 0.02 ; $p = 0.84$). In other words, students using DALITE ($n = 137$) in their college courses do not differ significantly in conceptual gains ($p = 0.38$) from students who used real-time Peer Instruction ($n = 188$). The results show a surprising similarity between four of the five groups and a small difference with a fifth (section T06). Overall differences are between DALITE groups are not statistically significant ($g_1 = 0.50$; $g_2 = 0.50$; $g_3 = 0.47$; $g_4 = 0.48$; $g_5 = 0.38$; $p = 0.06$) with four of the five groups being extremely similar and close to all the variation residing in the fifth group.

Building on DALITE rationales

Sixty-six DALITE questions from the database of 120 were made available to instructors, of which 48 were used across all 5 sections. Total numbers of DALITE assignments ranged between 12 to 15 with totals of 48 to 66 questions, respectively. In each case students were required to answer the question and produce rationales according to the DALITE script described earlier. A total of 6837 student-generated rationales were produced (see Table 1).

Table 1: Descriptive statistics on the rationales generated by students in the five sections, across the 3 colleges.

Sections	T06	T07	T08	T09	T10	Totals
# students	n=30	n=41	n=36	n=31	n=30	N=168
#Qs. Assigned by instructor	48	50	48	66	66	
Mean #Qs. answered per student	36	40	34	51	58	
Mode #Qs. answered per student	48	50	48	58	66	
% Qs. completed per student	75	80	70	78	88	
Total # of rationales	1081	1637	1206	1235	1678	6837
Avg. length of rationales (z-score avg.)	0.10	-0.51	1.88	1.14	2.26	

Of these five sections, T9 and T10 were assigned the most questions with section T10 being unique in that students had the highest completion rates of any section. Note that the same teacher taught sections T9 and T10, and was the most skilled with active learning pedagogy. Interestingly, while students in section T10 had the longest mean rationales, students in section T08 also wrote longer rationales. These trends were investigated, but not reported here.

Learning from DALITE

Is DALITE as effective as PI?

Our results show that DALITE, for this cohort, was as effective as in-class peer instruction (PI), measured by gains on the FCI. These findings are consistent with the instructors' perceptions that DALITE allowed them to replace the time spent on clicker questions (PI) with other group activity.

Did students engage in the DALITE script?

Most students completed the majority of the assigned DALITE questions. The length of an average rationale was in the range of 20–30 words (see Table 1). Questions dealing with linear dynamics (LinDyn) were particularly good at eliciting longer rationales (average lengths were statistically greater) as were graphical questions.

Why did students engage in DALITE?

Twenty-six post-instruction student-interviews were conducted from all five sections. This was a purposeful sampling therefore not all sections had the same number of students interviewed. We selected a range of students, some who wrote long rationales and others who wrote short ones. The interview data provides us with information about the specific features designed into DALITE and how students used them. These data suggest that students were motivated by a variety of reasons. The two most prominent were related to the following DALITE script features: (1) prompting for self-explanation; (2) providing an opportunity to compare with peers.

For instance, students referred to how they were “learning how to learn” when doing their DALITE assignments. In response to the question, Why did you write longer than average rationales, a student answered:

So I was trying to explain it to myself. I wanted to get all the points out and didn't want to leave anything out because I would print out the notes after and study that. So if I only had 2 sentences, I'd go back to it [the notes] and say I don't understand so when I wrote out my thought process when I would go back to it [DALITE]. Next time it would be a lot easier to understand the material. I believe that's why I write a lot.

Other students stated that DALITE provided a low-risk environment. Students were encouraged to express their understandings without the punitive specter of grades and judgement. Recall, there is no right or wrong assessment at the end of the DALITE script, instead students are encouraged to compare their rationale to that of an expert. One student in particular, who wrote progressively longer explanations throughout the term talked about how it encouraged her to express herself:

I think it was my confidence in my physics knowledge towards the end of the course... I was kind of tentative at first. And, I was kinna of like figuring out DALITE, theory, all that's happening. So I would write a little sentence but by the end I was so used to the process and had knowledge so I put everything down, everything I could possibly think of. It's almost [as if] I had lost the fear of being wrong, which is cool... It's like thinking out loud.

Another finding from the interviews was the growing sense of increased self-regulation and awareness of one's explanation. By reading their peer's rationales, some students began wanting to model good explanations both for themselves and others. One student stated that reading “choppy” rationales changed the way she wrote her own.

now I explain the concept behind everything, so I give more detailed rationales.... Since you have to present [your rationale], you have to say “ok this is what we think and why.” It organizes your thoughts.

Orchestration of DALITE: How DALITE interacts with other classroom activities

Two instructors (T08, T09 & T10) engaged their students with more writing outside of DALITE (using reflective writing and other activities). Students of those instructors write longer rationales and reported that having the opportunity to explain to themselves in writing helped prepare them for their class discussions. In fact, they state that their in-class discussions were “passionate”. Though they worked on DALITE alone, they would discuss their answers and some of their study habits with each other in class. Students in these groups enjoyed the authority the active learning pedagogy and classroom atmosphere provided. As a result, their ways of reasoning out problems together and as a “team” became a way of being: “*We were passionate about what we were doing, we wanted to do well, we wanted to do well together.*” However, this self-regulation seems to need to be supported by other classroom activities. In other words, students engaged in other writing activities valued in the classroom wrote more in DALITE and engaged more with the course materials.

Conclusion

As with all DBR projects, we provide guidelines for design. First, scripting works and can support self-explanation and reflection. Database questions can produce different levels of engagement (as evidenced by the length of rationales). Most importantly, DALITE can work differently depending on how it is orchestrated. Indeed, although writing ought to be influenced by linguistic proficiency, non-anglophone students were particularly engaged and produced longer rationales in sections where teachers place greater value on writing and pre-class preparation (reflective writing or Flip-JiTT). In short, orchestration is important for its uptake. Orchestration means bringing these activities and artifacts into the classroom and sending them back out, at the same time sending the message that you value both the doing as well as the way that it is done.

References

- Anderson, T., & Shattuck, J. (2012). Design-Based Research A Decade of Progress in Education Research?. *Educational Researcher*, 41(1), 16-25.
- Bransford, J. D., Franks, J. J., Vye, N. J., & Sherwood, R. D. (1989). New approaches to instruction: Because wisdom can't be told. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp.

- 470–497). New York, NY: Cambridge University Press.
- Charles, E.S., Tissenbaum, M., Whittaker, C., Lui, M., Dugdale, M., & Slotta, J.D. (2011). *Co-design of Collaborative Collective Knowledge Environment*. In Spada, H., Stahl, G., Miyake, N., Law, N. (Eds.) *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings. Volume II*. pp. 641-645. International Society of the Learning Sciences.
- Chi, M. T., Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improve understanding. *Cognitive science*, 18(3), 439-477.
- Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69(9), 970-977.
- Dillenbourg, P. & Fischer, F. (2007). Basics of computer-supported collaborative learning. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 21, pp.111-130.
- diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, 10(2-3), 105-225.
- Engle, R.A. (2006). Framing interactions to foster generative learning: A situative explanation of transfer in a community of learners classroom. *Journal of the Learning Sciences*, 15(4), 451-498.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 201319030.
- Henderson, C., & Dancy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics-Physics Education Research*, 3(2), 020102.
- Henderson, C. (2008). Promoting instructional change in new faculty: An evaluation of the physics and astronomy new faculty workshop. *American Journal of Physics*, 76(2), 179-187.
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Phys Teacher*, 30(3), 141-158.
- Lasry, N., Mazur, E., & Watkins, J. (2008). Peer instruction: From Harvard to the two-year college. *American Journal of Physics*, 76(11), 1066-1069.
- Lee, A., Ding, L., Reay, N. W., & Bao, L. (2011). Single-concept clicker question sequences. *The Physics Teacher*, 49(6), 385-389.
- Mazur, E. (1997). *Peer instruction* (pp. 9-18). Upper Saddle River, NJ: Prentice Hall.
- McDermott, L. C., & Redish, E. F. (1999). Resource letter: PER-1: Physics education research. *American journal of physics*, 67(9), 755-767.
- Meltzer, D. E., & Thornton, R. K. (2012). Resource letter ALIP-1: active-learning instruction in physics. *American journal of physics*, 80(6), 478-496.
- Sinatra, G. M., & Pintrich, P. R. (2003). The role of intentions in conceptual change learning.
- Stahl, G (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.
- Tissenbaum, M., & Slotta, J. D. (2009). A new framework for smart classroom research: Co-designing curriculum, research and technology. In *Proceedings of the 9th international conference on Computer supported collaborative learning-Volume 2* (pp. 91-93). International Society of the Learning Sciences.
- Turpen, C., & Finkelstein, N. D. (2009). Not all interactive engagement is the same: Variations in physics professors' implementation of Peer Instruction. *PRST-PER*, 5(2), 020101.
- Van Heuvelen, A. (1991). Learning to think like a physicist: A review of research-based instructional strategies. *American Journal of Physics*, 59(10), 891-897.
- Vosniadou, S (1994). Capturing and modeling the process of conceptual change. *Learning & Instruct*, 4(1), 45-69.

Acknowledgements

The research team wishes to thank the *Programme d'aide à la recherche sur l'enseignement et l'apprentissage* (PAREA; grant PA2011-06), the Ministry of Education (MELS & MERST) program, in the Province of Quebec; Dawson College, special grant from the Director General, for the development of DALITE; Jim Slotta & Mike Tissenbaum, from OISE; the software programmer at *Edu.8 Development*; and our research assistant Chao Zhang.