Practice and Evidence of Scholarship of Teaching and Learning in Higher Education
Special Issue: Threshold Concepts and Conceptual Difficulty
Vol. 12, No.2 April 2017, pp. 352 - 377

Threshold Concepts in Physics

David Harrison*
Department of Physics,
University of Toronto, Canada
david.harrison@utoronto.ca

Ruxandra Serbanescu
Department of Physics,
University of Toronto, Canada
sandra@physics.utoronto.ca

Abstract

In the last 25 years Physics Education Research has identified a number of fundamental ideas and concepts which beginning students have particular difficulty with, and found methods of instruction that are more effective than traditional pedagogy in helping students to understand the material. Here we discuss two of these ideas and concepts by two case studies. Case study 1 regards Newton's 1st Law of motion, and case study 2 is about the uncertainty of physical measurements. The analysis is from the perspective of *threshold concepts*, *troublesome knowledge*, and *liminality*. For each case study we discuss the research-based pedagogy used in teaching the material. We then add another perspective on these issues from Piagetian taxonomy. We then discuss the results of interviews with students about concepts that they struggled with, and ways that they found helped them go through the threshold to gain a deeper understanding of those difficult ideas.

Keywords: Physics Education Research, Newton's 1st Law, uncertainty in physical measurements, Piaget.

Introduction

Physics teachers in both secondary and post-secondary introductory courses have long suspected that many of our students have fundamental misconceptions about the nature of the physical universe and our description of that universe using mathematical language. About 25 years ago physics teachers began devising diagnostic instruments to identify those misconceptions, with the idea that by knowing more about the students' wrong ideas, we could address them directly in our courses. These instruments typically

---

* Corresponding Author

ISSN 1750-8428 (online) www.community.dur.ac.uk/pestlhe.org.uk

do not involve any significant calculations or the use of formulae to arrive at some algebraic or numeric answer: instead they focus on the concepts themselves. The best known of these instruments in the Force Concept Inventory (FCI), which was introduced by Hestenes, Wells and Swackhammer (1992), and was updated in 1995. The FCI is available from http://modeling.asu.edu/R&E/Research.html. The FCI has now been given to literally hundreds of thousands of students at a number of institutions worldwide.

The results of using these diagnostic instruments at the beginning of a course confirmed our suspicions: many of our students have surprisingly wrong ideas. In the next section we will discuss how at least some of these wrong ideas are threshold concepts.

Next, some teachers began giving the same diagnostic instrument at the beginning of the course, the "Pre-Course" and again at the end, the "Post-Course". The improvement in the scores from the Pre-Course to the Post-Course would show how much the students benefited from our "most excellent instruction". However, the results were underwhelming. Although many students had learned to take one or more formulae and use them to "plug and chug" to an answer to a particular problem, their understanding of the conceptual basis for the formalism had barely changed.

If what we were doing in our courses wasn't working in terms of the conceptual understanding of physics that we particularly value, then it seems obvious that we need to change our pedagogy. Using the diagnostic instruments in a Pre-Course/Post-Course protocol will then allow us to quantify whether or not the changed pedagogy actually works. This realization has led to a huge research field called Physics Education Research (PER). Many physics departments now have PER groups alongside the traditional research groups (high energy, condensed matter, atmospheric, etc.), and many offer PhD's in physics education. Figure 1 shows Physics Education Research groups that have registered with *PER Central* http://www.compadre.org/per/programs/. We will describe what types of pedagogy have been proven to be effective in more detail below, but for now will summarise the principle finding: students do not learn best by being lectured to. The best learning occurs when students interact with each other, particularly when those interactions are based on conceptually based activities using a guided-discovery model of instruction.

Figure 1. Physics Education Research Groups that have registered with PER Central (2014)

We should point out that the results of PER have been known intuitively by skilled educators since the time of Socrates. However PER allows us to prove that this type of pedagogy is effective. So in some sense we are using the techniques of physics research applied to education.

Threshold Concept 1: Newton's 1st Law

Newton's 1st Law of motion from his Mathematical Principle of Natural Philosophy states:

> *"Every body persists in its state of being at rest or of moving uniformly straight forward, except insofar as it is compelled to change its state by force impressed." (Newton, 1687)*

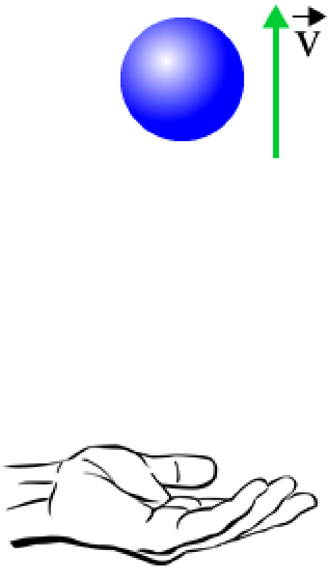
This view of the relation of forces to motion was quite different from the then accepted view, which was due to Aristotle. For Aristotle a body at rest was in its natural state. Heavy objects wanted to be at rest on the Earth, and light objects wanted to remain in the heavens. For a body to move in a straight line at a constant speed requires an external force. It is important to realize that Aristotelian dynamics is perfectly consistent with the students' lifelong observations about bodies in motion.

In the West, the Newtonian view was almost realized by Galileo in the early 17th century. Descartes realized what we now call Newton's 1st Law in 1633, but suppressed the result because of fears of the Inquisition. In China, Mo Tsu had a Newtonian view of the role of forces in the 3rd century BCE, when he wrote "The cessation of motion is due to the opposing force ... If there is no opposing force ... the motion will never stop."

Newton's 1st Law is now so well-known that it is the central theme of a current television commercial for an arthritis pain medication. However, just knowing the words of the Law is not the same as having passed through the threshold to actually understanding what the words mean.

An example of this lack of understanding is a common dialog that physics teachers have with students as they are studying Newton's Laws. Figure 2 shows the situation. You throw a ball straight up. Air resistance is negligible. While it is moving up with speed v, draw all the forces acting on the ball.

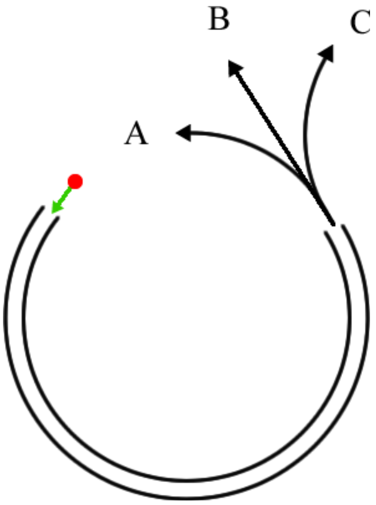
Figure 2. A hand has thrown a ball upwards



Since the ball is moving upwards but is slowing down, the only force acting on it is the force due to gravity, which acts down. Most students will correctly draw this force. But many of them will also draw a second force acting upwards on the ball. When asked, they will say something like "It is the force causing the ball to move upwards." In the Newtonian view, of course, this second force does not exist: there is only the force of gravity acting on the ball.

Another example of a similar misconception is a question from the Force Concept Inventory, shown in Figure 3. When the ball emerges from the track, there are no horizontal forces acting on it so the ball moves in a straight line and follows path B. But many students answer path A. Perhaps these students have some sort of quasi-idea about conservation of circular motion.

Figure 3. Three possible paths when the ball leaves the curved track.



A semi-circular track is mounted to a table top, and we are looking at the track from above. A small ball enters the left side of the track at high speed, and emerges from the right side. Which is closest to the path the ball follows?

Instead of just asking the question on the Force Concept Inventory, we in addition constructed a physical apparatus, and after posing the question to the class did the demonstration. Some students claim that they saw the ball follow path A! This is truly frightening: the students pre-conceived ideas were so strongly held that it effected what they believe they saw. We then laid a meter stick beside path B and repeated the demonstration. The ball clearly followed the straight line path right beside the meter stick. But some students claimed that we must have put magnets of something in the meter stick, because before we put the stick down the ball really did follow path A.

The fact that the ball really does follow path B with or without the meter stick is such a troublesome piece of knowledge that the students will actively reject it. When they are forced to confront the reality of path B, they will enter a psychological state which in Physics Education Research we commonly call "cognitive dissonance"; Piagetians will tend to use the word "disequilibration" to describe the same phenomenon.

A variation of the Force Concept Inventory question and in-class demonstration is to give the students the apparatus and have them do the experiment themselves. We have observed some students eventually getting frustrated. They complain, "I can't make the ball do what it is supposed to do."

A moment's reflection on what is happening to the students in these situations makes it clear that what we are asking them to do is to take down all the Aristotelian thought-forms that are based on a lifetime of experience. This is inherently:

- Difficult,
- Time consuming,
- Frightening.

These are, of course, all characteristics of threshold concepts.

Pedagogy 1

Above we stated that the key result of Physics Education Research is that students learn best by interacting with their peers while working with conceptually-based activities using a guided-discovery model of instruction. In this, the first of two sections on pedagogy, we will discuss how this is typically implemented. We believe that these same strategies can be easily modified for fields other than physics that confront their students with troublesome knowledge and threshold concepts.

Peer Instruction

If lectures are among the least effective forms of instruction, which they are, then if one is confronted with a large number of students in a lecture hall what can one do? A proven form of pedagogy is called *Peer Instruction*, which was introduced by Eric Mazur at Harvard in the early 1990s. (Mazur, 1997) In this method, the instructor briefly reviews the material and then poses a conceptually based question to the class, a ConcepTest. The curved track question of Figure 3 is an example of a suitable question. The students are given one or two minutes to think about the question and to individually "vote" on the right answer. The ideal question has about one-half of the class initially giving the wrong answer. Then the students are asked to discuss the question amongst themselves in groups of three or four students for two or three minutes, and to vote again on the right answer. Typically the percentage of students who get the right answer goes up dramatically. In addition, even students who initially got the correct answer are not very confident in the correctness of their answer. After Peer Instruction their confidence in the correctness of their answer also goes up dramatically. Further, when understanding of the concept is later tested on an examination, the new understanding of the concept is shown to persist.

For Peer Instruction to work, the students must have read the relevant sections of the textbook or course notes before the class. To insure that they do this, a short Pre-Class Reading Quiz on the material is given. The quiz is fairly trivial for students who have actually read the material, and a small grade is given for correct answers. Such tests can get up to 80% or more of the class to read the relevant material before class. (Heiner, Banet & Wieman, 2014)

Before introducing Peer Instruction in his courses, Mazur was giving traditional lectures to his students. In student evaluations, he was rated very highly. However, he also saw comments such as, "Mazur is a great prof., but physics still sucks!" After introducing Peer Instruction, student learning as measured by the diagnostic instruments increased dramatically. He says, "I have moved from being the sage on the stage to the guide on the side."

Physics by Inquiry

In post-secondary institutions, it is common for courses to have tutorials or recitation sections in addition to the lectures. In the physical sciences it is also common to have a laboratory. In courses using reformed pedagogy the tutorials, perhaps combined with the labs, are centred on students working together in small teams on conceptually-based activities, often involving physical apparatus and/or simulations. Often these are staffed by graduate student Teaching Assistants, who are trained in using a guided-discovery method of instruction, in which their role is to ask questions of the team to help guide them to the correct answer, but to avoid just answering a question.

A leading proponent of this type of instruction is Lillian McDermott at the University of Washington, who has developed a large number of activities. (McDermott & Shaffer,

2002) A master physics teacher, Edwin F. Taylor of MIT, commented that McDermott has "raised putting the student into a state of cognitive dissonance into an art form."

Studio Physics

A somewhat more radical approach to reforming the pedagogy of a course is to combine the classes, tutorials, and labs into a single entity. This is the approach taken by Priscilla Laws at Dickinson College (Laws, 2004), Joe Redish at the University of Maryland (Redish, 2003), Bob Beichner at North Carolina State http://scaleup.ncsu.edu and others. At the last count, over 150 institutions have adopted or adapted this approach.

The same pedagogy is being widely implemented in a variety of courses besides physics. Often these courses are called "Inverted" or "Upside-down" or "Flipped" classrooms: http://www.jstor.org/stable/1183333?seq=1

Threshold Concept 2: Uncertainty in Physical Measurements

Virtually every number used to describe the physical universe is uncertain. Learning to quantitatively deal with these uncertainties is part of the craft of an experimental scientist, both in the social sciences and the physical sciences. We pay special attention to teaching data analysis and uncertainties in many of our courses and teaching laboratories. The study of uncertainties is also called "error analysis". The international definition of measurement uncertainty is provided by the International Organization for Standardization (ISO) as the "parameter associated with the result of a measurement that characterizes the dispersion of the values that could be reasonably attributed to the measurand". (ISO, 1993) Our interest in doing a study on the concept of experimental uncertainty was motivated by the idea of comparing the assessment given in the Threshold Concepts literature with our own facts.

Wilson et al. have identified the measurement uncertainty as a **Threshold Concept in Physics**. (Wilson, Akerlind, Francis, Kirkup, McKenzie, Pearce & Sharma, 2010). The identification process took place in a one-day brainstorm meeting with five physicists from four Australian universities. The process assessed all the characteristics of a threshold concept: transformative, integrative, irreversible, boundary-making and troublesome. They found that the measurement uncertainty meets all of them. It is a common fact in the threshold concepts literature that instructors tried to use their own experience to assess students' difficulties in grasping troublesome concepts. Some of these concepts were being carefully identified to be threshold. According to Wilson, there are 5 stages of understanding of uncertainty, shown in Table 1.

Wilson carried out semi-structured interviews out with 24 randomly selected first year students from four universities. Students were asked to compare data sets, assess data spreads and identify factors that contributed to data scatter. Wilson's study suggested that very few students were able to quantify coherent ideas about data spread, but no quantitative data were provided to support this conclusion.

Our study

At the Department of Physics, University of Toronto, we introduce the experimental uncertainty in first year laboratories and Practicals settings. We teach: distribution of values in repeated experiments, types of errors, mathematical manipulations, etc., several times in the first and second year.

Table 1. Stages of understanding of the experimental uncertainty.

Stage 1	No conception of uncertainty, no thought of it in relation to experimental outcomes
	"I did an experiment and got this answer which is correct!"
Stage 2	Uncertainty is seen as mistakes
	"I did an experiment twice and got a different answer every time so I probably made a mistake or my instruments are broken"
Stage 3	Uncertainty is seen as a mean of quantifying how wrong you are
	"I know the right answer from the book, so my measurement is wrong"
Stage 4	Uncertainty is seen as something that must be planned for
	"I have to take many measurements in order to assess the uncertainty"
Stage 5	Uncertainty is a comprehensible, quantifiable result
	"I have to calculate the mean value and quantify the spread of variables"

In the second year of study, we teach the theory of uncertainties again, in a lab course environment (PHY224H). We introduce new elements and we use computation to implement the advanced concepts. Students do a number of specially designed exercises aimed at linking the theory of Error Analysis with practical experimental situations (Serbanescu, Kushner and Stanley (2011)). In order to assess students' knowledge, two Error Analysis tests were used at six weeks interval (pre- and post-instruction). The data discussed below were taken in 2013.
The tests included five questions: the first two were conceptual and carried 1 grade each. The others were numerical problems with 4 grades each. The tests were each worth 10% of the final grade of the course.

Experimental Uncertainty as a Threshold Concept (TC)

The following TC Question was identical in both tests. It was written by following the stages of understanding of uncertainty found by Wilson, A. et al. (2010) and presented in Table 1. Stages 2 to 5 correspond to options a) to d), below:

"How would you define the experimental uncertainty? Choose the statement that applies best in your opinion:

a) Uncertainty quantifies the mistakes you do
b) Uncertainty quantifies how wrong you are
c) If you make sufficient repeated experiments you can determine the uncertainty
d) Parameter attributed to a measurement which quantifies the variability in the method."

Analysis

Students' answers to the TC Question were correlated to the test grades. Records missing one of the two tests were deleted. The final sample size was 70.

Figures 4 and 5 show the results of the analysis as boxplots. The largest rectangle spans the lower and upper quartiles, and the horizontal line inside the box is the median. The vertical lines above and below the box extend to the greatest/smallest value that is less/greater than a heuristically defined cut-off. The cut-off is the median plus or minus 1.5 times the inter-quartile range.

Figure 4 Boxplots of Test 1 (pre-test) grades over answers to the TC Question. Answers a) - d) mean: a) = least knowledge to d) = most knowledge.

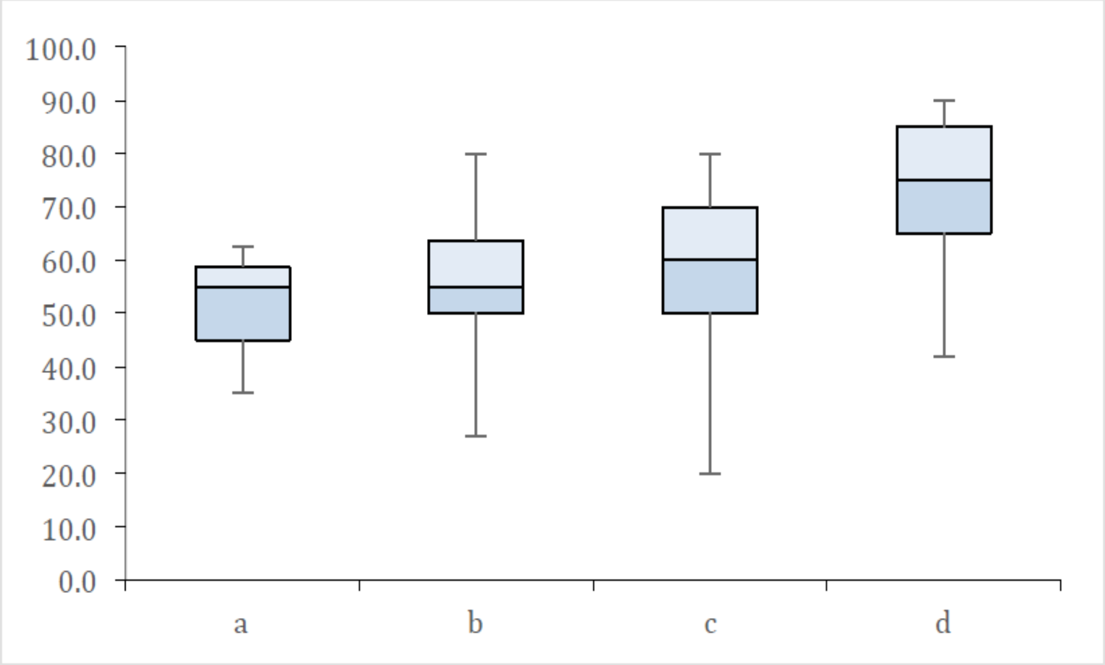
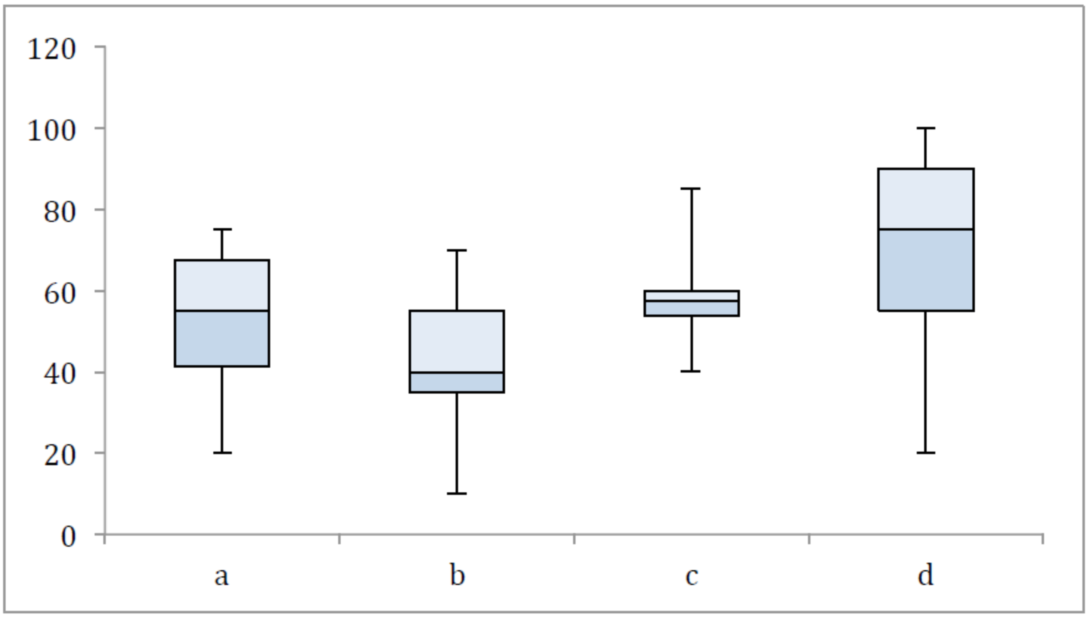


Figure 5. Boxplots of Test 2 (post-test) grades over answers to the TC Question. Answers a) - d) mean: a) = least knowledge to d) = most knowledge.



Further analysis and comments

We used pre- and post-tests with multiple questions in testing the TC Question. This is a methodology characteristic for PER. We did not interview students individually. A comparison between the answers to the TC Question in the pre- and post-tests reveals that the number of students who answered a) or b) stayed constant (17) regardless the enhanced instruction. On the other hand, the number of students who provided the right answer (d) increased from 36 to 45.

This was unexpected: the data presented in Figures 4 and 5 apparently show that the intensive instruction that took place between Test 1 and Test 2 did not have a significant effect on students' understanding of the concept of uncertainty, as reflected in the TC question.

30 students (42.8% of the class size) provided the right answer to the TC question in both tests. This group scored better than the class average in each of the two tests.

In testing the experimental uncertainty as a TC, we applied the PER method of multiple choice written tests (pre- and post-tests). Wilson's assessment cannot be proved clearly through this method.

To validate the experimental uncertainty as a TC, transformative thinking has to be revealed in real time. Multiple choice tests, based on identifying key elements from a dry set of definitions are not the right tools to do it.

We didn't interview the students, but fresh data (Fall 2016) provided a different insight into students' reasoning as we modified the TC Question to allow for a detailed answer in writing. In order to further try to validate the experimental uncertainty as a TC, a mixed methodology has to be used: students who performed poorly have to be interviewed, practical tasks may to be used to assess the newly acquired knowledge, and the question has to be rephrased. We noticed that the TC Question discussed above, taken from Wilson's theory, rather revealed the constant capability of better students to carry a coherent discourse.

Pedagogy 2

Piaget described the cognitive development of young people as consisting of four stages (Inhelder & Piaget, 1958):

1. Sensorimotor (birth - 24 months). Learns that he/she is separate from the external world. Learns about object permanence.
2. Pre-operational (2 - 7 years). Can represent objects as symbols which can be thought of separately from the object. Can "make believe." Wants the knowledge of knowing everything.
3. Concrete Operational (7 - 11 years). Can reason logically about concrete events or objects. Acquires concepts of conservation of number, area, volume, and orientation.
4. Formal Operational (11 - 17 years and onwards). Can reason logically about abstract formal concepts. Can reason with ratios. Can do separation and control of variables. Can think about different points of view or reference frames. Can think about thinking.

The ability to use the ways of thinking, the operations, associated with Formal Operations is clearly necessary to do physics in particular and science in general. However, as Arnett wrote: "research has shown that not all persons in all cultures reach formal operations, and most people do not use formal operations in all aspects of their lives". (Arnett, 2005)

As an example, here are two algebra problems:

Problem C

$$x = y + 3$$
$$x + y = 17$$

Solve for x and y.

(Answer: x = 10, y = 7)

Problem F

> Xavier is three years older than Yolanda. The sum of Xavier and Yolanda's ages is 17. How old are Xavier and Yolanda?
>
> (Answer: Xavier is 10, Yolanda is 7)

The manipulations to solve Problem C, little more than pushing symbols around on a piece of paper with a pencil, require only Concrete Operations. However, casting Problem F into the form of Problem C requires the type of abstraction that is a characteristic of Formal Operations. Of course, many if not most physics problems involve the same type of abstract thinking when casting a physical situation into a set of equations.

Lawson has developed a 24-question Classroom Test of Scientific Reasoning (CTSR) to probe whether students are at a Formal Operational stage of development. (Lawson, 1978) Giving the CTSR to students in introductory post-secondary physics courses shows that many of them are not capable of demonstrating Formal Operational ability. (Coletta, 2015. Harrison, 2014) There is also a positive correlation between performance on the CTSR and gains on the FCI for students Loyola Marymount University. (Coletta & Phillips, 2005). Coletta, Phillips, and Steinert (2007) added data on a positive correlation for students at Edward Little High School, Diff and Tache (2007) found a positive correlation for students at Santa Fe Community College, and Nieminen, Savinainen, and Viiri (2012) found a positive correlation for high school students in Finland.

A particularly troubling result of administering the CTSR is that, as described in Coletta (2015) and Harrison (2014), the male students tend to outperform the female students. There is also a "gender gap" in performance on the FCI. We should emphasise that we believe that the difference in performance is not due to causation, but rather because of cultural influences.

An important question, then, is: can we organize our courses to aid students in becoming Formal Operational, i.e. in learning to "think like a physicist"? There are some studies that indicate that the answer is yes.

In 2000 Lawson et al. demonstrated a normalised gain on the CTSR in a biology course for non-science majors ($p < 0.001$). Traditional courses begin with the theoretical concepts and then progress to more descriptive and hypothetical concepts. Lawson's course reversed the order: they start with the descriptive contents, progress to hypothetical concepts, and then finally to theoretical concepts.

In the United Kingdom a program called Cognitive Acceleration in Science Education (CASE) has had considerable success in stage promotion with students between ages 11 - 14 years. (Adey, 1999). CASE rests on five pillars:

1. Cognitive conflict. This occurs when a student encounters a problem that forces them to confront their misconceptions. Structured help from a teacher or particularly through interactions with other students helps the student gain at least an understanding of the source of the conflict.
2. Construction. The student must actively construct new ways of thinking.
3. Metacognition. The student is encouraged to think about his or her own thinking.
4. Concrete preparation. Just giving a student a cognitively challenging task is not enough. First there must be a phase of preparation in which the language and any apparatus to be used are introduced.
5. Bridging. The ways of thinking developed in a particular context must be linked to other contexts in science and experiences in real life.

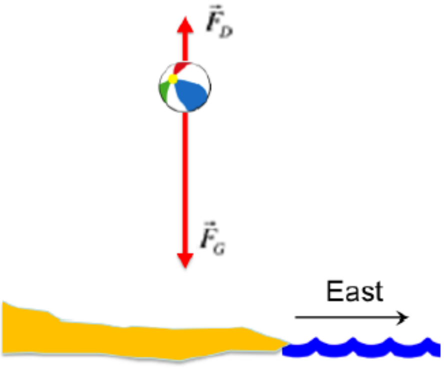
There is a video of CASE in action that nicely demonstrates how it is implemented. It is available at:
http://archive.teachfind.com/ttv/www.teachers.tv/videos/cognitive-acceleration.html
The similarity to the stages of liminality http://www.ee.ucl.ac.uk/~mflanaga/popupLiminality.html necessary for a student to pass through a threshold are striking.

As discussed above in the Pedagogy 1 section, interactive engagement pedagogy is already implementing the first 2 steps of CASE. Coletta (2015) describes explicit attempts to implement the other steps in post-secondary physics courses. The results were spectacular: Post-Course results on both the CTSR and the FCI were greatly improved. Further, the gender gap almost completely disappeared.

There is another issue that we are beginning to think about in our instructional strategies. When a researcher, whether in the sciences or other fields, begins thinking about some situation, they naturally use the language that is available to them. In physics, when people began thinking about the effect of forces acting on objects, they used the word "work" to describe the situation. Eventually it became clear that the description should be the force acting on some object times the distance over which the force acts. This became the physics definition of the word work. Thus, in physics if you hold a stationary heavy object you are not doing any work since the distance is zero, although in everyday life you would certainly say you were doing work and your muscles are becoming fatigued.

Another example is a question recently posed to a large (-1000 student) introductory physics course at the University of Toronto.

> A large, light beach ball is falling towards the beach on a windless day. The force of gravity of the ball, $\vec{F}_G$, is greater than the upward drag force from the air, $\vec{F}_D$. Which of the following directions is closest to the direction of the net force $\vec{F}_{net} = \vec{F}_G + \vec{F}_D$ on the ball?



A. North
B. East
C. South
D. West
E. The net force makes an angle of 90° with respect to all four of these directions.

(Answer: E)

The student answers were:

A. 1.2%
B. 1.7%
C. 57.4%

D. 0.8%
E. 38.9%

Clearly almost 60% of the students associate the word "down" with the direction South.

The problem of the signifier, such as in our case "work" or "down", not matching the thing that is signified has recently been explored by Land, Rattray, and Vivian (2014) We hope in future to use some of the insights of this work to further explore the difficulties our students experience with assimilating and using the new connections between the signifier and signified that we are asking them to master.

Finally, there is a further issue inherent in reformed instructional methods. As we stated at the Threshold Concept 1 section, the process of learning with these methods is time consuming. This means that typically the content of a reformed course must be reduced. However, in a typical traditional course the diagnostic instruments show that although the instructor may have covered a lot of material in the classes, the students didn't actually understand a lot of that material. As Redish and Hammer (2009) commented, "The idea that one has to cover a particular set of material, whether or not the students understand it, seems peculiar, but it is widespread."

Conclusion

We have discussed physics education from a few different perspectives. First, of course, is the framework of *threshold concepts, troublesome knowledge*, and *liminality*. We have found this framework to be extremely useful in thinking about effective pedagogy. However, Physics Education Research provides insight into student learning from a somewhat different perspective, and the methodology of that research has led to some evidence-based instructional strategies that are proven to work in physics, and which we believe can be easily adapted to other disciplines. Piagetian taxonomy leads us to think about learning in yet another way, similar but different than the ones already mentioned. Finally, we have investigated the mechanisms that lead to the sorts of transformative thinking that are necessary for a student to pass through the threshold to a fuller, deeper, and more satisfying understanding.

It is satisfying to note that although all of these different approaches to learning and instruction tend to use different vocabularies and methodologies, at the end of the day the conclusions drawn from them are really very similar to each other.

Acknowledgements

We thank Jason J.B. Harlow, Dept. of Physics, Univ. of Toronto, for sharing and allowing us to use his "beach ball" question in the Pedagogy 2 section.

References

Adey, P. (1999). The Science *of* Thinking, and Science *For* Thinking: A Description of Cognitive Acceleration Through Science Education. *Innodata Monographs – 2.* The International Bureau of Education: UNESCO.

Arnett, J. J. (2005). *Adolescence and Emerging Adulthood: A Cultural Approach* (3rd ed). Upper Saddle River: Prentice Hall.

Coletta, V. (2015). *Thinking in Physics: Strategies for Improving Scientific Reasoning, Conceptual Understanding, and Problem Solving in Introductory Physics.* New York: Pearson.

Coletta, V.P. & J.A. Phillips (2005). Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability, *American Journal of Physics*, *73*(12), 1172-1182.

Coletta, V. P., Phillips, J. A., & Steinert, J. J. (2007). Why You Should Measure Your Students' Reasoning Ability, *The Physics Teacher*, *45*, 235-238.

Diff, K., & Tache, N. (2007). *From FCI To CSEM To Lawson Test: A Report On Data Collected At A Community College,* available at http://www.compadre.org/portal/items/detail.cfm?ID=9054&Relations=1.

Harrison, D. M. (2014). *Factors correlated with students' scientific reasoning ability in an introductory university physics course,* available at http://www.upscale.utoronto.ca/PVB/Harrison/CTSR_Factors/CTSR_ Factors.pdf

Heiner, C. E., Banet, A. I., & Wieman, C. (2014). Preparing students for class: How to get 80% of students reading the textbook before class, *American Journal of Physics*, *82*(10), 989-996.

Hestenes, D., Wells, M., & Swackhammer, G. (1992). Force Concept Inventory, *The Physics Teacher*, *30*, 141-157.

Inhelder, B., & Piaget, J. (1958). *The Growth of Logical Thinking From Childhood to Adolescence* (A. Parsons. & S. Milgram, Trans), New York: Basic Books.

ISO (1993), *International Vocabulary of Basic and General Terms in Metrology (VIM),* Geneva, Switzerland. Statement taken from VIM3.9.

Land. R., Rattray, J., & Vivian, P. (2014). Learning in the liminal space: a semiotic approach to threshold concepts, *Higher Education*, *67*, 199-217.

Land, R., Meyer, J. H. F & Smith, J. (2008). *Threshold Concepts within the Discipline*, Rotterdam: Sense Publishers.

Laws, P. W. (2004). *Workshop Physics Activity Guide*, 4 volumes, New York: Wiley.

Lawson, A.E. (1978). The development and validation of a classroom test of formal reasoning, *Journal of Research in Science Teaching*, *15*(1): 11-14. Available from https://modelinginstruction.org/wp-content/uploads/2013/06/LawsonTest_4-2006.pdf

Lawson, A. E., Aklhoury, S., Benford, R., Clark, B. R., & Falconer, K. A. (2000). What Kinds of Scientific Concepts Exist? Concept Construction and Intellectual Development in College Biology, *Journal of Research in Science Teaching*, *37*(9), 996-1018.

Lippmann, R. (2003). *Students' Understanding of Measurement and Uncertainty in the Physics Laboratory: Social construction, underlying concepts, and quantitative analysis*, PhD Thesis, University of Maryland.

Mazur, E. (1997). *Peer Instruction: A User's Manual*, Upper Saddle River: Prentice-Hall.

McDermott, L. C., Shaffer, P. S. and the Physics Education Group (2002), *Tutorials in Introductory Physics*, Upper saddle River: Prentice-Hall.

Newton, I. (1687). *The Principia* (I. B. Cohen & A. Whitman, Trans.1999), Berkeley: University of California Press.

Nieminen, P., Savinainen, A., & Viiri, J. (2012). Relations between representational consistency, conceptual understanding of the force concept, and scientific reasoning. *Physical Review Special Topics - Physics Education Research, 8*(1), 010123.

Redish, E. F. (2003). Teaching Physics with the Physics Suite. New York: Wiley.

Redish, E. F., & Hammer, D. (2009). Reinventing college physics for biologists: Explicating an epistemological curriculum. *American Journal of Physics, 77*(7), 629-642. doi: 10.1119/1.3119150

Serbanescu, R. M., Kushner, P. J., & Stanley, S. (2011). Putting computation on a par with experiments and theory in the undergraduate physics curriculum. *American Journal of Physics, 79*(9), 919-924. doi: 10.1119/1.3593296

Wilson, A., Åkerlind, G., Francis, P., Kirkup, L., McKenzie, J., Pearce, D., & Sharma, M, D. (2010). Measurement Uncertainty as a Threshold Concept in Physics. In *Proceedings of the 16th UniServe Science Annual Conference,* 29 September-1 October 2010, pp. 98-103. University of Sydney: Australia