

Ashley B. Davidson, UNIVERSITY OF BRITISH COLUMBIA, adavidso@eos.ubc.ca
Christopher J. Addison, UNIVERSITY OF BRITISH COLUMBIA, chris.addison@ubc.ca
James Charbonneau, UNIVERSITY OF BRITISH COLUMBIA, james@phas.ubc.ca

Examining Course-Level Conceptual Connections Using a Card Sort Task: A Case Study in a First-Year, Interdisciplinary, Earth Science Laboratory Course

ABSTRACT

Universities are recognizing the need to prepare graduates to think conceptually and have the ability to take on complex, real-world problems. Strategies to assess conceptual knowledge are limited and often require more time and effort to complete than is accessible for most undergraduate courses. Card sorting is a very broad technique for understanding how people group concepts, but in higher education has typically been used to show a student's development towards expert-like thinking in a discipline as a whole. However, it typically does not give much insight into how we should change our teaching. In this paper, using the novel setting of two terms of a first-year, earth and ocean science lab that uses problem-based learning (PBL), we show how one can generate a card sort that is built using course learning goals and then use the analysis to make actionable improvements to course instruction. Using a card sort designed so that the expert sort corresponds to learning goals supported by the lab activities, we found that in both offerings of the course students generally moved towards expert-like sorting with a reduction in novice-like sorting. A striking feature stood out in both terms of the course, with one question scoring significantly lower than any other expert pairings, despite a change in the wording of that question between terms. This suggests that our course materials do not promote this specific conceptual connection that we had expected and gives us a clear place to look for issues in our course material. In a broader context, our results suggest that tailoring card sort questions to material at a course level, rather than at the discipline level, can provide a manageable, routine assessment of conceptual knowledge in students, while also providing feedback on the quality of course materials.

KEYWORDS

card sort task, conceptual connections, assessment tools, problem-based learning

INTRODUCTION

Background

Modern university graduates need to be able to adapt and solve complex, cross-disciplinary problems, yet many learners enter university without the conceptual knowledge necessary for this task (Lawless, Brown, and Boyer 2016, 54–55; Peters 2015; Smith et al. 2013). Students readily organize knowledge into disciplinary silos, reinforced by the departmental structure of universities, rather than organizing knowledge into a conceptual framework (Lawless, Brown, and Boyer 2016, 55). Strategies

such as problem-based learning (PBL) promote student learning, helping students to develop transferable skills and deal with open-ended problems comparable to the real world (Dadd 2009; Lawless, Brown, and Boyer 2016; Montgomery and Donaldson 2014). Assessing how PBL methods shift students' organization of knowledge is, however, hard to do with traditional assessment tools that can be used regularly in the classroom setting.

Card sort tasks have been used as a research tool to provide insights as to how participants organize their knowledge. Typically, participants are given a series of cards to sort based on the underlying principle or similarities of solution (Chi, Feltovich, and Glaser 1981, 124). The sorting itself may be done in person through one-on-one interviews (Chi, Feltovich, and Glaser 1981), as an in-class activity (Smith et al. 2013, 630), or through an online platform (Lapierre and Flynn 2020), with no clear bias observed between physical or virtual sorting (Bussolon, Russi, and Missier 2006). It is often specified that participants are not to solve the problems, just to sort them, with emphasis that this is not a test and there is no incorrect way to sort them. Participants may be asked to generate their own categories (open sort) or to sort problems into predefined categories (closed sort). Questions to be sorted are often pulled from textbooks relating to the discipline of interest, with some studies carefully choosing questions to contain a single surface feature and single deep feature. It is expected that experts in a field will tend to sort by underlying principle, whereas novices will tend to sort based on surface features in the problem (Chi, Feltovich, and Glaser 1981).

Studies using the card sort to investigate the transition from novice- to expert-like thinking have been used in several disciplines, as well as across several timelines, from tracking a single population over a semester to looking for trends between different populations at various stages in their career. Across disciplines, the majority of studies have looked at some combination of undergraduate students as novices and graduate students or faculty as the experts. Many disciplines have multiple studies with varying participants and timelines: in physics, where the applicability of the card sort tool was introduced looking at undergraduate students compared to graduate students (Chi, Feltovich, and Glaser 1981); in math, comparing first- and second-year undergraduate students (Fernandez-Plaza and Simpson 2016); in biology, studies have looked at biology faculty against non-biology majors (Smith et al. 2013), non-majors, early and advanced majors, graduate students and faculty to look at the transition over a career (Bissonnette et al. 2017), and tracking changes in undergraduates over a single semester (Hoskinson et al. 2017); similarly in chemistry, studies have looked at chemistry undergraduates against faculty (Krieter et al. 2016), as well as a comparison of students with no chemistry background, high school students, general chemistry undergraduates, upper-level chemistry undergraduates, graduate students, and faculty (Irby et al. 2016), and tracking changes in undergraduates over a single-semester, (Lapierre and Flynn 2020) and two-semester organic chemistry course (Galloway, Leung, and Flynn 2019). In general, these studies find that the card sorting activity is a suitable tool for investigating whether a participant displays more novice-like or expert-like thinking in their respective field.

One potential drawback of these studies is that the novice and expert levels, along with novice and expert groupings, are defined by the researcher. While typically looking at the difference between tested groups (e.g. undergraduate students versus graduate students), certain groups have had different roles in different studies; for example, in Chi, Feltovich, and Glaser (1981) graduate students are set as the experts (123–24), but when Irby et al. (2016) studied the spectrum from high school students to faculty, graduate students performed poorly and were observed to display more novice-like thinking than had been hypothesized. In cases where novices have been studied to look for an increase in expert-

like thinking over time, the card sort used had been developed for a true expert level, rather than a level one might expect an undergraduate to reach at the end of the course (Hoskinson et al. 2017, 4; Smith et al. 2013). While Hoskinson et al.'s (2017) study successfully recorded an increase in expert-like thinking in undergraduate students over a term, a greater level of detail in this progression may be possible with a lower "expert" threshold aligned with the course learning goals. With this in mind, we can think of how we would move towards using a card-sort task that is tailored for a class.

Goals of the card sort task in EOSC 111

EOSC 111 is an introductory, first-year, earth, ocean, and atmospheric sciences lab course that explores a variety of physical phenomena. This course is open to both majors and non-majors, with an emphasis on developing problem-solving skills and conceptual knowledge (e.g. general characteristics of wave propagation), as opposed to retaining specific knowledge. As part of the Carl Weinman initiative (Jones 2018), EOSC 111 labs were developed to have a high proportion of active learning and to use problem-based learning methods to provide a successful learning environment for students from a range of backgrounds and across all faculties. Students are provided with the basic knowledge required to complete the lab as pre-readings, and the instructor acts to facilitate learning through the lab activities, rather than present any new information during class. The overarching goals of this course are for students to gain practice making interpretations and drawing conclusions based on observations, and to recognize the interconnectedness of earth systems science. Ideally, these activities will help students to develop a conceptual framework, linking concepts (e.g. wave propagation) learned under different contexts (e.g. earthquakes and ocean waves) and avoiding the potential cognitive overload associated with the superficial features of each lab.

In order to gain insights as to whether the PBL lab activities were indeed promoting connections between concepts presented under different contexts, we used surveys to ask students how various aspects of the course may have benefitted them, or not. Overall, student satisfaction with the course was high, and across previous offerings of the course about 90% of students continually responded positively to surveys when asked about whether this course increased their confidence with new scientific concepts. However, while they felt more comfortable with new concepts, it was difficult to objectively assess whether students were truly making these conceptual connections or not. Teaching assistants (TAs) facilitating lab sections can make some assessment on most students through discussions, but it is not possible to achieve a detailed discussion with every student in each lab. Assessments involving long answers or interviews with students may provide insights as to the development of a conceptual framework, but these are rather labour intensive. Many courses may not have the hours to run these assessments regularly. The development of an assessment tool that offers a reasonably quick assessment for both students and instructors, yet still provides deeper information about how students are responding to course material, would be highly advantageous. One possible assessment technique for PBL that could be less labour intensive, yet still assess student ability to make conceptual connections, is the card sorting task.

Moving away from discipline-level assessment found in more typical card sorts and towards course-level assessment, we have developed a card sort task tailored to the course goals of EOSC 111, as a practical way to assess students' thinking in an earth science laboratory course employing PBL. We wanted to address two main research questions regarding this new application:

- (1) To what extent are current PBL methods used in EOSC 111 promoting conceptual connections?
- (2) What can a card sorting task tell us about the conceptual connections made within a course that can then be used to change how we teach the course?

If the card sort task gives us meaningful information about the cross-disciplinary connection students are making in class, this will provide an assessment technique for PBL that is less labour intensive than many current assessment options and capable of providing insights into student thinking and conceptual frameworks. This also provides a general framework for using card sorts to assess learning at the course level that can be easily applied to any course that has the goal of breaking down siloed knowledge.

MATERIALS AND METHODS

Research context

This research was conducted at the University of British Columbia, a research-intensive university and the only Canadian member of the Centre for the Integration of Research, Teaching, and Learning (CIRTL). CIRTL's mission is to improve undergraduate education through a culture of evaluating the effectiveness of teaching strategies for student learning. The card sort task was administered in EOSC 111 over two terms: Term 1 from September to December 2019 and Term 2 from January to April 2020. The card sort task was completed over the first five weeks of each term, and therefore changes to instruction as a result of COVID-19 did not occur until after all card sorts had been completed in Term 2. Each term had five separate sections, each with a different teaching assistant (TA), and registration in each section varied, with an average of 23 ± 6 students (full details in figure 1). EOSC 111 is open to both majors and non-majors, and often attracts a nearly equal number of students from both the faculty of science and the faculty of arts, with a smaller percentage of students from other faculties (see full breakdown in figure 2). Students were expected to complete the card sort task following the relevant lab in place of the usual post-lab survey; it was offered as an online survey through Qualtrics, linked to the survey section of the course's Canvas website. Students would be presented with nine questions and asked to sort them into three defined categories (closed sort) by dragging and dropping into the appropriate box. It was expected to take 10–15 minutes, similar to the post-lab surveys, as to not add to student workload. Timestamps on the surveys showed that the first sort took students the longest, averaging 11–13 minutes, with the following two sorts only taking an average of 4–5 minutes and 6–8 minutes, respectively. The card sort task was available for one week, opening after their lab was completed, and closing before the beginning of the next lab. Students received credit for completing the card sort task (~0.5% of their final grade, part of a 5% survey mark), however they did not need to consent to be part of the research project in order to receive credit. A response rate of greater than 90% was achieved across both terms, however, not all students completed all card sorts. Only data from students who completed at least two sorts was used in this study. Across both terms of the study, ~75% of students completed all three sorts, ~17% completed only two sorts, and the remaining 8% either did not respond, did not consent to the study or completed less than two sorts. In all cases, any data that came from the 8% of students who either did not consent to be part of the study or only completed one sort was removed from this analysis; a detailed breakdown of response rates by term is shown in figure 2.

Figure 1. Timeline of the card sort deployment across five sections from September to December 2019 (Term 1) and from January to April 2020 (Term 2). Card sorts were completed following the labs highlighted in green, in weeks one, two, and five, replacing the post-lab survey that occurred in all eight remaining labs. Note that in Term 2, the final card sort was completed in early February 2020, avoiding disruption from COVID-19.

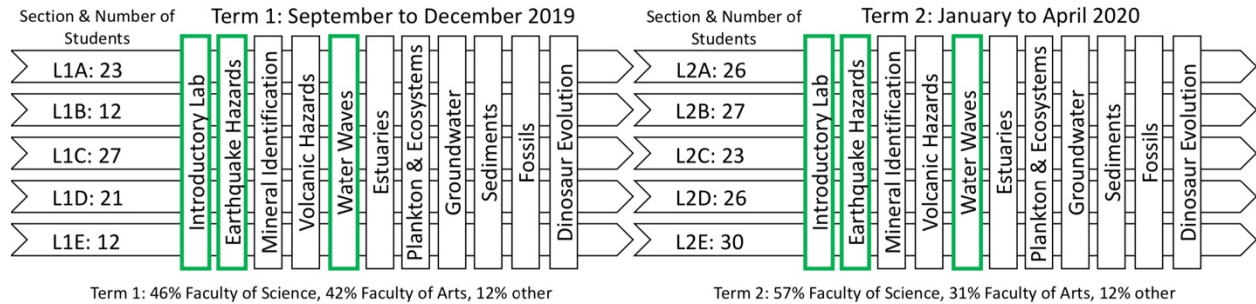
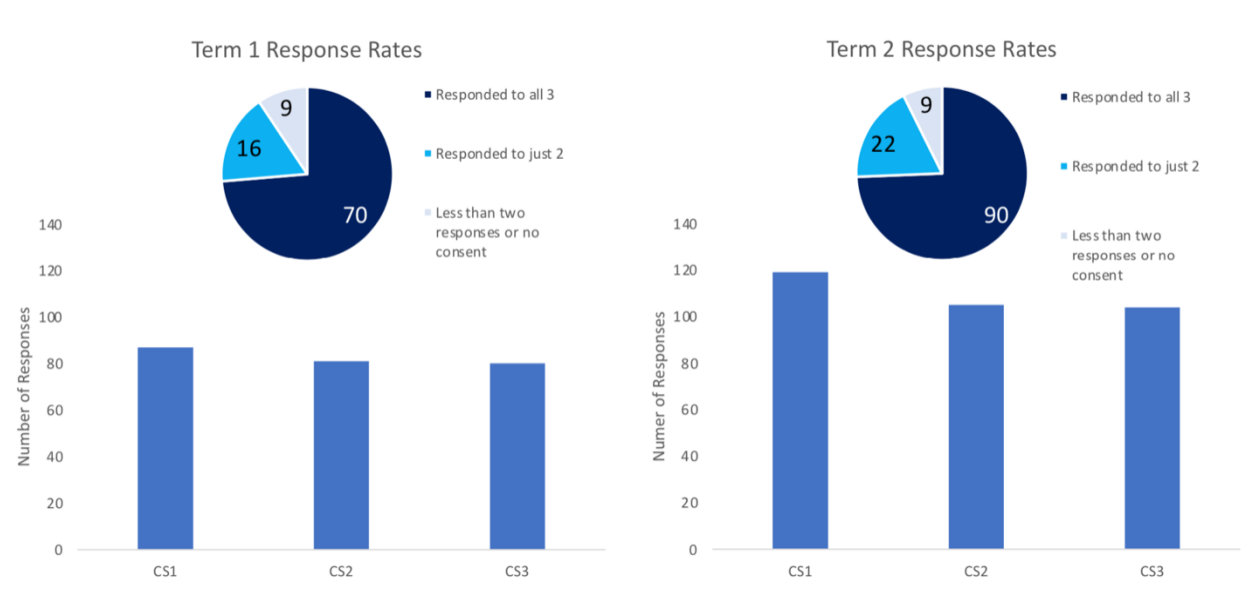


Figure 2. Response rates for Term 1 (September–December) and Term 2 (January–April); the bar graph shows the number of responses on each card sort (CS), while the pie chart indicates the proportion of students who either completed all three card sorts (navy blue), two card sorts (sky blue), or less than two card sorts/did not give consent for the study (light blue).



The card sort task

We carefully designed our card sorting task to target the learning goals of two labs that both explored wave concepts and to minimize any features beyond the intended surface and deep features of each question (e.g. a series of “Why?” questions, or answers that could be grouped as quantitative versus qualitative). The questions were developed to reflect the language students were exposed to during the laboratories. Some deep features were heavily focused on during a lab period, while others presented a more challenging connection the student needed to make on their own (see details below and in appendix 1). Surface features were largely based on the laboratory topics (e.g. earthquakes and ocean waves), while deep features were underlying concepts students were introduced to through laboratory activities. We also chose to include a surface feature “sound” that we did not cover at all in lab, to help

differentiate conceptual connections that students are making themselves, from ideas that may be linked in class.

The card sort task was presented to students three times over the term as a closed sort: week one, before any material was covered; week two, following the “Earthquakes Hazards and Determining Epicentres” lab (earthquakes); week five, following the “Wave Propagation” lab (waves). We expected the three-week break between card sorts 2 and 3 to be long enough that students would not simply remember their answers from the previous sort (Loftus 1985). We chose to use closed sorts to decrease the effort needed to process responses, to test the card sort’s suitability as an effective routine assessment tool. The card sort was accessed by students through an online interface, where they were instructed to sort problems based on the common underlying science principles, but not to solve them. Students were given the deep feature categories and asked to sort the nine questions by dragging each question into the corresponding category (see appendix 2 for student interface). The initial card sort was used to provide a baseline before we had introduced any material. As this was an introductory-level course, students were expected to exhibit novice-like thinking at the start of the course with this potentially being their first exposure to these concepts in a geoscience context (Petcovic and Libarkin 2007). Despite providing the deep feature categories in the closed sort, novices without prior knowledge of the concepts covered in course materials would be expected to focus on surface features as a means of sorting concepts, not recognizing the connections between deep features (Chi, Feltovich, and Glaser 1981). Over the following weeks, it was hypothesized that students would begin to shift towards more expert-like thinking, recognizing the deep feature as the underlying concept, which would be evident through an increased proportion of expert pairs, increased expert accuracy, and decreased expert edit distance as has been observed in discipline-based card sort studies (Bissonnette et al. 2017; Hoskinson et al. 2017; Smith et al. 2013,). Based on the question design, we expected this shift to be gradual, with certain expert pairings increasing after week two and others not until week five; a breakdown of the details behind each question can be found in appendix 1, with a summary of the expected results for each card sort and questions themselves (figure 3) in table 1, below:

Table 1. A summary of the predicted sorting results based on exposure to course materials with each card sort

Card sort	Unlikely correct	Maybe correct	Expect correct
1	No expectations		
2	B, D, H, G	A, C, F	E, I
3		A, C, F, G	E, I, D, H, B

For questions that are expected to be paired correctly, this means the language and content of the question were addressed directly in class; for example, in the earthquakes lab, which is completed immediately before card sort 2, students complete an activity focused on the resonance of different buildings and damage incurred, so we would expect they sort “E” under resonance. When we predict a question “maybe correct,” we cover the content of the question tangentially; again, with the earthquakes lab example, students complete activities on resonance frequency and so they may make connections to the other resonance frequency questions, “A” and “C.” Notably, laboratory activities do not explicitly focus on the deep feature connections between labs. We connect the surface features (e.g. earthquakes) to the deep feature (e.g. resonance frequency), but not to the other surface features (e.g. ocean waves),

and so by linking both questions “A” and “E” to resonance frequency, for example, students are making this connection themselves. Finally, if we predict a question is “unlikely correct” it is because it uses language different from what students were exposed to in the lab and/or we have not covered the content of the question yet. While we have predicted how the sorting of the questions may change after completing lab activities through the term, it is harder to predict the actual pairings. For example, on card sort 2 we may observe the correct pairing of “AE,” with students connecting the content of “A” with resonance and therefore sorting it in the same group as “E.” We will often be considering the pairing frequency and accuracy throughout the card sorts and will refer to expert and novice pairs; the expert pairs for resonance frequency, for example, are “AE,” “AC,” and “CE,” while the novice pairs for earthquakes are “BE,” “BI,” and “EI.”

Figure 3. Question set for the card sort task arranged with hypothesized deep features in columns and hypothesized surface features in rows. Note: in Term 1, question A was phrased slightly differently: “Ocean waves enter a bay and form a standing wave. Why did this occur?” Full details in appendix 1.

		Hypothesized Deep Features		
		Resonance Frequency	Solve with wave equation	Particle motion caused by wave
Hypothesized Surface Features	Ocean waves	A packet of waves enters a bay and the two sets of waves constructively interfere. Why did this occur? A	Two ocean waves have the same period, but different wavelengths. Which is faster? H	A seagull floats in the ocean, bobbing up and down on waves. Explain why the motion is only up and down? D
	Earthquakes	Explain why after an earthquake, one building is demolished and its neighbor is in good condition. E	What parameters would you need to describe the period of an earthquake? B	An earthquake occurs south of Vancouver. Seismographs record the strongest signal in the east-west direction. Why? I
	Sound	A performer sings a specific note, while directing their voice at a piece of glass. The glass vibrates and eventually shatters. How is this possible? C	Your neighbor is singing loudly. You happen to remember the speed of sound, could you determine the period of their voice? G	Why can sound travel through material in any state (solid, liquid and gas)? F

Data analysis

Similar to other card sort studies (Smith et al. 2013, 631–32), two main metrics were used to evaluate students’ responses to the card sorting task: edit distance and pairing accuracy. Edit distance calculates the number of moves necessary to change a student’s response to match either the hypothesized novice sort (novice edit distance), or the hypothesized expert sort (expert edit distance). For our set of questions, the maximum number of moves to reach either the expert or novice hypothesized sort was 6, and the minimum 0. Ideally, students’ novice edit distance would increase across the card sorts as they more frequently pair by deep features instead of surface features, and the expert edit distance would counter this, decreasing as they are progressing to more expert-like thinking.

Pairing accuracy was the other metric we considered; this differs from that of percent card pairings used by Smith et al. (2013) in that it calculates the proportion of correct pairings and subtracts the proportion of incorrect pairings. This metric was developed as part of a card sort study through the University of British Columbia's Science One program to account for both correct and incorrect pairings in a single metric (Addison, Charbonneau, and Dubois 2017). As with edit distance, there is both a novice pairing accuracy and expert pairing accuracy, where students' responses are compared to the respective hypothesized sorts. The maximum pairing accuracy is 1, while the minimum is -0.25; ideally, we would see novice pairing accuracy decreasing and expert pairing accuracy increasing over the course of term. Both of these metrics were calculated using the programming language R, with a script written by a researcher involved in the study.

In addition to these aggregate metrics, a compilation of the proportion of individual pairings recorded on each sort were investigated in depth to look for trends. For each term, we looked at comparisons of the responses between sections, across faculties, and how individuals progressed over the course of term. These comparisons were tabulated in Microsoft Excel, and we used the two-tailed student's *t* test function to compare various values, noting any significant differences as those with a *p*-value <0.05.

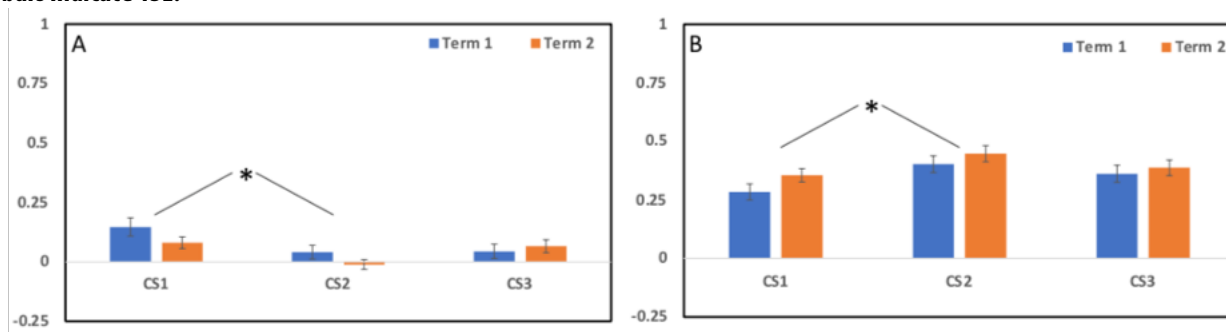
RESULTS AND DISCUSSION

The card sorting task, deployed over two terms, provides ample data to be evaluated using a number of metrics. First, we begin with the calculated pairing accuracy from each sort, followed by the edit distance for each sort. Following this, we breakdown the data further to look at the results at a pairing level, comparing the progression of novice and expert pairings from each term. Lastly, we look at how students progress over a term and discuss possible distinctions in our participant pool.

Overall scores

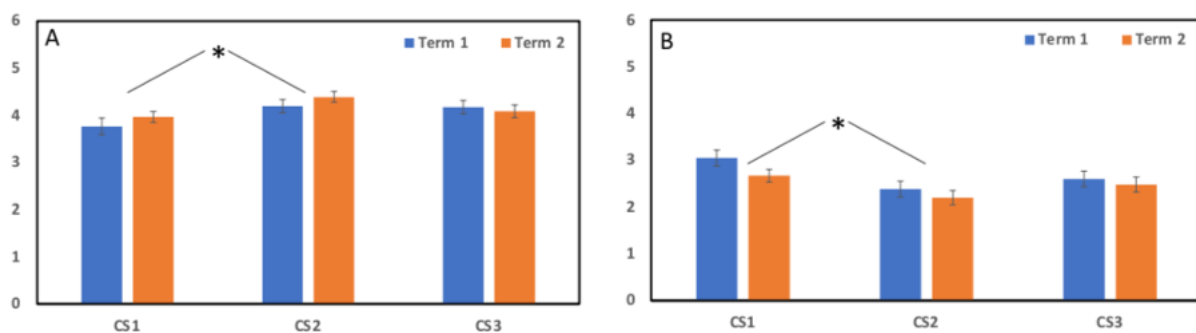
Considering pairing accuracy over the course of the card sorts each term, we can clearly see a decrease in novice accuracy and an increase in expert accuracy from card sort 1 (CS1) to card sort 2 (CS2) for both terms (figure 4). While this change is a significant difference, the changes from CS2 and card sort 3 (CS3), which all trend in the opposite direction, are not statistically significant. It is also notable that the January–April 2020 term starts off with more expert-like baseline scores but show possibly less retention in their scores by CS3 compared to the September–December 2019 term.

Figure 4. Novice (A) and expert (B) pairing accuracy scores with standard error plotted. Note that the maximum accuracy score is 1 and the minimum is -0.25. Asterisks indicate statistically significant differences between CS1 and CS2. The error bars indicate 1SE.



Similar to the trends in pairing accuracy, we see that in both terms we have a significant change in edit distance between CS1 and CS2, with both novice edit distance and expert edit distance moving in a direction that suggests more expert-like thinking (figure 5). Again, we see that between CS2 and CS3 there is some loss, with the trend moving in the opposite direction. We also observe that the spring 2020 term starts off with a more expert-like baseline, as they did with pairing accuracy. Overall, these metrics provide a quick overview of the results and show that over both terms, results suggest students are initially making conceptual connections as a consequence of the first lab, albeit with some loss of those connections over the full five-week period.

Figure 5. Novice (A) and expert (B) edit distance scores with standard error plotted. Note that the maximum edit distance is 6 and the minimum is 0. Asterisks indicate statistically significant differences between CS1 and CS2. The error bars indicate 1SE.



Detailed expectations of progression based on course material

If the method of teaching our labs was reaching students at a conceptual level, students should be able to make conceptual connections between card sort questions that fall outside of course material. The card sort activity was developed to ideally have students gradually increase the frequency of expert pairings based on the material taught over two labs, with some questions designed to include material outside of our lab content. We hypothesized increases in specific pairings after card sort 2 and 3, with certain pairings expected based on lab materials, and others likely if conceptual connections were being made.

Card sort 1 (CS1) provided a baseline to measure students' progression against. Card sort 2 (CS2) followed the earthquakes lab, where we would expect to see a breakdown in the surface "earthquake" pairings (BE, BI, EI) and an increase in the expert pairings (AE, CE, BG, BH, DI, FI). We do in fact see decreases across the novice pairings (figure 6), strongest in "EI" (earthquake/resonance frequency and earthquake/particle motion), and a greater decrease in Term 2 than in Term 1. We also see strong increases in the expert pairings "CE" (sound/resonance frequency and earthquake/resonance frequency), "DI" (ocean waves/particle motion and earthquake/particle motion), and "FI" (sound/particle motion and earthquake/particle motion) across both terms (figure 7); connections were immediately made to the "sound" questions, despite not being covered by any lab material, signifying that students are able to connect concepts on their own. Additionally, many students were able to make the connection between all three "particle motion caused by wave" questions, despite not yet covering the material in the waves lab that will explicitly describe question D (full breakdown of covered material by question in appendix 1). The expert pairings "BG" (earthquake/wave equation and sound/wave equation) and "BH" (earthquake/wave equation and ocean waves/wave equation) did not

show much change, which can be expected since these are part of the “solve with wave equation” deep feature, and the wave equation itself is not introduced until the waves lab. The pairing “AE” (ocean waves/resonance frequency and earthquakes/resonance frequency) also does not show much improvement and will be discussed in more detail later.

Card sort 3 (CS3) directly followed the waves lab, taking place three weeks after CS2 had been completed. We expected to see a breakdown in the surface “ocean wave” pairings (AD, AH, DH) and an increase in the expert pairings (AC, AE, DI, DF, BH, GH). In this case, we observed statistically insignificant changes in the pairings of “AD” (oceans waves/resonance frequency and ocean waves/particle motion) and “AH” (ocean waves/resonance frequency and ocean waves/wave equation) relative to CS2, with “DH” decreasing in Term 1, but relatively unchanged in Term 2 (figure 6). Similarly, we did not see clear increases in the expert pairs between CS2 and CS3 (figure 7); the pairs “DI,” “DF,” and “GH” all see increases between CS1 and CS2, but by CS3, there is little change or a slight decrease in these pairings. “BH” followed this pattern in Term 1 but showed very little change from CS1 in Term 2. Lastly, the expert pairs containing “A” (AC and AE) show a very low correct response rate across card sorts 1–3 and both terms compared with other expert pairings.

By looking at the proportion of correct pairings across each card sort, it appears that students are responding positively towards the course materials and PBL approach. Particularly after the earthquakes lab, the data indicates students are able to extend what they have learned directly in class to other applications with the development of related conceptual frameworks. It is also apparent that some of these connections are strongest immediately following the material, and there is some loss over the three weeks between card sorts. This may not be totally unexpected, as we do not consider any of the wave-based concepts covered in the card sorts in labs between the earthquakes and waves lab (switch to geology-based labs), so there is no reinforcement of the material between card sorts. Also notable is the difficulty in pairing question “A” correctly—in Term 1, this was thought to be related to unfamiliar language in the original question (the term “standing wave” which had not been used in class), so it was updated to reflect language used in the earthquakes lab (“constructive interference,” which was used interchangeably with resonance). Even with this update, the responses between the fall and spring terms are very similar. Like any of the questions with the surface feature “sound,” question “A” was not explicitly discussed or connected to in class activities. However, students did very well at sorting “sound” questions with the correct deep feature, making this connection themselves, yet they were not able to connect “A” with resonance. This points to a weakness in our activities regarding resonance frequency; while students seem to grasp the potential damage that can be caused as a result of resonance frequency, they have not quite linked this to the cause of damage itself, stemming from constructive wave interference. This observation provides us with direction to improve course material to help reinforce these conceptual connections between labs. Despite the variations between card sorts, it is clear when comparing figures 6 and 7 that overall students are thinking more like experts than novices with respect to our material.

Figure 6. The average proportion of correct novice pairings from (A) Term 1 and (B) Term 2. Averages are calculated from the results of five labs sections, with 1SE error bars.

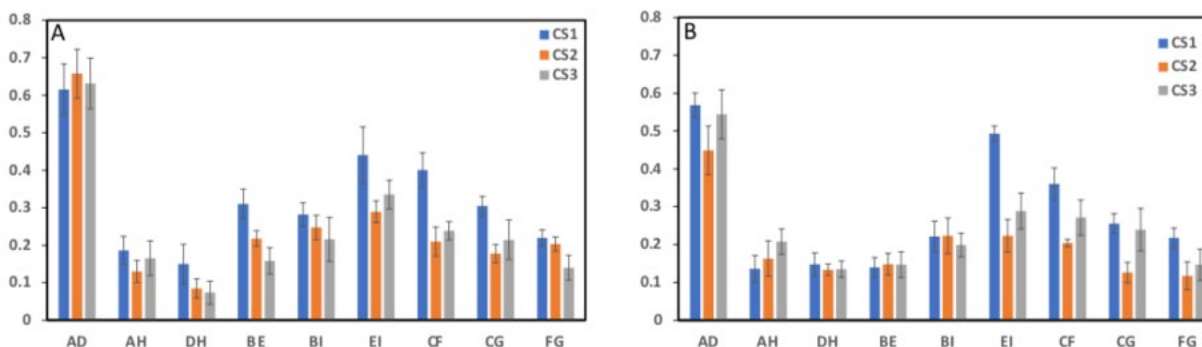
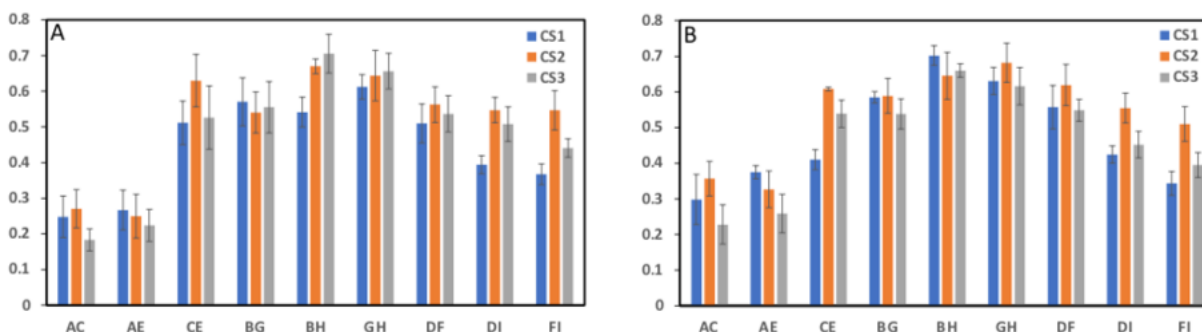


Figure 7. The average proportion of correct expert pairings from (A) Term 1 and (B) Term 2. Averages are calculated from the results of five labs sections, with 1SE error bars.



Different backgrounds

Given the average composition of this course contains a large proportion of students from both the Faculty of Science and Faculty of Arts, we wanted to assess the impact of student academic background on potential differences in learning as assessed by the card sort activity. We looked at both the proportion of correct pairings, and how individuals changed their card sort responses through term, whether they increased or decreased certain pairings.

Term 1 was very close to a 50-50 split between arts and science students with 46% of students from the Faculty of Science and 42% from the Faculty of Arts (12% were a mixture of other Faculties). The baseline pairings were relatively similar in Term 1, with novice pairings tending to be slightly higher for arts students (from 0% to 20% higher depending on the pair), while the expert pairings were slightly higher for science students (0–10% higher depending on the pair). After CS2, we see a decrease in almost all novice pairings for both groups (figure 8a) and an increase in most expert pairings (figure 9a), though this is a slightly higher increase for science students (up to 15–20% higher for pairs “BH” and “DI”). By CS3 we see increases across the novice pairings in both groups (figure 8a) and decreases in expert pairings in both groups (figure 9a).

Term 2 had a larger population of science students, with about one-third arts students (31%) and two-thirds science students (57%). The trends in baseline novice and expert pairings are similar to Term 1, but with science students showing a much greater proportion of correct expert pairings,

suggesting that some of these concepts addressed by the card sorts may have come up in classes during Term 1; for example, students in first-year physics may have already covered a unit on wave propagation, and may already see the deep conceptual connections between questions. As with Term 1, we see a decrease in novice pairings after CS2 (figure 8b), and an increase in expert pairings across both groups (figure 9b). After CS3, we again observe slight increases in the novice pairings (figure 8b) and decreases in the expert pairings (figure 9b).

The results from the card sort tasks over both terms show some differences between faculties, but this is likely a result of more recent, previous exposure to the material covered by our labs for students in the Faculty of Science, whereas these topics are likely newer for students coming from the Faculty of Arts. It does appear that by Term 2, science students may be coming in with the beginnings of a conceptual framework (in addition to wave-based knowledge), allowing them to begin connecting these questions before exposure to our learning materials. It may be more beneficial for science students to take this course in first term, to help them build a conceptual framework before they are introduced to too much content, or to avoid covering material they have seen previously.

Figure 8. The change in the proportion of a given novice pair from (A) Term 1 and (B) Term 2, for students pursuing a Bachelor's of Science degree (BSc) between CS1–CS2 (black), CS1–CS3 (black dotted), and for students pursuing a Bachelor's of Arts degree (BA) between CS1–CS2 (white), –CS3 (white striped). A negative change in proportion indicates less students chose that pairing on either CS2 or CS3, relative to CS1 and shows student learning, highlighted by the green area.

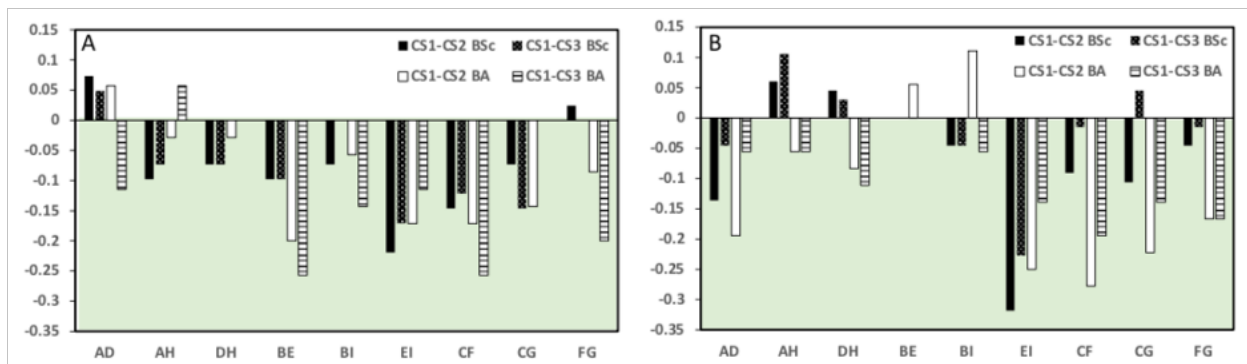
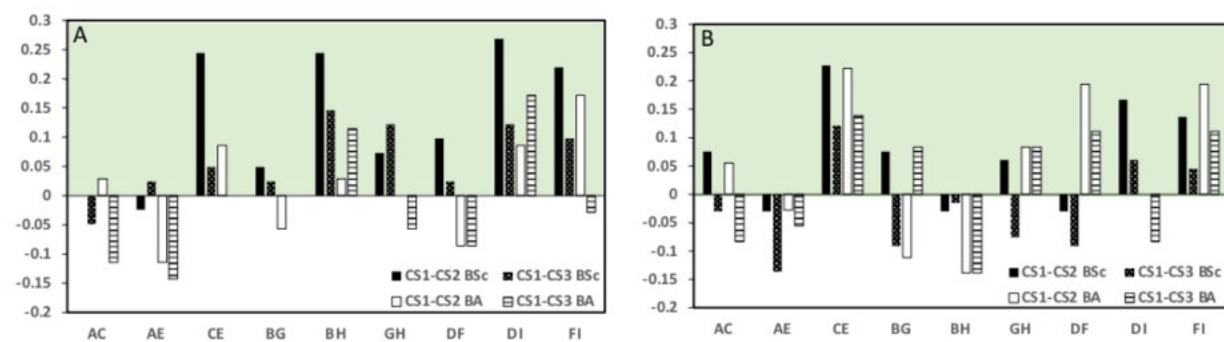


Figure 9. The change in the proportion of a given expert pair from (A) Term 1 and (B) Term 2, for BSc students between CS1–CS2 (black), CS1–CS3 (black dotted), and for BA students between CS1–CS2 (white), CS1–CS3 (white striped). A positive change in the proportion indicates more students chose that pairing on either CS2 or CS3, relative to CS1 and shows student learning, highlighted by the green area.



Changes in pairings by student

The analysis presented so far has considered the average performance of the entire cohort, or within significant groups within this cohort (students enrolled by faculty). But it is also worth considering how an individual student progresses through term and whether their pairings are persistent between sorts, or more fluid. Based on the design of the card sort questions, we would expect students to change pairings, with an overall decrease in novice pairings and an increase in expert pairings as lab activities change their knowledge constructs. The relative amount of persistent and fluid pairing will depend on the baseline knowledge of the student and how well-developed their conceptual framework regarding covered topics already is. In Term 2, a large proportion of students, ranging from 50% to 90% across all pairings and card sorts, did not change their pairings. When considering the difference between CS1 and CS3, novice pairings stayed constant for 50–80% of students (figure 10b), with increases ranging from 5% to 20% and decreases ranging from 5% to 35%. The largest decreases were observed for “EI,” “CF,” “CG,” and “FG” (sound and earthquake pairings). Expert pairings between CS1 and CS3 were retained for 60–80% of students (figure 11b), with increases ranging from 10% to 30% and decreases from 5% to 30%. The greatest increases were observed for “CE,” “FI,” “DF,” and “DI” (particle motion and resonance frequency pairings). When looking at these results from a faculty point of view, the overall trend discussed above is dictated by the larger proportion of science students in Term 2; arts students are much more likely to change their answers, with 40–70% of students keeping the same expert pairings between CS1 and CS3, decreases of 5–30%, and increases of 10–45%. Term 1 displayed very similar trends, though the difference between faculties was not apparent. A more consistent increase across expert pairings between CS1 and CS3 was observed (figure 11a), with the wave equation group of questions showing an increase between 20% to 30%, comparable with the particle motion group. With the overall proportion of “BG,” “BH,” and “GH” similar between terms (figure 7), yet lower increases in these pairs in Term 2, this suggests that more students in Term 2 had already been introduced to the wave equation before the waves lab, allowing them to make the connections from the start of term. This does make it more difficult to discern if the conceptual connections students are making are a direct result of our PBL activities but provides a baseline that we can begin to reference as we continue to assess conceptual connections in our course.

Figure 10. A comparison of the proportion of students who increased (blue) or decreased (grey) a given novice pair from CS1 to CS3. These are average values across all five sections with 1SE, for Term 1 (A), and Term 2 (B).

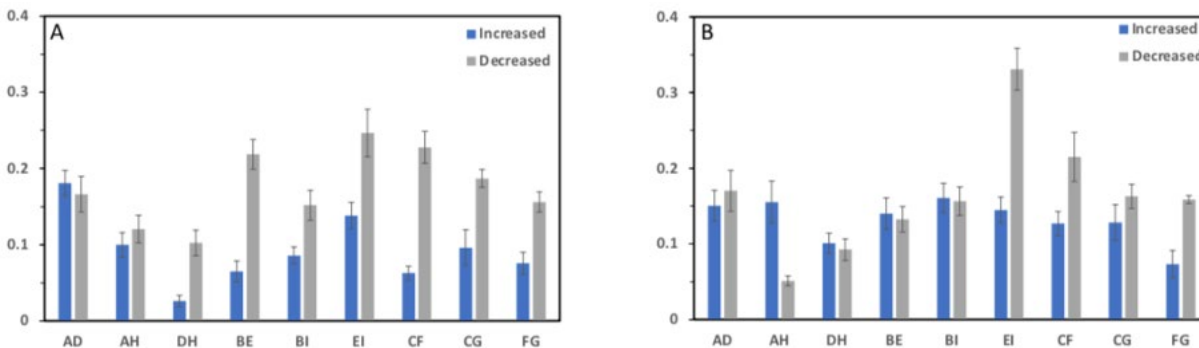
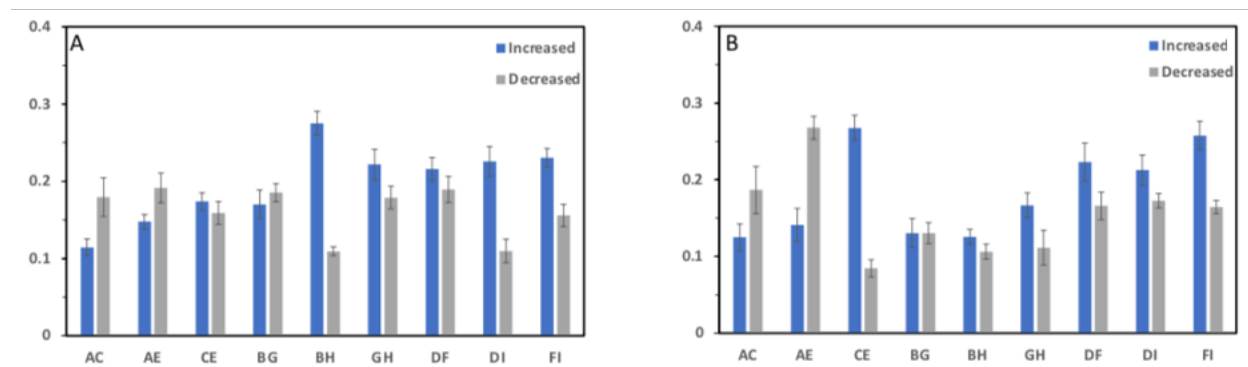


Figure 11. A comparison of the proportion of students who increased (blue) or decreased (grey) a given expert pair from CS1 to CS3. These are average values across all five sections with 1SE, for Term 1 (A), and Term 2 (B).



Comparison to self-reporting

One of the drivers to develop a sound assessment tool for conceptual knowledge stems from consistently positive responses to our optional end of term survey which asks students about their comfort with new scientific concepts and ideas. We continued to run this survey as a control and added a question allowing students to self-report on their experience with the card sort activity. The question regarding the card sort was “This term we tried out a card sorting activity in place of some post-lab surveys. Did you find that the act of sorting questions helped you make conceptual connections between labs that you otherwise wouldn’t have made?” (table 2). Greater than 50% of students responded positively to this question in both terms, with Term 2 reaching up to 67% positive response. This increased positive response may be related to a clearer set of instructions and expectations that were developed after feedback from Term 1. Both terms also saw an overwhelmingly positive response to the question “Do you feel that this course has helped to give you more confidence when approaching new scientific concepts and ideas?” with over 90% of students responding positively, consistent with results from previous terms.

Table 2. Results from the optional end of term survey offered each term

Survey Question	Definitely	A bit	Not really	Not at all	Response rate
(1) This term we tried out a card sorting activity in place of some post-lab surveys. Did you find that the act of sorting questions helped you make conceptual connections between labs that you otherwise wouldn’t have made?	16%	39%	37%	8%	84 out of 95; Term 1
	22%	45%	24%	8%	101 out of 121; Term 2
(2) Do you feel that this course has helped to give you more confidence when approaching new scientific concepts and ideas?	54%	40%	5%	1%	84 out of 95; Term 1
	52%	39%	7%	2%	101 out of 121; Term 2

Advancing curricular innovation based on the card sort task

Our goal in this study was to determine if the card sort task could be scaled down and tailored to be a routine assessment of student learning at a conceptual level. We were able to clearly identify where course materials fell short, as students were overall showing more expert-like than novice-like thinking throughout the series of card sorts. This expert-like thinking was evident through different measures, including edit distance (figure 4), which showed that students required fewer moves to reach the expert edit distance than the novice edit distance. Pairing accuracy showed similar trends as edit distance (figure 5), and the proportion of expert pairs was much greater than novice pairs across both terms and all card sorts (figures 6–9). When considering the results at a pairing level however, it became clear that expert and novice pairs involving question “A” showed a deviation in these trends. This shortcoming was related to resonance frequency, and in particular, we believe the actual physical cause of resonance frequency was not clear to students, though they connected with the potential results. We can see this through the lower proportion of correct expert pairs “AC” and “AE” (figure 7); students are making connections between “CE,” but something about “A” is not clear. Initially we thought this may be related to language and updated the question for Term 2 to match language in the earthquakes lab, however this did not improve the number of correct pairings. The main difference between the three questions is that “C” and “E” focus on the result of reaching the resonance frequency, whereas “A” focuses on what resonance frequency is. This suggests that students struggle to make the connection between constructive interference and resonance frequency or are not aware that a bay would have a resonance frequency. Given the content in the labs, the first option is most likely, as lab activities demonstrate resonance frequency in building models and discuss ways to avoid damage, but perhaps need a stronger connection to the root cause. As a result of the card sort activity, we will be changing the language and more strongly emphasizing the principles behind resonance frequency in the earthquakes lab to see if this increases conceptual connections the following term. In future terms, we may design a card sort task to target other shared concepts covered by our labs in order to evaluate student learning and identify any other weaknesses in current course materials. This approach can also be easily extended to other courses, both in EOAS and outside the discipline; by setting the learning goals of a course to be the expert level, the card sort task can be adapted to evaluate students’ conceptual knowledge and identify potential weakness in materials, a tool that can be utilized across disciplines.

LEARNING OUTCOMES AND CONCLUSIONS

We set out to determine if a card sort task designed around course learning goals was a suitable tool that could be used effectively to measure deficiencies in expected conceptual connections. We found that despite the wide number of variables in EOSC 111 (e.g. backgrounds, class size, instructors, etc.) compared to some other card sort studies, the card sort is in fact a strong tool for measuring conceptual connections throughout term, and also sensitive enough to note differences between terms, likely stemming from different student backgrounds. The card sort task provided an objective assessment of students’ conceptual knowledge and agrees with students’ positive self-assessment regarding comfort with new concepts, though a greater number of students have increased comfort with new concepts, than made conceptual connections tested by the card sort. We did find that retention of conceptual connections did not fare well over the entire five-week span and may necessitate changes to the course in order to improve students’ retention. In that sense, the card sort can be used both as an assessment of students’ conceptual knowledge, and as a targeted course development tool which allows

you to identify where concepts were not clear enough for students, with the ease of a weekly survey. In such cases where concepts are not clear to students, a marked difference in the frequency of pairing was observed, and lab materials can be revised in order to remedy this deficiency. In addition, we have shown that while a wealth of data can be obtained from a series of card sorting tasks, relatively simple metrics such as edit distance and pairing accuracy can provide a very quick and informative measure of students' progress. Overall, we've found that the card sort task is a manageable and highly adaptable assessment tool, that can be tailored to the needs of a given course across disciplines, rather than applied to assess a single discipline as has been done in previous studies, and that in itself, may help students make further conceptual connections.

ACKNOWLEDGMENTS

We would like to thank Patrick Dubois for their help in developing the survey interface and R scripts for analysis.

Ashley B. Davidson is a PhD Candidate and the course coordinator of EOSC 111 in the Department of Earth Ocean and Atmospheric Sciences at the University of British Columbia (CAN). <https://orcid.org/0000-0003-2141-6710>.

Christopher J. Addison is an Associate Professor of Teaching in the Department of Chemistry at the University of British Columbia (CAN). <https://orcid.org/0000-0002-5268-8682>.

James Charbonneau is an Associate Professor of Teaching and Director of the Science Gateway Programs at the University of British Columbia (CAN). <https://orcid.org/0000-0002-7142-5124>.

DISCLOSURE

ABD was funded by a TAR (Teaching as Research) Graduate Stipend Award from CITRL at the University of British Columbia.

ETHICS

This research was approved by the University of British Columbia's Behavioural Research Ethics Boards (H19-02359) and deemed minimal risk. Students were expected to complete the card sort as an assignment, however participation in the research was voluntary, the data used came only from students who consented to participate.

REFERENCES

- Addison, Christopher J., James Charbonneau, Patrick Dubois. 2017. "A Novel Card Sort Activity to Measure Interdisciplinary Thinking." International Society for the Scholarship of Teaching and Learning (ISSOTL) 2017, Calgary (International).
- Bissonnette, Sarah. A., Elijah D. Combs, Paul H. Nagami, Victor Byers, Juliana Fernandez, Dinh Le, Jaren Realin, Selina Woodham, Julia I. Smith, and Kimberly D. Tanner. 2017. "Using the Biology Card Sorting Task to Measure Changes in Conceptual Expertise During Postsecondary Biology Education." *CBE Life Sciences Education*, 16, no. 1 (March): 1–15. <http://doi.org/10.1187/cbe.16-09-0273>.
- Bussolon, Stephen, Barbara Russi, and Fabio Del Missier. 2006. "Online Card Sorting: As Good as the Paper Version." *Proceedings of the 13th European conference on Cognitive ergonomics: trust and control in complex socio-technical systems*, 250 (September): 113–14. <http://doi.org/10.1145/1274892.1274912>.
- Chi, Michelene T., Paul J. Feltovich, and Robert Glaser. 1981. "Categorization and Representation of Physics Problems by Experts and Novices." *Cognitive Science*, 5, no. 2 (April): 121–52. https://doi.org/10.1207/s15516709cog0502_2.

- Dadd, Kelsie. A. 2009. "Using Problem-Based Learning to Bring the Workplace into the Classroom." *Journal of Geoscience Education*, 57, no. 1 (January): 1–10. <http://doi.org/10.5408/1.3544224>.
- Fernandez-Plaza, José Antonio, and Adrian Simpson. 2016. "Three Concepts or One? Students' Understanding of Basic Limit Concepts." *Educational Studies in Mathematics*, 93, no. 3 (November): 315–32. <https://doi.org/10.1007/s10649-016-9707-6>.
- Galloway, Kelli R., Min Wah Leung, and Alison B. Flynn. 2019. "Patterns of Reactions: A Card Sort Task to Investigate Students' Organization of Organic Chemistry Reactions." *Chemistry Education Research and Practice*, 20, no. 1 (July): 30–52. <http://doi.org/10.1039/c8rp00120k>.
- Hoskinson, Anne-Marie, Jessica Middlemis Maher, Cody Bekkering, and Diane Ebert-May. 2017. "A Problem-Sorting Task Detects Changes in Undergraduate Biological Expertise Over a Single Semester." *CBE Life Sciences Education*, 16, no. 2 (October): 1–12. <http://doi.org/10.1187/cbe.16-05-0175>.
- Irby, Stefan M., Andy L. Phu, Emily J. Borda, Todd R. Haskell, Nicole Steed, and Zachary Meyer. 2016. "Use of a Card Sort Task to Assess Students' Ability to Coordinate Three Levels of Representation in Chemistry." *Chemistry Education Research and Practice*, 17, no. 2 (January): 337–52. <http://doi.org/10.1039/c5rp00150a>.
- Jones, Francis. 2018. "Comparing Student, Instructor, Classroom and Institutional Data to Evaluate a Seven-Year Department-Wide Science Education Initiative." *Assessment and Evaluation in Higher Education*, 43, no. 2 (June): 323–38. <http://doi.org/10.1080/02602938.2017.1343799>.
- Krieter, Felicia E., Ryan W. Julius, Kimberly D. Tanner, Seth D. Bush, and Gregory E. Scott. 2016. "Thinking Like a Chemist: Development of a Chemistry Card-Sorting Task to Probe Conceptual Expertise." *Journal of Chemical Education*, 93 no. 5 (March): 811–20. <http://doi.org/10.1021/acs.jchemed.5b00992>.
- Lapierre, Keith R., and Alison B. Flynn. 2020. "An Online Categorization Task to Investigate Changes in Students' Interpretations of Organic Chemistry Reactions." *Journal of Research in Science Teaching*, 57, no. 1 (January): 87–111. <http://doi.org/10.1002/tea.21586>.
- Lawless, Kimberly A., Scott W. Brown, and Mark A. Boyer. 2016. "Educating Students for STEM Literacy: GlobalEd2." In *Technology, Theory, and Practice in Interdisciplinary STEM Programs*, edited by Reneta D. Lansiquot, 53–82. New York: Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-56739-0>.
- Loftus, Geoffrey R. 1985. "Evaluating Forgetting Curves." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, no. 2: 397–406. <http://doi.org/10.1037//0278-7393.11.2.397>.
- Montgomery, Homer, and Katherine Donaldson. 2014. "Using Problem-Based Learning to Deliver a More Authentic Experience in Paleontology." *Journal of Geoscience Education*, 62, no. 4: 714–724. <http://doi.org/10.5408/13-085.1>.
- Petcovic, Heather L., and Julie C. Libarkin. 2007. "Research in Science Education: The Expert-Novice Continuum." *Journal of Geoscience Education*, 55, no. 4: 333–39. <https://doi.org/10.1080/10899995.2007.12028060>.
- Peters, Michael. 2015. "Using Cognitive Load Theory to Interpret Student Difficulties with a Problem-Based Learning Approach to Engineering Education: A Case Study." *Teaching Mathematics and Its Applications*, 34, no. 1 (March): 53–62. <http://doi.org/10.1093/teamat/hru031>.
- Smith, Julia I., Elijah D. Combs, Paul H. Nagami, Valerie M. Alto, Henry G. Goh, Muryam A. A. Gourdet, Christina M. Hough, et al. 2013. "Development of the Biology Card Sorting Task to Measure Conceptual Expertise in Biology." *CBE Life Sciences Education*, 12, no. 4 (December): 628–44. <http://doi.org/10.1187/cbe.13-05-0096>.

APPENDICES

Appendix 1. Breakdown of card sort questions and anticipated difficulty of each question

A	<p>Question: A packet of waves enters a bay and the two sets of waves constructively interfere. Why did this occur?</p> <p>Anticipated difficulty: In Term 1, the question was originally phrased as: “Ocean waves enter a bay and form a standing wave. Why did this occur?” The phrase “standing wave” was not explicitly discussed in class, so the wording was changed for Term 2 to be sure this was not causing question A to be paired correctly with low frequency. The term “constructively interfere” was used when discussing resonance frequency in earthquakes, but this scenario was not discussed in the waves lab. This is a challenging question that may be paired correctly after card sort 2.</p>
B	<p>Question: What parameters would you need to describe the period of an earthquake?</p> <p>Anticipated difficulty: This question is phrased in the same manner as students worked with in the waves lab, but they were not instructed to think of an earthquake in these terms. Students were not introduced to the wave equation until week five, so we expect them to make this pairing on card sort 3.</p>
C	<p>Question: A performer sings a specific note, while directing their voice at a piece of glass. The glass vibrates and eventually shatters. How is this possible?</p> <p>Anticipated difficulty: The language in this question does not parallel anything covered in their labs, and no comparisons to sound were made directly in lab. This is a challenging question that requires the student to make true conceptual connections, most likely after card sort 2.</p>
D	<p>Question: A seagull floats in the ocean, bobbing up and down on waves. Explain why the motion is only up and down?</p> <p>Anticipated difficulty: This language and concept are directly reflected in course material and it is expected students will pair it correctly after card sort 3.</p>
E	<p>Question: Explain why after an earthquake, one building is demolished and its neighbor is in good condition.</p> <p>Anticipated difficulty: This language and concept are directly reflected in course material and it is expected students will pair it correctly after card sort 2.</p>
F	<p>Question: Why can sound travel through material in any state (solid, liquid and gas)?</p>

	<p>Anticipated difficulty: The language in this question does not parallel anything covered in their labs, and no comparisons to sound were made directly in lab. This is a challenging question that requires the student to make true conceptual connections, most likely after card sort 2.</p>
G	<p>Question: Your neighbor is singing loudly. You happen to remember the speed of sound, could you determine the period of their voice?</p> <p>Anticipated difficulty: While no concepts dealing with sound were made in any of the labs, this question does contain similar language to the other wave equation questions, making it slightly more likely to be paired correctly than the other “sound” questions, likely after card sort 3.</p>
H	<p>Question: Two ocean waves have the same period, but different wavelengths. Which is faster?</p> <p>Anticipated difficulty: This language and concept are directly reflected in course material and it is expected students will pair it correctly after card sort 3.</p>
I	<p>Question: An earthquake occurs south of Vancouver. Seismographs record the strongest signal in the east-west direction. Why?</p> <p>Anticipated difficulty: This language and concept are directly reflected in course material and it is expected students will pair it correctly after card sort 2.</p>

Appendix 2. Card sort task student interface

Students are first presented with an introduction to the card sort task where they are told not to solve the problems, but to sort them based on the common underlying science principles by dragging each of the questions on the left into one of the groups provided on the right.

Please sort these same problems again into the three given groups, using the principles shown in the names of the groups.

Items	Particle motion caused by wave
Explain why after an earthquake, one building is demolished and its neighbor is in good condition.	
Two ocean waves have the same period, but different wavelengths. Which is faster?	
A packet of waves enters a bay and the two sets of waves constructively interfere. Why did this occur?	
A performer sings a specific note, while directing their voice at a piece of glass. The glass vibrates and eventually shatters. How is this possible?	
A seagull floats in the ocean, bobbing up and down on waves. Explain why the motion is only up and down?	
An earthquake occurs south of Vancouver. Seismographs record the strongest signal in the east-west direction. Why?	
Your neighbor is singing loudly. You happen to remember the speed of sound, could you determine the period of their voice?	
What parameters would you need to describe the period of an earthquake?	
Why can sound travel through material in any state (solid, liquid and gas)?	
	Solve with wave equation
	Resonance frequency



Copyright for the content of articles published in *Teaching & Learning Inquiry* resides with the authors, and copyright for the publication layout resides with the journal. These copyright holders have agreed that this article should be available on open access under a Creative Commons Attribution License 4.0 International (<https://creativecommons.org/licenses/by-nc/4.0/>). The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited, and to cite *Teaching & Learning Inquiry* as the original place of publication. Readers are free to share these materials—as long as appropriate credit is given, a link to the license is provided, and any changes are indicated.