

Engaging science academics with evidence based practices: Use of concept inventories in chemistry and physics across eight universities

Helen Georgiou^a, Manjula D. Sharma^b

Corresponding author: helengeo@uow.edu.au

^aSchool of Education, University of Wollongong, Wollongong, NSW, 2522, Australia

^bSchool of Physics, The University of Sydney, Sydney NSW 2006, Australia

Keywords: Conceptual surveys, concept inventories, active learning, chemistry education, Physics education, engaging science academics in evidence-based teaching, scholarship of teaching and learning (SOTL)

Abstract

There is ongoing research on how to improve student engagement and attainment in STEM in higher education, with active learning recognised as a feasible approach for several decades now. However, the uptake of active learning, and other evidence-based approaches, is inconsistent. This paper reports on one aspect of an Australian Government funded Fellowship; the specific scholarly practice of the use of concept inventories, widely associated with active learning, to engage academics in evidence-based practices in STEM disciplines. The ultimate aim was to equip lecturers with the tools to measure student attainment. In close collaboration with academics, pre- and post-tests were administered to students in a total of 12 different courses, constituting over 3000 individual student questionnaires collected across eight Australian Universities. We report on the implementation focusing on; engaging staff, the types of concept survey results made visible to staff not generally accustomed to seeing such results, and tentatively offer the possibility of national data on learning gains. Results show that the majority of lecturers engaged and continued the use of concept inventories. Our study demonstrates that concerted use of concept inventories might lead to increased uptake of evidence-based approaches with potential for improved teaching and learning in STEM disciplines.

Background

Recently, it has become clear that there is an expectation that individual academics, or lecturers, should engage in scholarly approaches to improve teaching and learning (Probert, 2013; Johnston, Hopkins, Varvell, Sharma, & Thornton, 2007; Vardi & Quin, 2011). This is sometimes referred to as 'Scholarship of Teaching and Learning' (SoTL) (Boyer, 1990). However, for most teaching-research academics in science, overcoming the tension between their teaching and research responsibilities, in order to more substantially engage with scholarly approaches to teaching is a significant challenge (Hemer, 2014; Lemass & Stace, 2010). There are issues even for those that *do* wish to engage more substantially, in terms of research in teaching, as this represents a change of field (Johnston et al., 2007; Matthews et al., 2015; Nicholls, 2004). Science lecturers thus might wish to engage with SoTL, a field of its own, but could also conduct research in other fields, such as 'Discipline Based Education Research' or Psychology, which each have different ideologies and methodologies, making it difficult for researchers to become enculturated into those fields (Dolan et al., 2018; Sharma & McShane, 2008).

One potentially less problematic entry into education research for science lecturers is in the area of ‘active learning’ (AL), which is a well-defined international research agenda that has a focus on improving student learning and experiences using evidence based approaches (Faust & Paulson, 1998). In STEM disciplines, the AL movement has been popularised by Nobel Laureate Carl Weimann and Harvard scholar Eric Mazur and has had considerable influence, especially in North America. AL strategies have been proven efficacious in the literature and are associated with higher student learning gains, fewer failures in science courses and higher retention across science degree programs (Freeman et al., 2014).

The popularity of AL strategies was established upon a foundation of particular instrument use and methodologies, namely, concept inventories (CI) and specifically, measurement of learning gains, which are able to reveal improvements in student learning. Despite the utility of local data such as assessments marks, student evaluation surveys and enrolment numbers, it was arguably CI use that provided the potency and universality to make the AL case and help it take hold (Hake, 1998). This is because these instruments test ‘conceptual understanding’, an underlying construct related to but not explicitly assessed in individual subjects (e.g., Thornton & Sokoloff, 1998). The central finding that AL strategies are far superior to ‘traditional’ lecturing is measured by gains on CIs and could be replicated over time and internationally.

CIs are developed and analysed using a range of quantitative and qualitative methods. For example, measures such as difficulty, reliability and validity of items on a test may be determined through classical test theory (Engelhardt, 2009). To measure gains, for either item or cohort, CIs can be administered as pre-tests and post-tests. Hake’s ‘normalized gain’ (gain) (1998) is commonly used to provide a quantitative measure for improvement. As was done with the Force Concept Inventory (FCI), gain scores may be compiled and thus provide some measure of the quality or success of instructional strategies used across different institutions (Hake, 1998). Figure 1 shows a ‘Hake plot’ where individual gain scores for each institution are plotted against their pre-test score. Different areas are highlighted, below ‘a’, to indicate low gains, between ‘b’ and ‘a’ to indicate medium gains and above ‘b’, indicating high gains.

CIs also identify students’ alternative conceptions or misconceptions, and this might be useful in terms of assessing prior understanding, identifying problem areas and focusing on conceptual understanding, in addition to using the tool as a way to measure overall gains (e.g., Francek, 2013; Georgiou & Sharma, 2015; Hake, 1998). In this case, descriptive statistics and hypothesis testing may then be used to diagnose existing alternative conceptions, or compare differences in test scores between groups. Such approaches provide more detail in terms of ‘what’ the problem is and ‘how’ you might develop teaching practices to address it.

The use of CIs does not occur without criticism. Critics suggest student understanding is trivialised by the assignment of numerical marks and that such inventories do not capture the process of ‘thinking’ (Smith & Tanner, 2010). There are also criticisms of the inconsistent application of Item Response Theory or Classical Test Theory to ensure validity and reliability and of the gain measurement (Wallace & Bailey, 2010). Nevertheless, CIs have a long history in higher education STEM (Treagust, 1988) and although their utility for measuring student understanding is now widely accepted (Libarkin, 2008), some expertise is required to implement them.

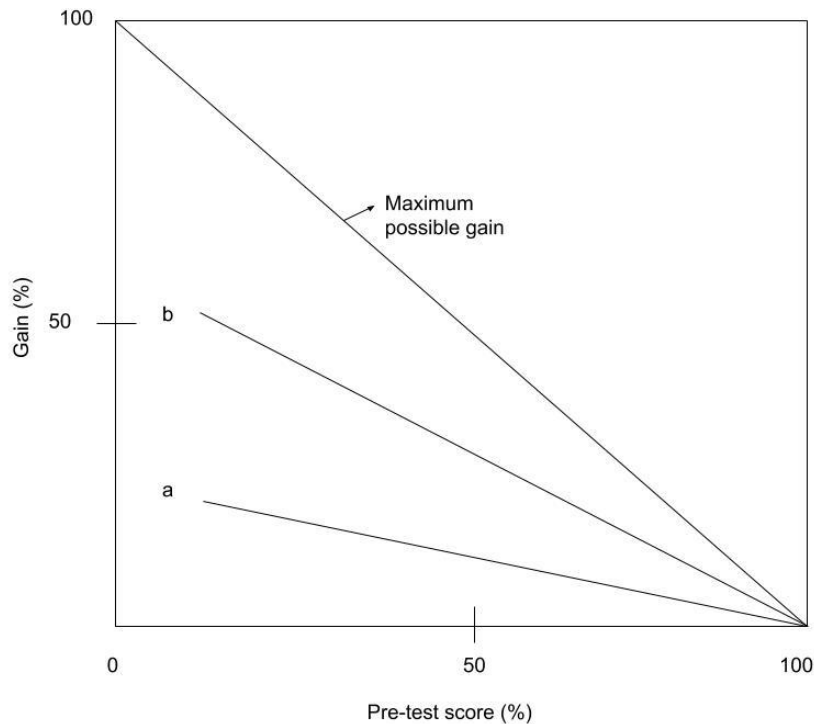


Figure 1: A 'Hake plot' where individual gain scores for each institution are plotted against their pre-test score; below 'a' indicates low gains, between 'b' and 'a' indicates medium gains and above 'b' indicates high gains.

In this study, AL and CIs were considered a good way to engage academics with evidence-based teaching and learning (Sharma & Georgiou, 2016). Generally, the uses of quantitative methods resonate with STEM faculty, who, due to their training, are familiar with the ethos of these approaches. That is, while STEM faculty may not be particularly familiar with the use of survey instruments, the quantitative aspect is more likely to appeal (Hendry, Georgiou, Lloyd, Tzioumis, Herkes & Sharma, 2020; Georgiou & Sharma, in press). Furthermore, White et al., (2015) demonstrate that embedding AL strategies within a teaching professional development program for 45 STEM academics resulted in transformative and persistent change in pedagogical approaches.

Aims

This paper reports on an Australian Government funded Fellowship which aimed to investigate and champion STEM academics' engagement with evidence-based teaching and learning practices. Specifically, this paper focuses on the use of CIs to equip lecturers with the tools to measure student attainment and to make visible the types of student learning difficulties that would otherwise not be explicitly assessed.

The aims of this paper therefore are:

1. To provide details of a program intended to engage lecturers in evidence-based teaching through the use of CIs and report on some ways participants engaged with this program.
2. To provide an overview of the utility of CIs; 'what CIs tell us'. Specifically, to report on the state of student understanding of scientific concepts across a diversity of universities in Australia, answering questions such as:

- What are the conceptual difficulties indicated by these CIs and which ones persist across courses and institutions (indicating something fundamentally difficult about these particular concepts)?
- Are the scores and gains similar/different across institutions/different sample groups (perhaps pointing to ‘good’ practices)?

Since we have results from a range of CIs from across Australia for both chemistry and physics, we also tentatively offer the possibility of national data on learning gains.

Method

Participant Recruitment

The second author, an established discipline-based educational researcher in physics at a research-intensive Australian university contacted potential participants who were known to have an interest in, or were responsible for teaching and learning programs (see Sharma & Georgiou, 2016 for more detail).

Data collection and analysis

The data are presented in two sections. The first presents an overview of the CI implementation and additional data collected in the form of interviews, site visits/observations and field notes, see Table 1. The second presents the data from four popular CIs used in tertiary science because they relate to fundamental concepts encountered in undergraduate courses and on topics that are known to be problematic for students (e.g. Georgiou, Sharma, O'Byrne, & McInnes, 2009; Georgiou & Sharma, 2010; Wattanakasiwich, Talaeb, Sharma, & Johnston, 2013).

Thermal concepts survey (TCS) covers heat, temperature, thermal processes (Wattanakasiwich et al., 2013) and may be administered in part (15 questions) or in full (35 questions). The use of TCS demonstrating learning gains with AL, including an analysis of lecture time spent on interactivity, can be found in Georgiou and Sharma (2015).

Force and Motion Concept Evaluation (FMCE), a 43-question survey covers kinematics, Newton's laws, energy and momentum (Thornton & Sokoloff, 1998). The use of FMCE demonstrating learning gains with Interactive Lecture Demonstrations as a form of AL is in Sharma et al. (2010).

Force Concept Inventory (FCI) is a 30-question survey covering Newton's laws and kinematics (Hestenes, Wells, & Swackhamer, 1992). The use of FCI demonstrating learning gains with 6,000 students establishes the low, medium and high gain areas; the higher the gain, the more pronounced the AL approach (Hake 1998).

Chemistry Concept Inventory (CCI) is a 22-question survey covering the transformation of matter, energetics, and representations in chemistry (Mulford & Robinson, 2002). The use of an adaptation of the CCI demonstrating learning gains with AL (Predict, Observe, Explain) is shown in Costu et al. (2010).

In total, 12 separate courses were analysed with 17 individual sample groups, as some institutions administered the surveys to different cohorts resulting in 3500 responses deemed appropriate for statistical analysis.

The analysis included: ‘cleaning’ data such as removing responses that were less than 50% completed; evaluating descriptive statistics, such as means and standard deviations for pre-

and post-tests; matching pre- and post- samples; calculating normalised gain (Eq. 1) and effect sizes (Eq. 2).

$$\langle g \rangle = \frac{\text{Post test class average} - \text{pre test class average}}{\text{total possible} - \text{pre test class average}}$$

Equation 1. Normalized gain measure (Hake 1998)

$$\text{Effect size} = \frac{\text{Difference in group means}}{\text{Variability of groups}}$$

Equation 2. Effect size (Preacher & Kelley, 2011)

This analysis therefore produced data on the question level (difficulty, distractor analysis, gain between pre- and post-test administration) and test level (descriptive statistics, normalised gain and effect size). Gain scores were plotted on a Hake graph (Figure 1). This paper reports on only some aspects of the analysis, relevant to the aims. The study has approval from The University of Sydney Human Research Ethics Committee (Approval number 2014/028).

Table 1: An overview of the CI implementation and additional data collected.

ID	Inventory	Details of administration	Additional data collection
B	FCI	Pre- and Post- collected contemporaneously at institution (online) as part of course and historical data also provided to us for analysis purposes. 2012 data provided only in terms of overall test score by student (not individual question)	<ul style="list-style-type: none"> • Interviews • Site visit (field notes, lecture observation)
C	CCI	Pre- and Post- administered online (Survey Monkey) contemporaneously.	<ul style="list-style-type: none"> • Interviews • Site visit (field notes)
D	FMCE	Both Pre- and Post- Administered in first lab and final lecture, pen and paper contemporaneously	<ul style="list-style-type: none"> • Site visits (field notes, lecture observation)
D	CCI	Pre- and Post- administered in first and final lab sessions, pen and paper contemporaneously	<ul style="list-style-type: none"> • Site visits (field notes, lecture observation)
D	TCS	Pre- and Post- administered in first lab session and final lecture, pen and paper, historical data provided only	<ul style="list-style-type: none"> • Site visits (field notes, lecture observation)
F	CCI	Pre- and Post- administered online within institution (Moodle) contemporaneously	<ul style="list-style-type: none"> • Interviews • Site visits (field notes)
F	TCS	Pre- and Post- was administered in first and final lectures, pen and paper contemporaneously	<ul style="list-style-type: none"> • Interviews • Site visits (field notes)
G	CCI	Pre- and Post- administered in first and final lab sessions on pen and paper contemporaneously	No additional data collected
H	FMCE	Pre- and Post- administered in first and final lab sessions on pen and paper contemporaneously	No additional data collected
I	CCI	Pre- and Post- administered in first and final lab sessions on pen and paper contemporaneously	<ul style="list-style-type: none"> • Interviews
J	FMCE	Pre- and Post- collected at institution (online) contemporaneously and only overall test scores provided for analysis	No additional data collected

Results

This section addresses the two main aims; first relating to the process of engaging lecturers in evidence-based development of teaching practices in an effort to improve learning outcomes and the second relating to what CIs tell us more generally.

CI implementation and lecturer engagement

The researchers supported the participant in the implementation and analysis of data. An initial meeting took place where a plan for their implementation was developed. The implementation varied from place to place; some participants took responsibility for more of the administration whilst others requested more support; some preferred the pen-and-paper options whilst others preferred to run the CI online; some conducted part of the analysis themselves whilst others relied wholly on the resources provided.

The analysis was performed by the researchers for all cases, with a solution available for the lecturers in the way of a fully functional *Excel* template, should they wish to perform a basic analysis in the future. The template was programmed to automatically fill once data were imported. A report containing a question-by-question summary, the means and standard deviations for both pre- and post-tests and gain relative to other samples completing the same test (de-identified) was also provided, see Figure 2 for an example. Where possible, a follow-up interview was conducted to discuss these resources (Table 1).

Of the nine separate implementations of CIs, six individual administrators consented to a follow up. Of those six, three had used CIs before (but were not planning to in the year of the program) and three had never used them before. Five of the six reported continued use of the CIs in follow-up interviews. The two examples discussed here, participant **F** (from University **F**) and participant **I** (from University **I**), have contrasting familiarity with CIs and are at different points of transformation. Participant **F** has just embarked and participant **I** is versed in educational transformations as well as CIs. We draw on additional data collected as shown in Table 1 (interviews and field notes from site visits) to elaborate on their experiences.

University F: Participant **F** was in the initial stages of transformation with no prior experience with CIs in chemistry. Furthermore, the participant was a research scientist given responsibilities for first year chemistry, hence was not familiar with educational research or SoTL. The CIs were implemented online over Moodle for first year chemistry students in two streams (for non-majors and majors). The participant used the CIs from the 2012 data to procure a 'baseline' (with results acting like 'pre-tests'). They then used the results to create online pre-lecture questions as a first step towards flipping lectures, thus engaging in evidence-based teaching. The participant was able to use what they had done to attract funding to continue using CIs and enhance their pedagogy; gathering evidence illustrating 'shifts' in student learning/experiences complementing those gathered through assessment, retention and institutional surveys. This study tracked the use of CIs by participant **F** through to 2016, with the results providing feedback on newly implemented pedagogies, such as flipped lectures and other AL activities. The results from the CIs made visible students' conceptual difficulties influencing the pedagogies of the teaching team as well as providing personal development for the participant who was in the process of preparing for promotion.



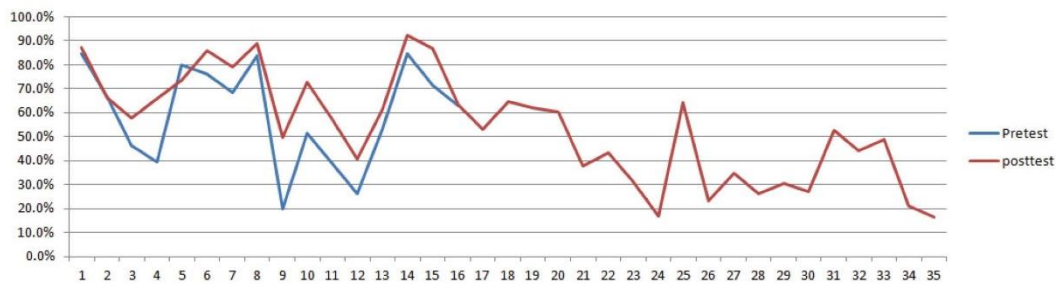
CONCEPTUAL SURVEY REPORT

Contact: Helen Georgiou, Georgiou@physics.usyd.edu.au, Manjula Shama, Sharma@physics.usyd.edu.au

Overview

	Pre /16	Post /16	Post /35
N	1033	737	737
Mean	9.54	11.28	18.86
St Dev	2.81	2.92	5.00
Gain	0.27		

Individual questions



How do we compare?

Yellow star indicates gain measurement for international sample. Blue stars are a similar Australian university. Red squares are from an international sample.

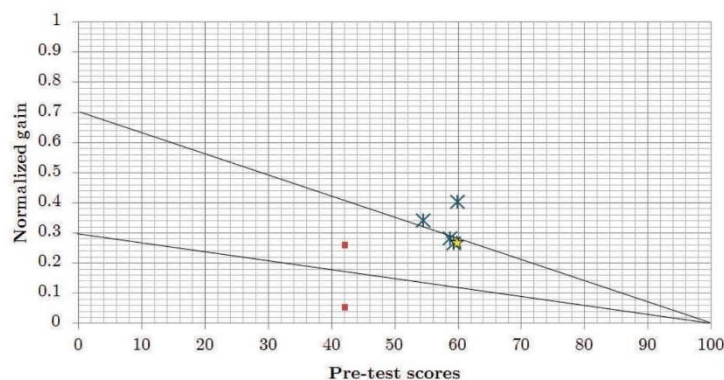


Figure 2: Example of report received by participants.

University I: Participant I was deeply immersed in transformation as their department was undertaking a teaching re-invigoration program in chemistry, which included different staffing structures, teacher training and engagement in SoTL. Participant I, a staff heavily invested in teaching with prior experience with CIs was not planning on administering them since they had additional responsibilities consuming their time. However, given the opportunity to be in a national project, they did implement CIs and the results showed

improved gains; the highest (Table 2), thus providing participant **I** with substantial evidence of the efficacy of their re-invigorated programs. The teaching team gained confidence and continued on their trajectory. What they focused on was detail on individual items responses (Figure 3) as these are considered pivotal to conceptual understanding underpinning further disciplinary learning. They searched/developed AL strategies aligned with specific items, integrating them into their courses. These results were used to justify and market AL pedagogies across the faculty.

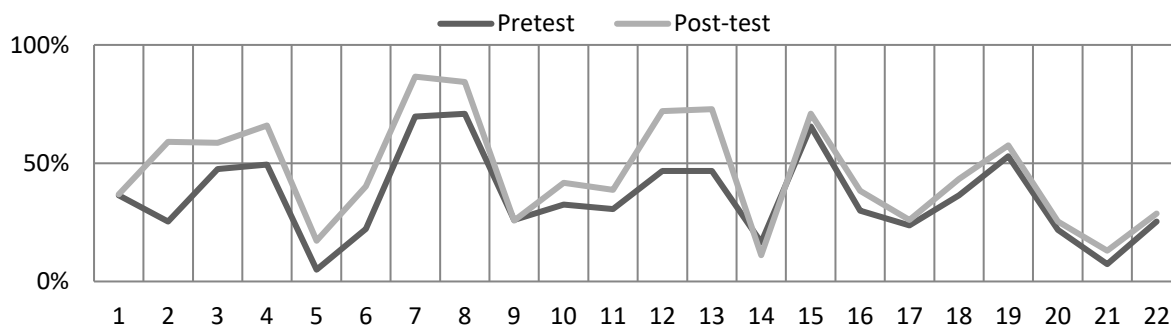


Figure 3: Pre and post test scores for the CCI for University I

What CIs tell us more generally

Conceptual difficulties

We provide a summary of the most ‘difficult’ or ‘troublesome’ questions across the different samples. The criteria for selecting these questions include: relatively low proportion correct in pre- or post- tests and/or extremely low or negative gains. As well as the correct answer with a brief explanation of the concept, the most common response is provided to give a sense of the misconceptions or alternative conceptions revealed. For some of the questions, notable observations are also provided, where unusual activity was noticed in the statistical analysis.

Chemistry concept inventory

The three areas that proved particularly troublesome in the chemistry CI were: conservation of mass and reaction ratios, understanding of molar mass, and, precipitation and solutions. We present the first two.

In question 5, a representation of six molecules of O_2 and six sulfur atoms in ‘a closed container’ is presented and students must choose from one of five representations for the following reaction: $2S + 3O_2 \rightarrow 2SO_3$ when the mixture reacts ‘as completely as possible’. The correct answer depicts four molecules of SO_3 with two S atoms remaining. This question was a low scoring question for two of the six universities tested. One university had extremely low scores (5% correct). The more common selection showed six molecules of SO_3 . This demonstrates an understanding of molecules but not the conservation of mass and ratios. There are only enough oxygen atoms to make four molecules of SO_3 so the two additional S atoms remain unchanged. Interestingly, in the post scores of one university, students changed their answers from the incorrect option above, not to the correct option but to another incorrect option which depicted two S_2O_6 molecules and two S atoms. This was noteworthy as it was a large number of students ($n=600$). The switch demonstrates the understanding of conservation of mass and ratios to the detriment of understanding of molecules.

Question 14 shows a full stop on the page “.” and asks for an estimate of the total number of carbon atoms required to fill it. The options include 4, 200, 30 000 000 and Avogadro’s number 6.02×10^{23} . This question was consistently low scoring (for all universities) and very low (10% correct) for one. The correct answer is 30 000 000. The more common selection was Avogadro’s number, which represents 12 grams of carbon, which corresponds to a large spoonful.

Thermal concepts survey

In this survey, two areas proved conceptually difficult; first understanding and calculating specific heat, and the second relates to ideal gas behaviours, specifically to the relationship between temperature and pressure.

In question 3, which refers to two cups of water:

Cup A: 100g, heated to 75°C from room temp

Cup B: 200g, heated to 50°C from room temp

Students were asked: “When both are allowed to cool back down to room temp, which one had more heat transferred from it?” with options: Cup A, Cup B, Both same, not enough to tell. This question had low pre scores from both Universities administering it. There was some improvement in this question, with post scores around 50-60%. The correct answer is both the same (this requires understanding of the specific heat equation). The more common selection was Cup B, with 200g of water.

Question 9 focuses on the image shown in Figure 4 with explanatory text (a syringe contains an ideal gas and has a frictionless piston of mass M ; it is moved from a beaker of cold water to a beaker of hot water).

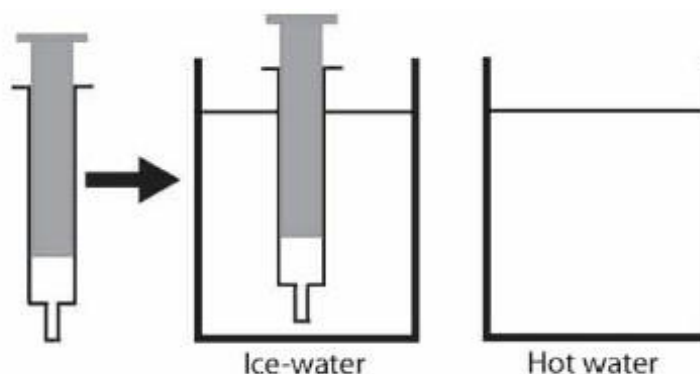


Figure 4: Image used for question 9 on the TCI. The pressure of the gas inside the syringe does not change when moved from ice-water to hot water because the piston is in mechanical equilibrium (it is free to move). Many respondents indicated that the pressure increased.

The question asks about the change in gas pressure; does it increase, decrease or is there no change? This question exhibited low pre-test scores and modest gains for both universities. The correct answer is that there is no change. The gas will expand but the pressure is still atmospheric pressure, since the piston is allowed to move freely. The most common selection was that the pressure increased. This conflates an increase with pressure with temperature, without taking into account the volume change.

Force Motion Concept Evaluation

This inventory revealed misunderstandings about Newton's third law and identifying the force(s) involved in a projectile's motion.

In question 34, which forms part of a series that explore the forces between cars and trucks when they collide, students are asked to identify what happens when a car and the truck (same weight) collide, with the truck being stationary when the car hits it. There are several selections possible with different combinations of the greater/lesser force on the two objects. This was a low scoring and low gain in both samples (the third sample did not have individual data). The correct answer was that the force on the car and the truck is the same magnitude. The most common response was that the force the car exerts on the truck is greater than the force the truck exerts on the car.

The previous question, where the truck and cars are moving towards each other at the same speed, resulted in a higher proportion of correct answers. The same is true of preceding questions which involved a truck and a car where they were different weights, except in this case, students also linked weight to force (at the same speed, the larger weight exerted the larger force) as well as speed (the higher speed car exerted the greater force)

In questions 11-13, students were asked to consider the forces acting at various periods in a coin toss:

Q11: The coin moving upward

Q12: The coin at its highest point

Q13: The coin moving downward

This set of questions was low scoring and had low gains for both samples. The correct answer was that the force is down and constant. The most popular response was up and decreasing, zero, down and increasing; a conflation between force and speed/velocity.

Gains and Overall test scores

At the institutional level, the inventory results may be used to compare the effect of the teaching program on students' conceptual understanding. The 'normalised gain' and 'effect size' measurements, shown in Table 2, provide some information about improvements in student learning. There are some notable results, including CCI from University **C** and University **G**, and FCI-B from University **B**, with relatively low (or negative) gains.

Table 2. Summary data from implementation of concept inventories. All data were collected in 2014 except for B and D which were collected in 2012-2014 and 2011-2012, respectively. B, D and F samples included multiple sections/classes. The ‘C’ and ‘F’ sections for the FCI-C and FCI-D from University B indicate the physics major and non-major cohorts, respectively.

University	Concept Inventory	Pre-test mean% (SD)%	Post-test mean% (SD)%	Both-N	Gain	Effect size
I	CCI	35.8 (16.5)	46.1 (16.9)	261	0.16	0.62
C	CCI	57.9 (22.3)	57.5 (23.1)	26	-0.01	-0.016
F	CCI-A	63.5 (19.7)	67.2 (20.2)	32	0.10	0.19
F	CCI-B	65.1 (19.8)	69.8 (20.6)	37	0.13	0.24
D	CCI	62.2 (19.6)	60 (18.3)	49	-0.06	-0.11
G	CCI	44.0 (16.5)	47.1 (17.8)	608	0.06	0.19
H	FMCE	35.1 (25.9)	37.5 (29)	84	0.04	0.09
D	FMCE-A	30.2 (27.2)	49.5 (30.5)	10	0.28	0.71
D	FMCE-B	44.2 (26)	54.8 (28.6)	424	0.19	0.41
J	FMCE	50.9 (32.3)	85.9 (21.9)	181	0.71	1.08
F	TCS	57.0 (16.1)	66.9 (16.8)	560	0.23	0.61
D	TCS-A	58.6 (16.8)	70.4 (14.4)	55	0.28	0.69
D	TCS-C	54.5 (21.3)	70.1 (13.8)	60	0.34	0.73
D	TCS-B	59.9 (22)	76.0 (13.6)	34	0.4	0.73
D	TCS-D	59.3 (18.9)	71.1 (13.6)	63	0.27	0.69
B	FCI-A 2012	69.6 (21.6)	79 (18.4)	286	0.31	0.43
B	FCI-B 2013	75.3 (19.4)	77.6 (19.4)	234	0.05	0.07
B	FCI-C 2014 C	76.4 (19.1)	80.8 (17)	342	0.19	0.23
B	FCI-D 2014 F	40.3 (20.7)	52.2 (22)	202	0.20	0.58

These data were plotted on a plot emulating the Hake plot in order to begin planning for a database to visualise low, medium and high gains (Figure 5). The demarcations from Hake’s paper are included as a reference only.

Findings of interest from these data include the relatively low gains on the CCI overall, the very high and very low gains on FMCE and the closeness of the TCS results. Following these up was not within the scope of this study, and despite the large sample size, the individual data points are not considerable enough to draw any substantial conclusions. However, these data provide a reasonable starting point to seed discussions of relatively ‘low’ ‘medium’ and ‘high’ gain areas.

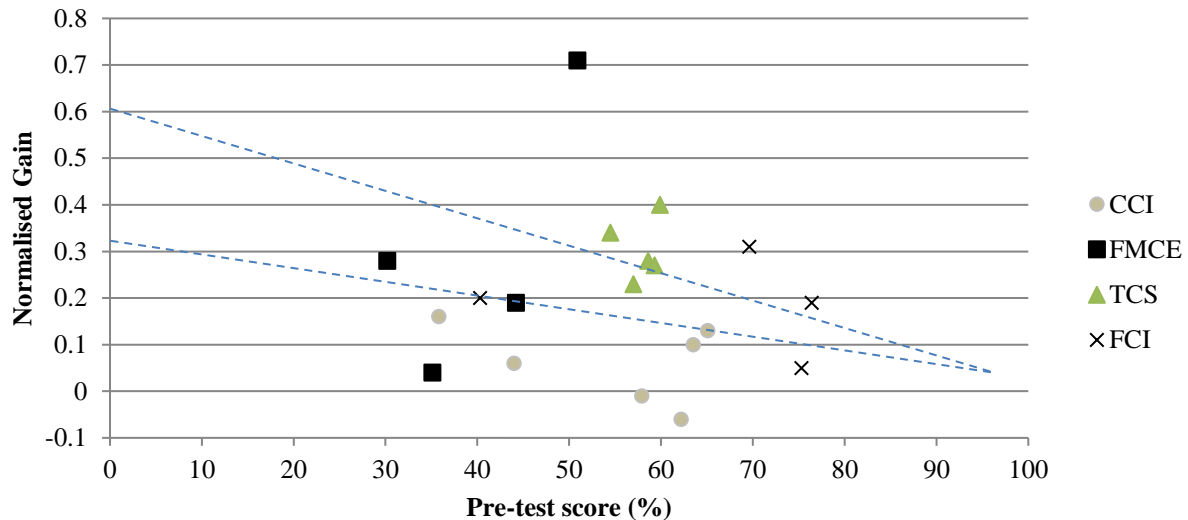


Figure 5: Emulating the Hake plot for the various conceptual inventories used in this study

Discussion

Participants were provided assistance in the use of CIs to engage in evidence-based teaching development to improve student learning. Of the 12 different courses where CI data collection occurred, various levels of engagement with the corresponding participants were achieved. There are many possible advantages of inducing participants who are open to change into evidence-based education (White et al., 2015), however, uptake is reported as difficult to achieve (Borrego & Henderson, 2014). As part of this study, we report that the majority of participants continued to use CIs after the implementation. Two illustrative examples show that this approach has utility in engaging lecturers in evidence-based teaching development and that this development is useful in terms of personal professional development of teaching. The first, participant **F** used and continued to use CIs to establish a base-line and determine the efficacy of a flipped lecture innovation at their institution. For participant **I**, the utility was in the facilitation of the CI use for course assessment, as this university was engaging in a school-wide teaching transformation.

In terms of the details of the CI results themselves, we elaborate on student difficulties which might help in terms of course/subject development. As with the extensive literature on misconceptions in science (e.g., Hill et al., 2015; Malik, Angstromann & Wilson, 2019; Schultz et al., 2017), this study revealed that students, even at university level, still have difficulty understanding basic concepts and these misunderstandings remain even after completing their courses. Though the data do not necessarily reveal novel understanding about individual misconceptions, the large data set offers a unique opportunity to view compiled results, where different outcomes can be considered in the context of others. For instance, they show us that students begin with and develop their conceptual understanding differently in different contexts. These questions are based on concepts that are considered fundamental assumed knowledge at the undergraduate level. Therefore, misunderstandings of these kinds of concepts, if unchecked, may cause conceptual difficulties to compound in further study (Georgiou et al., 2009). Perhaps the most enduring effect of this study is the realisation amongst participants and teaching team members who are not versed in AL and/or CIs that students continue to struggle with basic concepts; clearly shown through CIs. While those

versed in CIs can debate and critique CIs, to obtain lasting change and uptake of AL, student learning difficulties need to be made visible to ‘others’.

Some of the interesting findings that appear in the data include the difference between groups in first year courses and differences across universities, capturing the possibility of the increased efficacy of certain AL approaches and cases of high and low/no gains. Revealing individual misconceptions across the national sample may also provide interesting insights to readers.

The large data set may be used to contribute to a database to establish a baseline. While this study does not have similar ‘power’ to that of the Hake (1988) study, it tentatively offers the possibility of national data on learning gains across different CIs. The Hake study had used data from several thousand students on the national level to delineate ‘low’, ‘medium’ and ‘high’ gain areas as well as provide a direct connection between the use of AL approaches and a shift toward higher course average normalised gains. While this study supports the known premise that increasing engagement through AL improves student learning as ascertained through learning gains, the message for the participants was about the use of CIs as ways to determine that *their* intervention and/or version of flipped learning had been successful. This insight is likely the most powerful lesson from the use of a particular inventory across many institutions. The second lesson is around ‘value and return on investment’; what do we expect as a ‘reasonable’ course average normalised gain on a particular inventory when AL approaches have been utilised? Our tentative attempt raises the notion that the delineation between ‘low’, ‘medium’ and ‘high’ gain areas are likely different for different surveys, hence what is a ‘reasonable’ average normalised gain is nuanced.

More data are needed to establish whether the diagonal lines in the Hake study (Figure 6) for the United States apply for the Australian context (Karim & Cid, 2020), and if they need to be shifted for the other inventories. Despite the fact that more work needs to be done, what is clear is that the course average normalised gain shifting in one way indicates improved conceptual understanding on a particular inventory. At the very least, this inchoate graph allows one to compare with others who have improved course average normalised gains using a particular CI.

Limitations

This research provides a broad view of data collected as part of a National Fellowship. As such, it was important to present the data such that it represented the process and reflected the aim of the Fellowship. This approach would have been different had the research occurred primarily as part of a research project, and thus, a range of limitations must be addressed. In terms of the first aim, related to engaging academics in the use of CIs, we were not able to collect data from all participating academics. It is therefore possible that a self-selection effect overestimates the degree to which CIs were maintained after the intervention. With respect to the second aim, where we present the CI results and plot these data on a Hake graph, we acknowledge that we cannot make claims related to the ‘quality’ of the gains, as Hake did in the original 1998 paper.

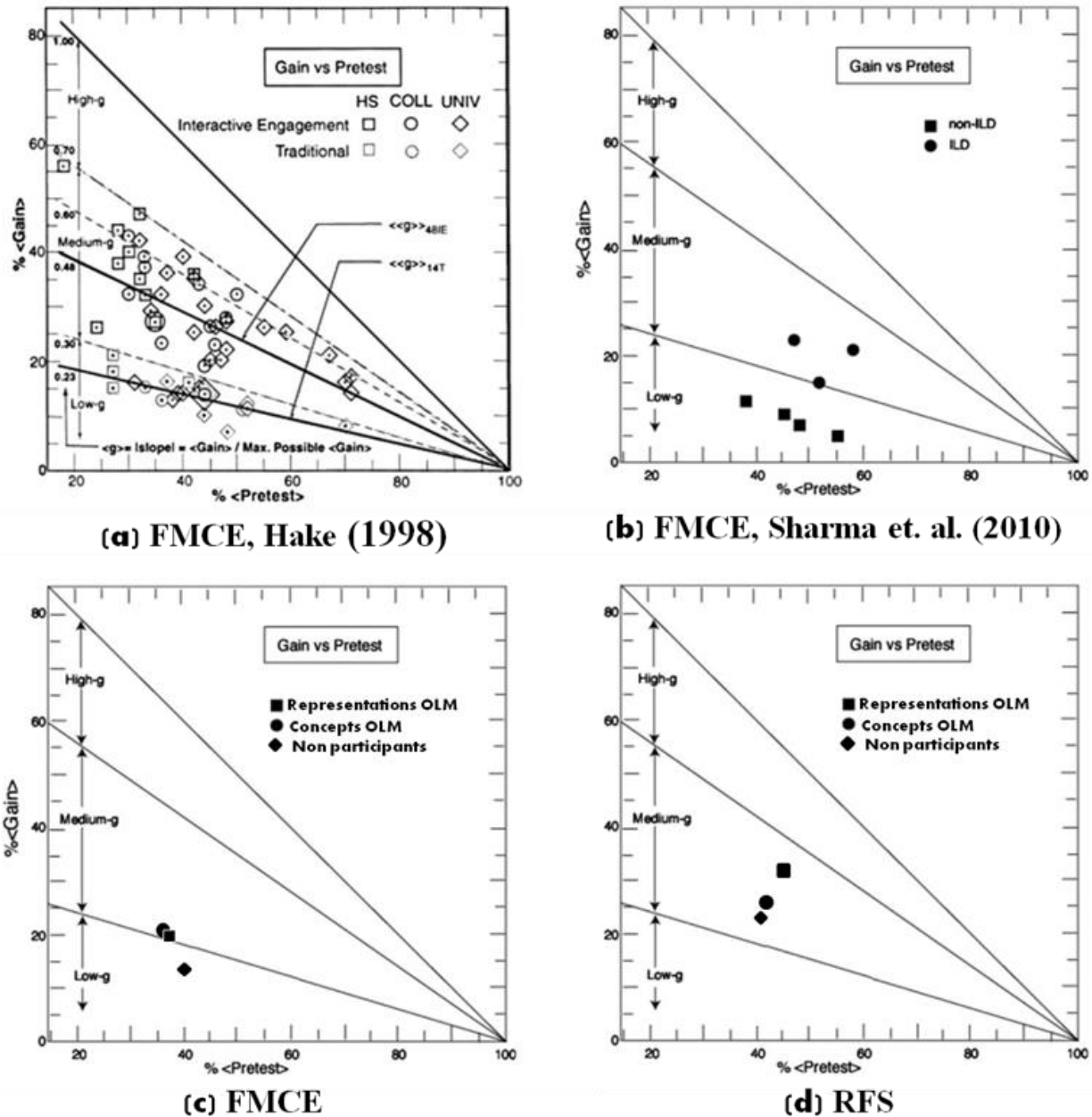


Figure 6: Hake plots indicating course average normalised gains for different inventories from Hill, Sharma and Johnston (2015)

Conclusions and implications

This paper presents data from a considerably large administration of CIs in physics and chemistry at the university level. The purpose of this administration is to engage lecturers in an evidence-based approach to improving courses and therefore student learning outcome. It is known that taking a scholarly and collaborative approach to improving student learning will result in the greatest gains, however, in the university context, there are several unique but significant problems or barriers. In this paper, we aim to remove some of these barriers to improve engagement, which included assisting with the administration of CIs. Results indicated that participants were using CIs more widely/frequently. We also report on the utility of CIs in order to demonstrate how they can be useful, with results illustrating, for

example that significant conceptual difficulties still exist amongst university students, but that these could be shifted in some cases. The results are also presented on a ‘normalised gain’ plot, in order to lay the groundwork for a database of sorts, that reports on low, medium and high gains amongst different courses and universities in Australia. The use of CIs, though not a panacea, offers significant potential in the tertiary science sector, and this study demonstrated how this potential is developed, as a first step.

Acknowledgements

This work was supported by A National Teaching Fellowship for XXXX from the decommissioned Office for Learning and Teaching, funded by the Australian Government. We would like to acknowledge all lecturer and student participants of this study, as well as the reviewers, who provided valuable feedback in the development of this paper.

References

- Borrego, M., & Henderson, C. (2014). Increasing the use of evidence-based teaching in STEM higher education: A comparison of eight change strategies. *Journal of Engineering Education*, 103(2), 220-252.
- Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Princeton University Press, 3175 Princeton Pike, Lawrenceville.
- Coştu, B., Ayas, A., & Niaz, M. (2010). Promoting conceptual change in first year students’ understanding of evaporation. *Chemistry Education Research and Practice*, 11(1), 5-16.
- Dolan, E. L., Elliott, S. L., Henderson, C., Curran-Everett, D., John, K. S., & Ortiz, P. A. (2018). Evaluating discipline-based education research for promotion and tenure. *Innovative Higher Education*, 43(1), 31-39.
- Engelhardt, P. V. (2009). An introduction to classical test theory as applied to conceptual multiple-choice tests. In C. H. Harper & K.A (Eds.), *Reviews in PER: Getting started in PER*. College Park, MD: American Association of Physics Teachers.
- Faust, J. L., & Paulson, D. R. (1998). Active learning in the college classroom. *Journal on excellence in college teaching*, 9(2), 3-24.
- Francek, M. (2013). A Compilation and Review of over 500 Geoscience Misconceptions. *International Journal of Science Education*, 35, 31-64. doi:10.1080/09500693.2012.736644
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1319030111
- Georgiou, H., Sharma, M., O’Byrne, J., & McInnes, B. (2009). *University students’ conceptions about familiar thermodynamic processes and the implications for instruction*. Paper presented at the Uniserve Science, Sydney.
- Georgiou, H., & Sharma, M. D. (2010). A Report on a Preliminary Diagnostic for Identifying Thermal Physics Conceptions of Tertiary Students. *International Journal of Innovation in Science and Mathematics Education*, 18, 32-51.
- Georgiou, H., & Sharma, M. D. (2015). Does using active learning in thermodynamics lectures improve students’ conceptual understanding and learning experiences? *European Journal of Physics*, 36(1), 015020.
- Georgiou, H., & Sharma, M. D. (in press 2021) Enacting and sustaining change in undergraduate STEM education: a multiple case study across 6 universities. *Proceedings of the Singapore National Academy of Science on STEM Education (PSNAS)*
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66, 64-74. doi:10.1119/1.18809
- Hemer, S. R. (2014). Finding time for quality teaching: an ethnographic study of academic workloads in the social sciences and their impact on teaching practices. *Higher Education Research & Development*, 33(3), 483-495.
- Hendry, G. D., Georgiou, H., Lloyd, H., Tzioumis, V., Herkes, S. & Sharma, M.D. (2020). “It’s hard to grow when you’re stuck on your own”: Enhancing teaching through a peer observation and review of teaching program, *International Journal for Academic Development*, <https://doi.org/10.1080/1360144X.2020.1819816>
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The physics teacher*, 30(3), 141-158.

- Hill, M., Sharma, M. D., & Johnston, H. (2015). How online learning modules can improve the representational fluency and conceptual understanding of university physics students. *European Journal of Physics*, 36(4), 045019.
- Johnston, H. M., Hopkins, A. M., Varvell, K. E., Sharma, M. D., & Thornton, R. K. (2007). The research-teaching nexus in physics: scholarship into teaching and learning. *Australian Physics*, 44(2), 66-70.
- Kanim, S., & Cid, X. C. (2020). Demographics of physics education research. *Physical Review Physics Education Research*, 16(2), 020106.
- Lemass, B., & Stace, R. (2010). Towards teaching and research parity: A teaching quality and reward framework. *Perspectives*, 14(1), 21-27.
- Libarkin, J. (2008). *Concept inventories in higher education science*. Paper presented at the BOSE Conf.
- Malik, U., Angstmann, E. J., & Wilson, K. (2019). Learning and Conceptual Change in Thermal Physics Concepts: An Examination by Gender. *International Journal of Innovation in Science and Mathematics Education*, 27(1).
- Mathison, K. (2015). Effects of the performance management context on Australian academics' engagement with the scholarship of teaching and learning: a pilot study. *The Australian Educational Researcher*, 42(1), 97-116.
- Matthews, K. E., Crampton, A., Hill, M., Johnson, E. D., Sharma, M. D., & Varsavsky, C. (2015). Social network perspectives reveal strength of academic developers as weak ties. *International Journal for Academic Development*, 20(3), 238-251.
- Mulford, D. R., & Robinson, W. R. (2002). An inventory for alternate conceptions among first-semester general chemistry students. *Journal of chemical education*, 79(6), 739.
- Nicholls, G. (2004). Scholarship in teaching as a core professional value: what does this mean to the academic? *Teaching in Higher Education*, 9(1), 29-42.
- Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychological methods*, 16(2), 93.
- Probert, B. (2013). Teaching-focused academic appointments in Australian universities: recognition, specialisation, or stratification? Report, Office for Learning and Teaching. Retrieved Feb 28, 2020 from <https://apo.org.au/sites/default/files/resource-files/2013/01/apo-nid35663-1168771.pdf>
- Schultz, M., Lawrie, G. A., Bailey, C. H., Bedford, S. B., Dargaville, T. R., O'Brien, G., ... & Wright, A. H. (2017). Evaluation of diagnostic tools that tertiary teachers can apply to profile their students' conceptions. *International Journal of Science Education*, 39(5), 565-586.
- Sharma, M. D., & Georgiou, H. (2016). More active lecture approaches in science and mathematics: Using expert cultural capital to drive change. *Report submitted to the Australian Government Office for Learning and Teaching*. Retrieved Feb 28, 2020 from <http://www.olt.gov.au/resource-library>
- Sharma, M. D., Johnston, I. D., Johnston, H., Varvell, K., Robertson, G., Hopkins, A., . . . Thornton, R. (2010). Use of interactive lecture demonstrations: A ten year study. *Physical Review Special Topics-Physics Education Research*, 6. doi:020119 10.1103/PhysRevSTPER.6.020119
- Sharma, M. D., & McShane, K. (2008). A methodological framework for understanding and describing discipline-based scholarship of teaching in higher education through design-based research. *Higher Education Research & Development*, 27(3), 257-270.
- Smith, J. I., & Tanner, K. (2010). The problem of revealing how students think: concept inventories and beyond. *CBE-Life Sciences Education*, 9(1), 1-5.
- Thornton, R. K., & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. *American Journal of Physics*, 66, 338-352. doi:10.1119/1.18863
- Treagust, D. F. (1988). Development and Use of Diagnostic-tests to Evaluate Students Misconceptions in Science. *International Journal of Science Education*, 10, 159-169. doi:10.1080/0950069880100204
- Vardi, I., & Quin, R. (2011). Promotion and the Scholarship of Teaching and Learning. *Higher Education Research & Development*, 30(1), 39-49.
- Wallace, C. S., & Bailey, J. M. (2010). Do concept inventories actually measure anything. *Astronomy Education Review*, 9(1), 010116.
- Wattanakasiwich, P., Talaeb, P., Sharma, M., & Johnston, I. (2013). Construction and Implementation of a Conceptual Survey in Thermodynamics. *International Journal of Innovation in Science and Mathematics Education*, 21(1).
- White, P. J., Larson, I., Styles, K., Yuriev, E., Evans, D. R., Rangachari, P., . . . Davie, B. (2015). Adopting an active learning approach to teaching in a research-intensive higher education context transformed staff teaching attitudes and behaviours. *Higher Education Research & Development*, 1-15.