

Assessing Students' Deep Conceptual Understanding in Physical Sciences: an Example on Sinking and Floating

Ji Shen¹ · Ou Lydia Liu² · Hsin-Yi Chang³

Received: 19 March 2015 / Accepted: 2 August 2015 / Published online: 23 September 2015
© Ministry of Science and Technology, Taiwan 2015

Abstract This paper presents a transformative modeling framework that guides the development of assessment to measure students' deep understanding in physical sciences. The framework emphasizes 3 types of connections that students need to make when learning physical sciences: (1) linking physical states, processes, and explanatory models, (2) integrating multiple explanatory models, and (3) connecting scientific models to concrete experiences. We carried out a 2-phase exploratory study that helped further develop and refine the framework. In the first phase, we developed 3 items on sinking and floating and pilot tested them with 18 undergraduate students. Analysis of student responses revealed various student misconceptions and the different connections students made among science ideas. Based on the findings, we revised the assessment, modified the instruction, and collected data from another cohort of 26 students. The second cohort of students showed significant improvement of understanding of sinking and floating after instruction. Implications and limitations of how our assessment framework can be used to improve students' conceptual understanding in science are discussed.

Electronic supplementary material The online version of this article (doi:10.1007/s10763-015-9680-z) contains supplementary material, which is available to authorized users.

✉ Ji Shen
ji.shen1221@gmail.com

Ou Lydia Liu
lliu@ets.org

Hsin-Yi Chang
hsinyichang@nknuc.nknu.edu.tw

¹ Department of Teaching and Learning, University of Miami, 324B Merrick, 5202 University Drive, Coral Gables, FL 33124, USA

² Educational Testing Service MS 16-R, 660 Rosedale Rd, Princeton, NJ 08541, USA

³ Graduate Institute of Science Education, National Kaohsiung Normal University, No. 62, Shengjhong Rd., Yanchao District, Kaohsiung 824, Taiwan

Keywords Explanatory models · Modeling assessment · Physical science · Sinking and floating · Transformative

Introduction

It is challenging to design assessment items that effectively capture students' prior experience and deep understanding of science concepts. The complicated dimensions of science learning contest many traditional approaches in terms of assessing for deep understanding. An individual item in a science assessment instrument is typically designed to measure one aspect of the target science concept. However, science learning is such a complex and dynamic process that multidimensional knowledge and abilities of the students should be considered to achieve the goal of assessment for learning (National Research Council [NRC], 2001, 2014). One such core ability is making meaningful connections among ideas. It is equally challenging for teachers to extract useful information from student responses to provide quality feedback to students and improve instruction. Much has been documented about the important role of teacher-student and student-student interactions in classrooms in enacting formative assessments. For example, drawing on a wide range of research, Black and Wiliam (2009) elaborated on a theory of formative assessment that centered around the interaction among the teacher, the learners, and their peers. Facilitating a productive interaction between the teacher and the students starts with a good mechanism of knowing what students know, including the types of conceptual connections they can make (NRC, 2001).

To address these challenges, we propose an assessment framework that may help design assessments to better capture students' deep conceptual understanding in physical sciences and provide the instructor with specific information to improve their learning. In the following, we first elaborate on the learning theory on which our assessment framework is based. We then report an exploratory study to showcase how to use the framework to develop specific assessment items and associated scoring techniques using the topic of sinking and floating.

The Transformative Modeling Framework

We take the position that a good assessment needs to be built upon a sound theory on how students learn (NRC, 2001). There have been many good assessment approaches in science education in this regard (NRC, 2014). Among these, two approaches informed our assessment design most significantly: DIAGNOSER and Knowledge Integration. The former acknowledges the many facets a student knows about a specific science topic and attempts to diagnose the status of these knowledge facets (Hunt & Minstrell, 1994; Minstrell, 2000); the latter focuses on understanding how a student makes connections among relevant science ideas or facets (Liu, Lee & Linn, 2011; Shen & Linn, 2011).

Our assessment framework was driven by the transformative modeling (TM) framework, developed to capture (and promote in corresponding instruction) students' deep understanding of abstract and complex topics in science. The TM framework originated

from model-based learning and teaching perspectives that encourage students to develop, use, and evaluate models to describe, explain, and predict science phenomena (e.g. Chinn & Samarapungavan, 2008; Clement, 2000; Frederiksen, White & Gutwill, 1999; Lehrer & Schauble, 2006; Namdar & Shen, 2015; Schwarz et al., 2009).

The TM theory describes students' conceptual learning as a process of making sense of the world through a chain of operations on physical and symbolic materials that constitute explanatory models (Shen & Confrey, 2007; Shen & Jackson, 2013). An *explanatory model* in this paper is defined as a cohesive set of elements that students use to describe, explain, predict, and communicate with others about a natural phenomenon, an event, or an entity. Consistent with the Knowledge Integration framework, TM emphasizes the logical connections among the explanatory elements that students use, and acknowledges the diverse resources and origin of these elements. These explanatory elements may seem fragmented to a novice, similar to the DIAGNOSER perspective.

The central argument of the TM theory is that the elements and their connections in an explanatory model are constructed so that they can be *transformed* into different media and representations under a variety of contexts, but still convey shared meanings that cut across these contexts. As such, TM frames a transformation as the most fundamental process that links and explains other modeling processes. For example, the TM perspective views the process of constructing a new model as transforming existing models and new observations to a qualitatively different construct. Under the TM framework, learning scientific knowledge is to acquire the capacity to transform a variety of constructs through physical and symbolic materials available in different contexts. This includes transformation between alternative models (Shen & Confrey, 2007, 2010).

Figure 1 portrays structurally how we consider transformation between alternative explanatory models using a generic example. Selected states (S1, S2, S3) in a given context (A) are connected via processes (P12, P23) and can be fully or partially explained by different explanatory models M(X) and M(Y). The diagram in Fig. 1 illustrating the key states, processes, and explanatory models (SPM) and their connections is named the SPM diagram. We identify three important aspects of conceptual understanding that facilitate quality transformations in physical sciences: (1) linking physical states and processes of the target domain with explanatory models, (2) integrating multiple explanatory models, and (3) connecting scientific models to concrete experience. We elaborate on these aspects one by one in the following sections.

Linking Physical States and Processes with Explanatory Models

The idea of linking physical states and processes with explanatory models arises from answering three basic questions concerning a scientific inquiry. A student may start an inquiry process with the very basic question of "What is happening?" The student may transform the observations into data, inscriptions, and digital records. During such a transformation, a rich set of scientific observations of the key states helps the student identify the interacting elements and the relevant state variables that constitute the key properties of the target system. Then, the student may ask "How does it happen?" They need to understand that a physical process of a system simply consists of a set of

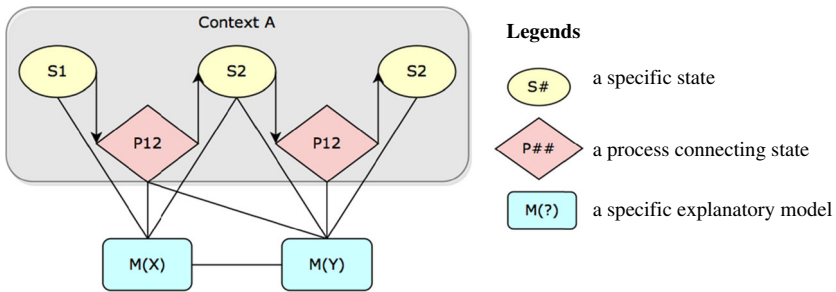


Fig. 1 The types of connections highlighted in the transformative modeling framework: selected states ($S1$, $S2$, and $S3$) and processes ($P12$ and $P23$) in a particular context (A) can be partially explained by two different models, $M(X)$ and $M(Y)$

changing states over time. Many of these processes lead to a physical equilibrium (e.g. thermal equilibrium) in which the system maintains a stable state or to an emerging state (e.g. superconducting) in which the system appears in a qualitatively different state. Eventually, the student may ask “Why does it happen?” An explanatory model is then applied or constructed to theorize the patterns of how the key states emerge and transform (e.g. statistical mechanics is used to explain phase change).

The what-how-why question description previously presented is only a simplified example of how learning can take place during scientific inquiry. In reality, students may not follow this linear sequence. Instead, they may go back and forth in exploring the what, how, and why of a phenomenon. For instance, a student may be curious about the phenomenon of a balloon sticking to a wall and wonder *why* (Shen & Linn, 2011). The student may recall that an electrical *state* of an object may be positively charged, negatively charged, or neutral, and may also have learned the *explanatory model* of why charges move and how like charges repel and opposite charges attract. Therefore, the student may initially explain that when a balloon sticks to a wall this phenomenon is the result of both objects having an opposite charge. However, this explanation may not be true because the student does not pay attention to *what* the initial electrical *states* of the two objects are (e.g. the wall is neutral initially) and the *process* of *how* they stick to each other. Therefore, a comprehensive understanding integrates the initial and final states of the objects involved, the processes of how the objects are charged and how induction occurs, and why the interaction between charges can explain these states and processes.

Integrating Multiple Explanatory Models

Building on different perspectives and experiences, a learner may draw on different models to explain the same phenomenon (Shen & Confrey, 2010). Some of these models may be consistent with the scientific models whereas other models may be in conflict with the accepted normative understanding of how things work. Students typically start with simple and concrete models such as physical models, and then advance to more abstract and theoretical models, such as mathematical models.

We argue that when using more advanced models to explain complex science topics, students need to integrate and transform the models they have already learned in order to understand an observed phenomenon. For instance, to explain and predict how

charged and/or neutral materials interact with each other, students may use a charge-based model that uses two types of charges, positive and negative, to account for electrostatic attraction and repulsion; a particle-based model that uses the movement of particles and their interactions to account for electrostatic interactions; or an energy-based model that uses an electric field, the conservation of total energy, and transformation of energy to account for phenomena caused by static electricity. A deep understanding of static electricity provides students with an opportunity to integrate these models so that they can explain an observed phenomenon (e.g. a balloon sticking to a wall) and allows students' understanding of the phenomenon to be transformed (Shen & Linn, 2011).

Connecting Scientific Models to Concrete Experience

Students learn science more effectively and meaningfully if their learning is anchored to their personal experiences (Linn, Lee, Tinker, Husic & Chiu, 2006). In the TM framework, applying a scientific model to explain a set of everyday experiences involves transforming the model in various forms in order to make connections to one's otherwise isolated sets of experience. This transformation is an indicator of students' ability to develop an integrated understanding of the underlying concepts. In this way, students are more ready to recognize the important features of a new context and transform existing knowledge to interpret the new context. For example, many instances of everyday experiences are related to static electricity, e.g. feeling an electric shock when grabbing a metal doorknob and having fluffy hair when combed on a dry day. An in-depth understanding of electrostatics not only requires students to make sense of these phenomena inside and outside of the classroom but also helps them transform the world around them (Shen & Linn, 2011).

Exploratory Study on Sinking and Floating

In the following section, we report an exploratory study through which we illustrate how the TM framework can be used to assess students' deep understanding of the concepts of sinking and floating. In this study, we ask three central questions: (1) How can students' understanding of different explanatory models using TM assessment items help them explain the phenomenon of sinking and floating? (2) How do students integrate different explanatory models to explain sinking and floating? (3) What lessons can be learned from the TM assessment framework and its implementation in the case of sinking and floating and how such lessons can benefit the assessment design in physical sciences?

Transformative Modeling Assessment Items on Sinking and Floating

Sinking and floating is a well-researched topic (e.g. Ruiz-Primo & Furtak, 2007). To develop the TM assessment items for this topic, we first followed the DIAGNOSER fashion and summarized the key facets regarding the states and processes of sinking and floating, as well as the basic definitions using the density/force model. Using this knowledge base and following the TM framework, we constructed a total of five

constructed-response items about sinking and floating. The specific items and corresponding SPM diagrams and the knowledge base table can be found in the [electronic supplemental materials](#) on the IJSME website.

Student Participants and the Instructional Context

We conducted an exploratory study with two convenience samples of undergraduate students who were in a middle grades teacher preparation program in a public university in southeast USA. We did not include the two cohorts for comparison purposes. The first cohort was used to pilot test the initial three TM assessment items and to reveal the types of responses these items could elicit. We expected that the data analysis would take much longer time than normal. The analysis results from the first cohort were used to inform the design of the assessment and modification of instruction for the second cohort.

These two cohorts of students were enrolled in an algebra-based, physical science course in two consecutive academic years. The course met 3 days per week, 2 h per day, for 15 weeks. The first cohort consisted of 18 students (17 female, 1 male); the second cohort consisted of 26 students (21 female, 5 male). The first author taught the course. Both cohorts spent approximately 30 h in learning Newtonian mechanics and 6 h in learning, sinking, and floating including a lab on Archimedes's principle. Students in both cohorts had learned the density model in pre-college science classes, as revealed by their responses to the questions in class. This study focuses on examining their responses to the five TM assessment items, which were administered differently for the two cohorts (Table 1).

The first cohort started the sinking and floating unit by identifying and discussing the forces and directions exerted on objects floating, suspending, or sinking in water. They received a lecture that covered key concepts including the connection between density and buoyancy in sinking and floating. They observed the phenomena of a Cartesian diver as a class demonstration. They also conducted a lab confirming Archimedeian's principle (see [electronic supplementary materials](#)). They reviewed the sinking and floating assessment instrument (Yin, Tomita & Shavelson, 2008) designed for middle school students in class. Additional reading related to sinking and floating was assigned from the textbook *Conceptual Physics* (Hewitt, 2005). At the end of instruction, the students took an in-class quiz that tested their understanding of three items: Released-block, Cartesian-diver, and Two-ball.

Table 1 Implementation of assessment items in two cohorts of students

	Pre-measure	Treatment	Post-measure	Purpose of study
Cohort 1 (<i>n</i> = 18)	N/A	Sinking and floating unit	Items Released-block, Cartesian-diver, Two-ball	Development of TM assessment items
Cohort 2 (<i>n</i> = 26)	Item Fish	Sinking and floating unit and 2 items as homework (Released-block, Cartesian-diver)	Item Submarine Item Two-ball	Implementation of TM to improve instruction and learning

Based on the analyses of the responses from the first cohort, the following changes were made previous to our investigation of the second cohort: (1) we refined the three assessment items, (2) added two new items, (3) changed the way in which we implemented the items, and (4) modified instruction. In terms of item revisions, we added an explicit instruction “use both the density model and the force model” in all the items. Since the three items administered to the first cohort did not explicitly ask students to use both models, students tended to use only one model in a single item. We decided to add an addition instruction to motivate students to use both models, so that we could evaluate students’ ability to integrate both explanatory models. In terms of implementation, we added the item Fish as a pre-test right before the students received the instruction because the Fish item presented a familiar context to all students because it was covered in their reading. We assigned the items Released-block and Cartesian-diver as unit homework because we prepared in-class, hands-on activities to help students go over these two items (see section “Modified Instruction for the Second Cohort” to understand how we modified the instruction after showing the results of the first cohort). We used the items Submarine and Two-ball as a post-test, about 2 weeks after the conclusion of instruction of the unit, with the hope that the students could transfer what they have learned to relatively new contexts.

Data Analysis

We collected students’ written responses and recorded them in Excel files. For the first cohort of students, we coded the types of models that the students used into four categories: density model (D), force model (F), alternative model (A), and mixed model (M). A density model uses the concept of density to explain sinking and floating; a force model uses Newtonian mechanics to explain sinking and floating; an alternative model is one that is neither the density model nor the force model; and a mixed model combines the use of both the density model and the force model. This coding was not applied to the second cohort as they were explicitly asked to use both the density model and the force model.

We compared each student response with the normative SPM diagrams (see [electronic supplementary materials](#)). The diagrammatic representations helped us visualize students’ understanding in terms of the SPM connections. Finally, we assigned a SPM score (max of 5) to each response using a six-level rubric (see [electronic supplementary materials](#)). Adapting a Knowledge Integration rubric structure (Shen & Linn, 2011), the SPM scoring rubric highlights the connections valued in the TM framework. The authors first coded a set of randomly selected sample responses ($n = 30$) independently and found the percent agreement to be higher than 75 %. The authors then discussed any discrepancies that were found in the sample responses and reached a consensus on a coding procedure and a rubric that could be used to interpret students’ responses. Using the coding procedure, the first author coded the rest of the responses.

Findings and Discussion

Findings from the First Cohort. We implemented three items (Released-block, Cartesian-diver, Two-ball) in a quiz and gave this quiz to the first cohort after the

instruction of the unit. The Released-block item tended to elicit the use of only the force model among 11 students while 7 students chose to use a mixed model approach. The Cartesian-diver item and the fourth sub-question regarding the Two-ball item tended to elicit among students the use of the density model among half of the students while the other half of the students used a mixed model. For the Two-ball item, 11 students used the density model and 7 students used either the force model or an alternative model.

The majority of the students were able to apply the density model to explain sinking and floating. For instance, all students answered the second sub-question of the Two-ball item correctly, indicating that all the students understood that a change of the fluid environment does not change the density of an object, and that a sinker has a greater density than a floater with the same volume. However, most students did not demonstrate a mastery of the force model in the context of sinking and floating. The SPM analyses revealed several weaknesses in students' understanding.

For the Released-block item, a majority of the students noticed the initial and the final floating states and understood their implications in terms of density, but many of the students failed to understand the process of how the block came to a full stop. Only two students explicitly mentioned the process when the block emerged from the water. Both students used upward acceleration to explain why the block passed its equilibrium position. For instance, one student explained:

When it (the block) reaches the surface, due to its upward acceleration, it comes out of the water slightly higher than its normal floating depth. The block then pops up and down on the water surface until it stops.

This kind of interpretation may originate from the notion that force is proportional to velocity and that no-force implies no-motion, i.e. the "impetus" theory that has been well documented in physics education (e.g. diSessa, 1993; McCloskey, 1983).

For the Cartesian-diver item, although many students connected the process of changing water levels inside the test tube to its sinking or floating state, their responses revealed a lack of integration of the force model and the density model. Some students had the misconception that the buoyancy of an object is inversely proportional to its density. For example, one student who used the mixed model explained the sinking and floating of the test tube stated:

...When you release the bottle ... the extra water in the test tube is released, decompressing the air within the tube. With less water, the tube becomes less dense, so its buoyancy increases, allowing it to float back up to the surface, since its density is less than the water surrounding it.

It was very likely that this student had the concept of "relative buoyancy" in mind: i.e. equating buoyancy to the ratio of buoyancy over its weight.

Although all of the students answered the second sub-question of the Two-ball item correctly, their explanations revealed problems in understanding the concept of buoyancy. For the first sub-question, eight students stated that ball A had greater buoyancy because it was floating. These students confused the sinking and floating states with the explanatory model; they equated buoyancy, a construct in an explanatory model, with the state of floating. They may think that buoyancy is a property of an object, which is

opposed to the interaction between an object and its surrounding fluid. Some students' responses to the Two-ball item also demonstrated the misconception that buoyancy of an object is inversely proportional to its density. For the third and fourth sub-questions, none of the students were able to provide a complete and correct answer. All students predicted that the buoyancies of the two balls would change in the same way. Four students thought that the buoyancies of the balls would decrease; eight students thought that the buoyancies of the balls would increase; and five students thought that the buoyancies of the balls would stay the same. One student thought that the buoyancies of the balls would change but did not indicate how. This reflects a common weakness among students when they attempt to resolve complex problems. Such students tend to ignore associated changes when one change is introduced into a part of a physical system. In the case of this study, many students only paid attention to one changing variable (density of the liquid), but ignored the fact that the submerged volume of the floater also changed. On the fourth sub-question, most students predicted that both balls would float "better" or "higher." These students did not distinguish the two balls as representing two different cases. A typical student's response is "the two balls will float better in the salt water, since the difference between the densities of the balls and salt water is greater, both balls will float higher than they did in plain water." Such a response indicates an inability of the student to understand that the complex state of ball B after adding salt depends on the density of the ball and of the salt water.

In terms of the SPM score, the students did better on the items Released-block and Cartesian-diver, $M_SPM = 3.3$, $SD = 0.05$, and $M_SPM = 3.0$, $SD = 1.0$, than the item Two-ball, $M_SPM = 2.5$, $SD = 0.7$. The SPM scores were based on the combination of student responses to the sub-questions. A SPM score of 3 (out of 5) indicates that a student may partially connect an explanatory model to explain relevant states and processes.

Modified Instruction for the Second Cohort. We recognized that even though many students from the first cohort were able to partially explain sinking and floating using the two models, their performance on the integration of the density and the force model was problematic. Based on the results of the first cohort, we modified the instruction and the assessment implementation, with the hope that the change would help the second cohort of students better integrate the two explanatory models. Specifically, the instructor highlighted the SPM framework by presenting Fig. 1 to the students. Emphasis was made on the sinking or floating states; the processes linking these states; and the application of the two models to explain these states and processes in the sinking and floating context. Similar to the first cohort, students showed inadequate understanding when responding to the pre-test Fish item, the two homework Released-block items, and the Cartesian-diver. The instruction provided feedback to the students to address their misunderstandings. For instance, they did not address the oscillating process of the Released-block item. In response, the instructor showed a video in class to emphasize the nuanced process using slow-motion playback. The instructor also illustrated how the density model and force model could be applied to fully explain the whole process. Also, the students were asked to make a simple Cartesian-diver using a 2-l plastic bottle and a glass test tube in small groups during class, and discussed how they could explain the Cartesian-diver by using the two explanatory models. We asked the students to carefully draw the force diagrams when conducting the Cartesian-diver

activity and the lab confirming Archimedean's principle to help the students visualize the forces.

Findings from the Second Cohort (the Pre- and Post-tests). We used the item Fish as a pre-test and items Submarine and Two-ball as a post-test for the second cohort. In the pre-test, the students showed a naïve understanding of buoyancy (Mean_SPM = 1.7, SD = 0.9). Their naïve understanding of how to integrate density and buoyancy was partially due to their inadequate understanding of the concept of buoyancy itself. Several students were not clear about the direction of buoyancy. As one student claimed “the buoyant force always pushes the fish down.” A couple of other students stated that the buoyant force may point up or down, depending on its density. Another student claimed “buoyancy will help push the fish down when the fish is denser and push the fish up when it is less dense.” Similar to the first cohort, many students thought that buoyancy is a term equivalent to the floating status of an object. One student wrote, for instance, “floating means a greater buoyancy;” another student wrote, “for a sunk fish there is no buoyant force.” Many students incorrectly linked the density model to the force model and thought that density was a measure of weight or mass. They did not understand that density is a measure of an intensive property of a material, whereas a force describes the interaction between two things. Other students thought that a denser object displaces more water, therefore having greater buoyancy.

Students demonstrated advanced understanding when responding to the Submarine item, one of the post-test items. The average SPM for this item is 4.1 (SD = 0.7) and indicated that the majority of students could apply one explanatory model or integrate both models to successfully explain the key states and processes about sinking and floating. Since the content of the post-test Submarine item is similar to the content of the pre-test Fish item, we performed a Wilcoxon signed-rank test and discovered a significance between the results of both items ($W = 325, p < .001$).

When comparing the students' pre-test and post-test responses, we noticed a remarkable improvement in demonstrating their understanding of sinking and floating using the two models. For instance, a student who previously equated density with weight when responding to the item Fish, “By making its force by density greater than the force by buoyancy the fish can sink,” now recognized that

The density of an object is the relationship between its mass and volume (m/v). To move down the water, the submarine increases its density to letting in water, thus increases its mass.... The same idea applies to the buoyancy of the submarine. By increasing its weight (by adding water, it can counter the buoyant force), it sinks....

Another noteworthy result is the mention of the Cartesian-diver case by six students when explaining the motion of the submarine. This shows that the students could transform the Cartesian-diver as a concrete model to explain another phenomenon and, as a result, link two contexts in a meaningful way.

For the Two-ball item, the students also demonstrated an improved understanding: 24 out of 26 students predicted that ball B has a greater buoyancy (recall that the rate was 10 out of 18 for the first cohort); all students understood that ball B has a greater density; the mean SPM score, based on responses to the third and fourth sub-questions,

was 3.7 (SD = 0.7). This result was statistically better than the result for the Fish item as revealed by the Wilcoxon signed-rank test results ($W = 276, p < .001$). There was still much room for improvement though. Many students simply treated the two balls the same and believed that the buoyancy for both balls will increase, which is similar to the belief of the first cohort.

Summary in Response to the Research Questions

This exploratory study illustrates that the TM framework has great potential in guiding the development of assessment that benefits instruction and learning the concepts of sinking and floating. In response to the first question (How can students' understanding of different explanatory models using TM assessment items help them explain the phenomenon of sinking and floating?), our results show that when learning about sinking and floating, many students only focused on static states of sinking or floating but ignored some critical processes that link these key states, e.g. the oscillation process when the block moves out of water in the Released-block item. As a result, students may favor explanatory models that only apply to static states. Overlooking the key processes may hinder students' understanding in linking these processes with more powerful scientific models, e.g. unable to explain why a moving block comes to a full stop.

With respect to the second research question (How do students integrate different explanatory models to explain sinking and floating?), our results reveal that after the students may have learned both density and force models, the students tended to only draw on one model to explain sinking and floating and displayed an unwillingness to use both models. Learning only isolated models over time may further prevent students from developing a coherent understanding of related science concepts. It is important to diagnose the different causes that lead to isolated understanding during instruction, e.g. unable to grasp a key concept in one model; incorrectly linking concepts from different models; and linking different models to different aspects of a phenomenon.

Our third research question was "What lessons can be learned from the TM assessment framework and its implementation in the case of sinking and floating and how can such lessons benefit the assessment design in physical sciences?" First, the sample TM assessment items were able to capture students' ideas of the dynamic processes in sinking and floating in light of multiple explanatory models. This is critical in teaching and learning about physical sciences. Students are asked to predict and then explain a change in a complex yet familiar physical system, which prompts them to make meaningful connections through explaining states and processes of the target phenomenon via multiple explanatory models. Using assessments that measure student ability to connect multiple explanatory models on the same topic provides a new channel to connect otherwise isolated science units.

Second, our study also demonstrates that the diagnostic information extracted from the SPM analyses could benefit the instruction on sinking and floating. Taking advantage of the first-cohort results, we revamped the instruction, which resulted in a significant improvement in students' understanding in the second cohort. Providing scaffolding to draw students' attention to the nuanced processes may improve their observational skills and help them integrate multiple scientific models, as well as connecting explanatory models to concrete experiences. One strategy we employed

was to explicitly address the states, processes, and explanatory models and compare and contrast these models for different cases. We also videotaped some of the physical processes, allowing students to zoom into and playback in slow motion the critical processes and discuss these processes using multiple scientific models. Another strategy was to give explicit prompts in assessments (Davis, 2003), as we did for the second cohort. The instruction in the homework problems asking students to use both models propelled students to think about the integration of the two models. Moreover, we provided timely feedback to students on their homework for them to further reflect on the assessment.

Third, whereas a typical science assessment often asks students to apply an explanatory model in a conventional context (e.g. investigating force and motion in the context of free-falling objects in air), our sinking and floating items expanded the discussion of an object's motion in a non-conventional medium, water. Results from these items revealed that although the students in our study had previously learned force and motion, many of them were not able to link Newtonian mechanics to fully explain the motion involved in sinking and floating. It has been argued that applying and transferring prior knowledge to new contexts are extremely important for students to gain a deeper understanding of the science content (e.g. Bransford, Brown & Cocking, 2000; Bransford & Schwartz, 1999). We stress that it is critical to consider including "non-conventional" contexts for assessing transfer. We further argue that it is equally important to assess knowledge not only in a "transfer" mode but also in a "transformation" mode: i.e. asking students to transform both their knowledge (e.g. restructuring prior knowledge) and potentially the new context (e.g. physically manipulating the sinking and floating status of an object).

Lastly, reflecting on our own experience, we summarize the design principles that can be used to develop future TM assessment items: (a) the items should elicit constructed responses from students to enable them to evaluate their reasoning and justification; (b) the items should present phenomena that involve a process of changing states; (c) the items should present phenomena that trigger multiple facets of knowledge and can be explained by multiple explanatory models (especially the ones covered in school curriculum), and explicitly ask students to apply these models; and (d) the items should present multiple phenomena that can be explained by the same explanatory model.

Conclusion and Limitation

In this study, we propose an assessment framework that focuses on three aspects of conceptual understanding: (1) linking physical states and processes with underlying explanatory models, (2) integrating multiple explanatory models, and (3) connecting scientific models with concrete experience. We illustrate the framework with an exploratory study in the context of sinking and floating. Detailed analyses revealed multiple sources of student misconceptions. Through refined assessment and instruction, a new cohort of students were able to make significant improvements in integrating scientific modeling in explaining why things sink or float.

Given the exploratory nature, there are many limitations of the study. We only included two small and convenient samples to test the items. The five items used in this

study should not be taken as a coherent instrument; more rigorous psychometric studies with large samples should be conducted to make inferences at the instrumental level.

We only applied the TM framework to the topic of sinking and floating. Although we believe that the framework, in principle, can be applied to other physical sciences topics that involve complex processes and multiple explanatory models, this requires hard thinking and a series of empirical investigations. When applying to a new topic, one has to carefully lay out the knowledge facets regarding the key states and linking processes, the specific and relevant explanatory models, and the instances of experience to which students are expected to connect.

Our work is an ongoing contribution to advancing assessment for learning in science classrooms. However, there are still many practical challenges. Although the TM design enables teachers to elicit rich information from their students, the analysis (or scoring) is time consuming. The analysis approach used in this current study may not be practical for teachers to use with students in a classroom setting. Educators who want to adopt this approach in classroom settings need to develop creative ways to elicit and process students' responses efficiently and effectively. For instance, we used constructed response items to elicit students' rich responses. But the framework could be extended to other formats, such as computerized modeling problems, e.g. Quellmalz & Pellegrino, 2009. How to use advanced technology (Liu, Brew, Blackmore, Gerard & Madhok, 2014; Quellmalz & Pellegrino, 2009) to reduce the effort needed for eliciting and analyzing student responses, but at the same time retain the essence of the TM design, may be an important question to answer by future investigators.

References

- Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5–31.
- Bransford, J. D., Brown, A. L. & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school* (expanded edition). Washington, DC: National Academy Press.
- Bransford, J. D. & Schwartz, D. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, 24, 61–100.
- Chinn, A. C. & Samarapungavan, A. (2008). Learning to use scientific models: Multiple dimensions of conceptual change. In R. A. Duschl & R. E. Grandy (Eds.), *Teaching scientific inquiry* (pp. 191–225). Rotterdam, Netherlands: Sense.
- Clement, J. (2000). Model based learning as a key research area for science education. *International Journal of Science Education*, 22(9), 1041–1053.
- Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *The Journal of the Learning Sciences*, 12(1), 91–142.
- diSessa, A. (1993). Towards an epistemology of physics. *Cognition and Instruction*, 10(2/3), 105–225.
- Frederiksen, J., White, B. & Gutwill, J. (1999). Dynamic mental models in learning science: The importance of constructing derivational linkages among models. *Journal of Research in Science Teaching*, 36(7), 806–836.
- Hewitt, P. (2005). *Conceptual physics* (10th ed.). Boston, MA: Wesley.
- Hunt, E. & Minstrell, J. (1994). A cognitive approach to the teaching of physics. In K. McGilly (Ed.), *Classroom lessons: Integration cognitive theory and classroom practice*. Cambridge, MA: MIT Press.
- Lehrer, R. & Schauble, L. (2006). Cultivating model-based reasoning in science education. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 371–388). New York, NY: Cambridge University Press.

- Linn, M. C., Lee, H.-S., Tinker, R., Husic, F. & Chiu, J. L. (2006). Teaching and assessing knowledge integration in science. *Science*, 313, 1049–1050.
- Liu, O. L., Brew, C., Blackmore, J., Gerard, L. & Madhok, J. (2014). Automated scoring for inquiry science assessment: Prospects and obstacles. *Educational Measurement: Issues and Practice*, 33(2), 19–28.
- Liu, O. L., Lee, H. S. & Linn, M. C. (2011). Measuring knowledge integration: Validation of four-year assessments. *Journal of Research in Science Teaching*, 48(9), 1079–1107.
- McCloskey, M. (1983). Intuitive physics. *Scientific American*, 248(4), 122–130.
- Minstrell, J. (2000). Student thinking and related assessment: Creating a facet assessment-based learning environment. In N. Raju, J. Pellegrino, L. Jones & K. Mitchell (Eds.), *Grading the nation's report card: Research from the evaluation of NAEP*. Washington, DC: National Academy Press.
- Namdar, B. & Shen, J. (2015). Modeling oriented assessment in K-12 science education: A synthesis of research from 1980 to 2013 and new directions. *International Journal of Science Education*, 37(7), 993–1023.
- National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.
- National Research Council (2014). *Developing assessments for the next generation science standards*. Washington, DC: The National Academies Press.
- Quellmalz, E. & Pellegrino, J. W. (2009). Technology and testing. *Science*, 323, 75–79.
- Ruiz-Primo, M. A. & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44(1), 57–84.
- Schwarz, C.V., Reiser, B.J., Davis, E.A., Kenyon, L., Acher, A., Fortus, D., ... Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632–654.
- Shen, J. & Confrey, J. (2007). From conceptual change to transformative modeling: A case study of an elementary teacher in learning astronomy. *Science Education*, 91(6), 948–966.
- Shen, J. & Confrey, J. (2010). Justifying alternative models in learning the solar system: A case study on K-8 science teachers' understanding of frames of reference. *International Journal of Science Education*, 32(1), 1–29.
- Shen, J., & Jackson, D. (2013). Measuring volume of tree: A problem-driven, modeling-based lesson for preservice science teachers. *Journal of Science Teacher Education*, 24(2), 225–247.
- Shen, J., & Linn, M.C. (2011). Connecting scientific explanations and everyday observations: A technology enhanced curriculum on modeling static electricity. *International Journal of Science Education*, 33(12), 1597–1623.
- Yin, Y., Tomita, K. M. & Shavelson, R. J. (2008). Diagnosing and dealing with student misconceptions about 'sinking and floating.'. *Science Scope*, 31(8), 34–39.