Reviewing affective, behavioural, and cognitive learning gains in higher education

Jekaterina Rogaten[1], Bart Rienties[1]*, Rhona Sharpe[2], Simon Cross[1], Denise Whitelock[1], Simon Lygo-Baker[3], Allison Littlejohn[1]

[1] Institute of Educational Technology, Open University UK, MK7 6AA, Milton Keynes, United Kingdom.

[2] Department of Technology Enhanced Learning, University of Surrey, GU2 7XH, Guildford, United Kingdom.

[3] Department of Higher Education, University of Surrey, GU2 7XH, Guildford, United Kingdom.

* Corresponding author: bart.rienties@open.ac.uk

The notion of learning gains has increasingly become apparent within the higher education literature and is gaining traction in government policies in the US and UK, reflecting a desire to "quantify" the impact of university study and teaching excellence in particular. Given the increased policy focus on learning gains, it is essential that researchers and policy makers are informed about common used approaches to measure learning gains, and their expected range and magnitude. Therefore, a systematic literature search was undertaken. 52 studies (n = 41,009) were coded into affective, behavioural, and/or cognitive learning gains.
The review found a rich but diverse variety of adopted methodologies and approaches to "measure" Affective, Behavioural, and Cognitive (ABC) learning gains. Nonetheless, there is a lack of consistency in the ways in which learning gains are currently measured and reported. These inconsistencies and limitations hamper effective comparisons of learning gains and teaching excellence. We recommend a greater emphasis on longitudinal measurement of learning gains using validated approaches.

Keywords: Learning gains, higher education, cognitive measures, affective measures, behavioural measures.

Reviewing affective, behavioural, and cognitive learning gains in higher education

The notion of learning gains has increasingly become apparent within the higher education literature and is gaining traction in government policies in the US and UK, reflecting a desire to "quantify" the impact of university study and teaching excellence in particular. Given the increased policy focus on learning gains, it is essential that researchers and policy makers are informed about common used approaches to measure learning gains, and their expected range and magnitude. Therefore, a systematic literature search was undertaken. 52 studies (n = 41,009) were coded into affective, behavioural, and/or cognitive learning gains.
The review found a rich but diverse variety of adopted methodologies and approaches to "measure" Affective, Behavioural, and Cognitive (ABC) learning gains. Nonetheless, there is a lack of consistency in the ways in which learning gains are currently measured and reported. These inconsistencies and limitations hamper effective comparisons of learning gains and teaching excellence. We recommend a greater emphasis on longitudinal measurement of learning gains using validated approaches.

Keywords: Learning gains, higher education, cognitive measures, affective measures, behavioural measures.

Introduction

Over the years researchers and practitioners have been developing and testing a range of measurement approaches aiming to effectively capture improvements in student learning in higher education (Hake 1998; Cahill et al. 2014; Coates 2016). One approach commonly used is the notion of learning gain, which is defined as growth or change in knowledge, skills, and abilities over time that can be linked to the desired learning outcomes or learning goals of the course (e.g., Cronbach and Furby 1970; Arum and Roksa 2014; Rogaten, Rienties, and Whitelock, 2017; Roohr, Liu, and Liu 2017). Recent calls about the need for clear and transparent measures of learning to assess the value of higher education in both the United States (U.S. Department of Education 2006) and the UK (Department for Business, Innovation & Skills 2016) have promoted the use of learning gains. Several large scale national studies (Wabash National Study of Liberal Arts Education 2012; Higher Education Funding Council for England 2015) are using learning gains in order to investigate factors that impact educational outcomes. Educational researchers can contribute to these national debates by investigating how learning gains can be validly and reliably measured on a large scale, and to what extent learning gains can appraise the quality of education provision and provide evidence of impact on graduate outcomes.

Although the concept of learning gains is not new and has been actively debated in psychology and education fields (e.g., Dimitrov & Rumrill Jr, 2003; Bao, 2006; Rocconi & Ethington, 2009), learning gains and the ways in which researchers have measured them, and the reported magnitude of learning gains seems to be widely diverse. For example, in one of the largest longitudinal studies on learning gains in the US, the Wabash National Study of Liberal Arts Education, a range of learning gains (e.g., moral reasoning, critical thinking) were measured amongst 17000 students across 50 US colleges and universities at three separate time intervals (Arum and Roksa 2014). Pascarella and Blaich (2013) found that although several groups of students reported positive learning gains in critical thinking, moral reasoning, engagement, and responsible leadership, other learning gains were negligible, such as positive attitudes to literacy, and political and social involvement. In contrast, Andrews, Leonard, Colgrove and Kalinowski (2011) assessed cognitive learning gains amongst 8000+ biology students using two "objective" knowledge tests and found only small average learning gains.

Despite the increase in interest and volume of research examining learning gains and increased attention by policy makers to measure the "value added" of education, there is no comprehensive and systematic literature review on learning gains in higher education. As such, this review aims to fill this gap by analysing empirical research findings reported in the literature into students' learning gains. More importantly, we will provide three recommendations how learning gains may be applied for assessing teaching excellence.

Categorising different types of learning gains

In line with Everson (2017), one of the challenges in conducting a systematic literature review is in examining what are the similarities and differences among what learning gains were measured and how they were assessed. Given the increased emphasis on higher education's role in developing appropriate attitudes and skills in addition to cognitive development, an increasing number of researchers have started to move beyond focussing on measuring cognitive learning gains (Douglass, Thomson, and Zhao 2012; Pascarella and Blaich 2013; Emke, Butler, and Larsen 2016).

In order to be able to distinguish the different categories and approaches that learning gains studies have developed and to make sure that comparisons can be made among studies looking at similar types of learning gains, we propose that learning gains research needs a coherent classification system to be able to distinguish a more nuanced understanding of different types and approaches to learning gains. In psychology and education, a large number of studies have used the lens of Affective, Behaviour, and Cognitive (ABC) learning to understand, unpack and explain the complex, multidimensional notions of learning and cognitive development (e.g., Ostrom, 1969; Panksepp, 2003; Jindal-Snape & Rienties, 2016). For example, Ostrom (1969) used the ABC approach to classify different attitudes, while Panksepp (2003) used ABCs to classify brain activity. Therefore, in this article we will apply an ABC classification to distinguish three types of learning gains: 1) affective learning gains which encompass change in attitudes, satisfaction, and well-being, 2) behavioural learning gains which encompass changes in students' behavioural skills over time, such as engagement, leadership skills, team work and 3) cognitive learning gains which encompass gains in abilities related to cognitive development, such as understanding, knowledge, critical, analytical and scientific thinking. Of course we acknowledge that these ABC learning gains are interrelated, and should be seen both from an individual classification as well as from an intertwined, holistic perspective.

Affective learning gains

Several researchers have looked into measuring affective learning gains, which can be defined as a change in affect related states during a course, such as confidence, motivation and attitudes. For example, Beck and Blumer (2012) measured students' confidence in designing an experiment in their laboratory class in a sample of 38 biology students. Confidence was measured using a 12-item self-reported survey administered at the beginning and end of a semester. Similarly, Mathabathe and Potgieter (2014) measured students' knowledge in stoichiometry and how confident 91 students in science and engineering were in their knowledge. They found that overall students' confidence in their knowledge of stoichiometry improved, but the results also highlighted that gains in confidence is a complex issue. Mathabathe and Potgieter (2014) found that a large proportion of students were overconfident in their knowledge, and overconfidence increased over time, whereas realistic confidence (confidence corresponding with the level of knowledge) decreased as a result of the course. Importantly, this study demonstrated that affective learning gains had an effect of cognitive learning gains, as students who became overconfident in their knowledge showed lower knowledge gains.

Behavioural learning gains

Recently researchers have extended the focus within learning gains field to also include behavioural learning gains, which focus more strongly on skills than knowledge, and include for example engagement, leadership skills, study skills, or team working skills (e.g., Strijbos, Engels, & Struyven, 2015; de Hei, Strijbos, Sjoer, & Admiraal, 2016). For example, Varsavsky et al. (2014) looked at learning gains in team working amongst 400 science students developing skills to work with others to accomplish a shared task, whereby students reported improvements in teamwork skills, but they rated the importance of the skill significantly higher than their improvement. Similarly, Cabrera, Colbeck and Terenzini (2001) assessed learning gains in group work skills in 1,258 Engineering students from seven different schools using students' retrospective judgements about their learning gains, and found that students reported moderate learning gains in group work skills.

Cognitive learning gains

Finally, cognitive learning gains can be defined as development in knowledge, understanding, and cognitive abilities. The "classic" way to measure cognitive learning gains is to assess changes in knowledge or conceptual learning using existing standardized tools at the start and end of a semester (Roohr, Liu, and Liu 2017). This method is particularly popular in the "harder" sciences, where acquisition and application of knowledge are the expected outcome of study (Gok 2012; Pentecost and Barbera 2013; Cahill et al. 2014).

In addition to objectively measuring cognitive learning gains, recently several researchers have adopted different, perhaps subtler approaches to measure cognitive learning gains. For example, Douglass, Thomson and Zhao (2012) measured students' perceptions of desired learning gains in quantitative skills, writing and reading skills, critical thinking, and knowledge in a sample of 12,500 sciences and humanities students using the Student Experience in the Research University Survey (SERU-S). In all aspects students reported positive learning gains, i.e., increases of quantitative skills, writing, reading, critical thinking, and knowledge in the field. Coil, Wenderoth, Cunningham and Dirks (2010) assessed learning gains of 196 students in scientific writing, reading, metacognitive monitoring of own learning and satisfaction in biological sciences. They found an overall improvement in students' writing and reading.

Research questions

As highlighted by the studies above, a wide range of learning gains approaches, research designs, and measurements have been employed in the last two decades. Some researchers used objective testing and employed pre-post test designs, while others relied on students' self-reported retrospective judgement of own learning gains, making it difficult to draw direct comparisons between numerous studies assessing affective, behavioural, and/or cognitive learning gains. As such, in order to reach a better understanding of learning gains in higher education, it is important to develop a critical, systematic overview of the learning gains literature. Our ABC classification approach aims to categorise, contrast, and unpack the range of learning gains approaches used in higher education. In research question 1 we will distinguish the kinds of approaches used to measure learning gains (i.e., pre-post objective testing, pre-post subjective testing, cross-sectional testing). In research question 2 we will focus on the reported magnitudes of learning gains across the ABC categorisation.

RQ1: What kinds of approaches are used to measure affective, behavioural, and cognitive learning gains in higher education?

RQ2: What is the magnitude of affective, behavioural and cognitive learning gains in higher education?

Method

Search strategies

The literature search was conducted in the period of February-April 2016 in ERIC and Web of Science core collection. In addition, an ancestry approach was used to exhaustively search for peer-reviewed empirical studies in the reference lists of relevant articles. The time frame of the search was set between 2000 and 2016. The keywords search included: “learning gain*”, higher education, college, graduate, not school, not child. No other preconditioned criteria were used in the search. In total 231 studies were identified, of which 208 were from Web of Science collection, and 23 studies were from ERIC. The first selection included peer reviewed theoretical articles, empirical studies, qualitative and case studies, conference papers, PhD theses, books, and book chapters. The next step of our review included a comprehensive evaluation of the abstracts, resulting in 20 studies being excluded as they did not assess learning gains.

Inclusion/exclusion criteria

The remaining 211 potentially relevant articles were then reviewed using the following selection criteria:

1) Empirical studies which used learning gain as a dependent variable.
2) The sample comprised of undergraduate and/or postgraduate students studying for a higher education qualification.
3) Studies of at least three weeks duration. As such, most experimental/lab studies that used learning gain as dependent variable were excluded.
4) Sufficient quantitative data reported. At least Mean and Standard Deviation or Standard Error for the dependent variable was reported in a numeric format (i.e., studies that only included graphical representations where no accurate values could be obtained were excluded).
5) The student sample was not used in any other study where learning gain was a dependent variable.

After a careful review of the potentially relevant articles, 52 non-overlapping empirical studies met all the inclusion criteria. The articles that were excluded failed to meet the selection criteria. Within the selected studies a total of 114 independent student samples were identified totalling 41,009 higher education students.

Coding

A coding scheme for classifying the key variables from each of the research papers was developed to reflect differences in methods used within learning gains research, and identify whether the respective study assessed affective, behavioural and/or cognitive learning gains. Firstly, the initial coding scheme was tested and adjusted with four studies being coded by five researchers to determine the validity of the coding scheme. Secondly, 23 studies were double coded to determine the reliability of the revised coding scheme, whereby there was 100% agreement in coding learning gains as affective, behavioural and cognitive, and an average of 82% agreement on coding method, measures, computation of learning gain, student sample, time between pre-test and post-test. The remainder of studies were coded by the first author and any discussion points were agreed with the second author.

Affective learning gains included learning gains in attitude, confidence, enjoyment, enthusiasm for a topic, feeling comfortable with complex ideas, interest in a topic, motivation, satisfaction, and self-efficacy. Behavioural learning gains included ability to work independently, applied conceptual understanding, effort and engagement, leadership skills, learning gains in team/group working skills, practical competence, resource management, responsibility, preparation skills, and time management skills. Cognitive learning gains included learning gains in students' ability to evaluate and create knowledge, analytical, autonomous cognition, critical and ethical thinking, creative and higher order thinking, discipline specific skills, knowledge and understanding of the topic, oral and written communication, problem solving, scientific reasoning, and statistical and research skills/knowledge. Overall, there were twenty eight studies that examined only cognitive learning gains, twelve studies that examined a combination of affective and cognitive learning gains, five studies that examined a combination of cognitive and behavioural learning gains, and seven studies examined a combination of affective, behavioural and cognitive learning gains.

Analysis

In order to reflect the diversity of learning gains research, all studies were firstly separated into self-reported measures and objective measures of learning gains (e.g., standardised multiple choice tests). In order to ensure consistency, all self-reported learning gains were converted into percentages for cross-sectional studies. Following that, the average percentage of self-reported learning gains were computed separately for affective, behavioural, and cognitive learning gains.

For studies that used pre-post test design, average normalised learning gains (Hake 1998) were computed. In the studies included in this review, only cognitive learning gains were assessed using both self-reported and objective testing, whereas behavioural and affective learning gains were only assessed using self-reported surveys, as no studies were available that reported on objective testing. Average normalised learning gains were estimated for each study using means of pre-test, post-test and maximum post-test scores (Hake 1998). Although the average normalised learning is usually computed using normalised gain for each student, the computation using overall group means is also acceptable and is similar to the average normalised gain based on individuals' scores if the sample size is above 50 students (Bao 2006). The average normalised gain is considered small if it is below 0.3, moderate if it is between 0.3 and 0.7, and high if it is above 0.7 (Hake 1998).

Results

Main characteristics of selected studies

In total 52 empirical studies were included totalling 41,009 higher education students. The majority of student samples were from US universities (77%; e.g., Andrews et al., 2011; Pentecost & Barbera, 2013), while 8% were from Australian universities (e.g., Varsavsky et al., 2014; Hill, Sharma, & Johnston, 2015). Other studies originated from Turkey (Gok 2012; Yalaki 2010), Kuwait (Anderson 2006), Germany (Woltering et al. 2009), Mexico (Shuster and Peterson 2009), South Africa (Mathabathe and Potgieter 2014), Taiwan (Cheng, Liang, and Tsai 2015) and China (Liu, Liu, and Chi 2014). Most studies focussed on undergraduate courses, in particular first year courses. In terms of RQ1, the main distinction between the studies was their methodology, as indicated in Table 1. For a detailed analysis for each study included, and their respective research design(s), we refer to Table 2.

Pre-post tests were the most common method used in 36 studies totalling 79 student samples (70% of all student samples; e.g., Georgiou & Sharma, 2015; Emke, Butler, & Larsen, 2016). A comparison between two samples of students were made in 23 studies (e.g., Cahill et al., 2014; Mortensen & Nicholson, 2015) totalling 64% of all student samples. In addition, six studies used multiple samples (e.g., Pentecost & Barbera, 2013; Campbell et al., 2014;) totalling 23% of all student samples. Finally, 10 studies used single student samples (e.g., Andrews et al., 2011; Buriak & Potter, 2014) totalling 13% of all student samples. All research studies that used pre-post design could be further divided into those that used the same measure/test during pre-test and post-test assessments (65 student samples, 85.5%; e.g., Andrews et al., 2011; Cahill et al., 2014), and those that used similar/compatible measures/tests (11 student samples 14.5%; e.g., Getha-Eby, Beery, O'Brien, & Xu, 2015; Emke et al., 2016). The majority of studies using pre- post test design assessed learning gains through standardised and objective tests/inventories (61 student samples; 80%) and a smaller proportion of studies used self-reported questionnaires or a combination of objective tests/inventories and self-reported questionnaires (15 student samples; 20%).

➔ Insert Table 1 about here

In total, 18 studies comprising of 35 student samples (31% of all student samples) used a cross-sectional design and self-reported questionnaires to assess students' learning gains. Self-reported perception of learning gains were mainly measured using validated self-reported scales (25 student samples, 71%) like CLASS (e.g., Gok, 2012; Cahill et al., 2014), or SALG (e.g., Gill & Mullarkey, 2015; Ojennus, 2016). Comparatively fewer studies used self-reported questionnaires that were developed by a research team to specifically assess their students' learning gains (10 student samples, 29%; e.g., Liu et al., 2014; Matthews, Adams, & Goos, 2015).

Overall, as is clear from Table 1 most studies identified in this review focussed on cognitive learning gains, followed by behavioural, and affective learning gains. While recently more studies have focussed on combining cognitive with affective and behavioural gains, Table 1 clearly highlights that no affective or behavioural studies have used pre-post objective measurements of learning gains. This is an important methodological research design omission, given that universities are increasingly pressed to provide "solid" evidence of affective and behavioural gains beyond cognitive gains to both local and national governments, as well as students and parents as customers of university services (Woodall, Hiller, and Resnick 2014; Everson 2017).

Outcome of studies using self-reported retrospective measures

In order to address RQ2, we first will unpack 28 studies using self-reported measures (51 independent samples, n = 19,509), after which we will unpack 22 studies using objective measures (39 student samples, n = 18,024). Of the 28 self-reported studies, thirty-five samples were assessed using cross-sectional design (e.g., Douglass, Thomson, & Zhao, 2012; Matthews et al., 2015), sixteen samples were assessed in pre-post test design (e.g., Hatch et al., 2014; Stolk & Martello, 2015), out of which eight samples were assessed using both standardised tests and self-reported measures and both pre-post testing and cross-sectional design (e.g., Mathabathe & Potgieter, 2014; Mortensen & Nicholson, 2015).

Affective self-reported learning gains

Affective learning gains were measured in 19 studies (e.g., Moorer, 2009; Strayhorn, 2010) comprising 28 student samples totalling 3,333 higher education students. For example, Cahill et al (2014) examined affective learning gains by assessing attitudes towards physics and learning physics of 921 undergraduate physics students using a 42-item self-reported Colorado Learning Attitudes about Science Survey (CLASS). The results showed a decrease in positive attitude towards physics, which ranged between -3.9% to -7.6%. Furthermore, Stolk and Martello (2015) looked at 114 undergraduate science students' learning gains in motivation using the Situational Motivation Scale (SIMS). Students' motivations were monitored throughout the semester and results showed that mean SIMS subscale scores for intrinsic motivation (range 4.91 and 5.06), identified regulation (range 4.87and 4.98), and self-determination (range 7.97 and 8.49) were significantly higher than for external regulation (range 3.26 and 3.34) and amotivation (range 1.73 and 1.63), indicating that students reported higher learning gains in adaptive motivational orientation.

Overall, there were nine affective learning gain studies that employed pre-post test designs. Most research reported positive change, but three studies reported negative change. The average normalised gain for those studies that reported affective learning gain was $\langle g \rangle = 0.39$ which is according to Hake (1998) a moderate learning gain. There were 10 cross-sectional studies totalling 1,772 students who on average reported 77.7% affective learning gain. As highlighted in Figure 1a, a wide range of affective learning gains was reported amongst these 19 studies, ranging from -15% to 98%. Cross-sectional studies reported on average higher affective gains than pre-post designs.

➔ Insert Figure 1a b and c about here

Behavioural self-reported learning gains

Behavioural learning gains were measured in 13 studies (e.g., Casem, 2006; Varsavsky et al., 2014) comprising 23 student samples totalling 4,268 higher education students. With the exception of one study (Stolk and Martello 2015), the remaining 12 studies used a cross-sectional design for measuring behavioural learning gains. In these cross-sectional studies, students on average reported 77.0% behavioural learning gain. Although there are fewer studies which used behavioural change as a measure of learning gains, several researchers have argued that skills development is just as important as knowledge acquisition. Varsavsky and colleagues (2014) and Stolk and Martello (2015) both criticised higher education courses for continuing to prioritize decontextualized knowledge over learning gains in skills (Stolk & Martello, 2015), even though students rated the importance and usefulness of competences, such as teamwork, higher than subject specific knowledge (Varsavsky, Matthews, and Hodgson 2014). As indicated in Figure 1b, again a substantial range of learning gains were reported in these 13 studies, ranging from 38% to 95%.

Cognitive self-reported learning gains

Cognitive learning gains were measured in 22 studies (e.g., Seymour, Daffinrud, Wiese, & Hunter, 2000; Ojennus, 2016), comprising 39 student samples, totalling 18,024 higher education students. Pre-post testing was used in four studies (Wattiaux and Crump 2006; Lim, Hosack, and Vogt 2012; Hatch et al. 2014; Stolk and Martello 2015;), and two studies used a form of pre-post testing through reflection (Douglass, Thomson, and Zhao 2012; Nagel et al. 2012) all totalling to seven student samples. Only in one sample (Stolk and Martello 2015) did students report lower cognitive ability at the post-test than at the pre-test, but the difference was not significant. As such, this particular sample was not included in the remainder of the analysis. The average of normalised learning gain estimated on six independent pre-post test student samples was <g> = 0.34, which is considered moderate gain (Hake 1998).

The rest of the studies (32 samples) used cross-sectional designs and assessed students' perceptions of their cognitive learning gains. On average, students reported 74.0% in cognitive learning gain. A wide range of reported % learning gains was found, ranging from 11% to 96%. Studies that used self-reported retrospective measures of learning gains reported higher learning gains in cross-sectional studies than in pre-post test design studies. Furthermore, given that only (some) cognitive learning gains studies used objective pre-post measurements of learning gains, it is perhaps not surprising that these cognitive learning gains were on average lower than self-reported affective and behavioural learning gains, see Figure 1c.

Cognitive learning gains measured by objective testing

As was indicated before in Table 1, objective testing was only used on cognitive learning gains using pre-post design. In total, 32 studies used objective testing analysing 71 student samples totalling 22,004 higher education students. Out of 32 studies five used a combination of self-reported and objective measures of learning gains totalling eight student samples (12%), while six studies used other types of non-standardised testing totalling 10 student samples (9%). Cognitive learning gains were mainly measured by standardised tests, such as the Chemical Concepts Inventory (Pentecost and Barbera 2013) or Force and Motion Concept Evaluation (Pollock 2006; Hill, Sharma, and Johnston 2015). For example, Gok (2012) examined students' conceptual learning gains of electricity and magnetism in 138 physics students using a Conceptual Survey of Electricity and Magnetism (CSEM) at the beginning (pre-test) and end (post-test) of a semester. The results showed that students' learning gains *g* ranged from 0.36 to 0.62, showing medium learning gain. Using a much larger sample, Pentecost and Barbera (2013) looked at learning gains in content knowledge and reasoning in 2,392 undergraduate chemistry students from four different universities. They used the Chemical Concepts Inventory which consisted of 22 multiple choice questions and found that across universities g only ranged from 0.04 to 0.14, indicating low learning gain.

Overall, the results of pre-post testing showed positive cognitive learning gains, with the exception of two studies that used graded assessment test results (academic performance) to assess students learning gains (Yalaki 2010; Jensen, Kummer, and Godoy 2015). Average normalised learning gain was $\langle g \rangle = 0.34$ which is considered a moderate learning gain (Hake 1998).

As is widely debated in education, ideally a particular educational treatment should be compared and contrasted with a comparison or control condition in order to determine the direction, and size of the potential positive or negative effect of that treatment on learning processes and academic outcomes (Cook & Campbell, 1979; Torgerson & Torgerson, 2008). Therefore, as a final part we analysed data of studies that included a comparison/control condition. Unfortunately, none of the reviewed studies used a Randomised Control Trial (RCT) design, and primarily relied on comparing the treatment group with an alternative group (e.g., a previous cohort, another class taught by a different teacher). Follow-up analyses of treatment studies which compared a change in curricular or module design (treatment) enhanced students learning in comparison to traditional lectures (control) found that students performed better in the treatment condition <g> = 0.39 than in control condition <g> = 0.26. Again a wide range of learning gains were found, whereby average normalised learning gains *g* ranged from -.20 to .81, see Figure 2 for detailed breakdown. In other words, substantial variation in reported learning gain results were found in more experimentally designed studies, highlighting that the magnitude of ABC learning gains is wide, diverse and complex.

➔ Insert Figure 2 about here

Discussion

The main aim of this systematic literature review was to map out the research on learning gains in higher education in order to develop a more comprehensive understanding of reported learning gains in the literature, and to assess the potential of learning gains approaches as an "appropriate" measure of the value of higher education (Woodall, Hiller, and Resnick 2014; Everson 2017; Roohr, Liu, and Liu 2017). This is especially urgent for a UK context, as the current government aims to link financial support to universities with (self-reported) measures of teaching excellence, which in part may be measured by (self-reported and objective) learning gains. Without appropriate, validated, and relevant measures of generic and/or discipline-specific learning gains which are tested and validated in robust experimental as well as naturalistic design studies, using "inappropriate" proxies of students' learning gains might lead to severe financial and reputational implications for universities, and sub-optimal provision of teaching excellence to students.

Following a robust and thorough coding and analysis of affective, behavioural, and cognitive learning gains across 52 empirical learning gain studies, our systematic review found that most studies focussed on cognitive learning gains, mostly using pre-post objective testing. However, there were some studies examining a combination of affective, behavioural, and cognitive learning gains, but in their majority they employed cross-sectional design, and mostly relied on self-reported retrospective estimates of learning gains, with obvious potential limitations (RQ1).

In terms of RQ2, for the cognitive dimension the average magnitude of learning gains in studies using “objective” and retrospective self-reported measures was moderate. In other words, students on average reported moderate growth in their cognitive ability, primarily knowledge, understanding, higher order thinking, metacognition, problem solving, scientific reasoning, discipline specific skills, oral and written communication, statistical and research skills/knowledge, autonomous cognition and metacognition. Similarly, for the affective and behavioural dimension, students on average also reported moderate learning gains in cross-sectional and pre-post test design studies. Interestingly, a number of studies found negative affective gains, which were mainly observed on the measures of learning attitude and attitude towards the subject (Lim, Hosack, and Vogt 2012; Cahill et al. 2014).

As evidenced by our systematic review and the detailed breakdown of varied research designs in Table 2, the 52 studies reported a vast range of learning gains, highlighting on the one hand the inherent and well-recognised complexity of higher education teaching practice (Beetham and Sharpe 2013; Rienties, Brouwer, and Lygo-Baker 2013; Richard Arum and Roksa 2014; Woodall, Hiller, and Resnick 2014; Coates 2016; Roohr, Liu, and Liu 2017), and on the other hand the methodological complexities of defining what actually constitutes a learning gain, and how this could potentially be measured appropriately (Pascarella and Blaich 2013; Rogaten, Rienties, and Whitelock 2017; Everson 2017). The reported magnitude of learning gains amongst these 52 studies was surprisingly diverse and wide ranging. At the same time, as highlighted by Everson (2016) we need to make it clear to teachers and students what we mean when measuring value added.

In terms of research design options, the results of our systematic review indicated that learning gains seemed more visible with a pre-post test design, regardless of whether objective or self-reported measures were used. These findings suggest that asking students whether they have developed knowledge, understanding and certain skills and competences might be as accurate as testing them using “objective” tests. One possible factor influencing pre-post self-report of learning gains is students’ ability to detect and reflect on their own progress. It is possible that when students are asked to assess their learning gains at the beginning of a course, it may help students to establish a baseline of their knowledge, skills and competences prior to any learning taking place (Douglass, Thomson, and Zhao 2012). In addition, whether or not students would have gained in terms of ABC learning gains over time, irrespective of the educational treatment, often could not be assessed in the majority of reviewed research designs.

As indicated previously, relatively few studies used quasi-experimental designs by introducing a control or comparison condition. Those studies who did found on average some positive cognitive learning gains, but the average effect size of the treatment vs. control condition was relatively small (<Δg> = .14). Furthermore, a strong divergence in the direction and magnitude of reported cognitive learning gains were found in these treatment vs. control studies. Of course without a randomisation of treatment vs control conditions, as done in Randomised Control Trials to prevent obvious Hawthorne or John Henry potential biases, the jury might still be out in terms of the actual size and scope of these treatment studies. In other words, how learning gains were measured (i.e., pre-post, post-test only; self-reported vs objective; widely validated vs. specific own-developed instruments), with which research design (e.g., cross-sectional, quasi-experimental, RCT), and which types of learning gains were targeted (affective, behaviour, cognitive) could have a substantial influence on the reported effect size, and robustness of the findings.

Limitations

This systematic literature review has a number of limitations that should be considered when interpreting its findings. Firstly, this review is limited by the inclusion criteria that were applied. As such, the findings can only be interpreted within the empirical learning gains research in higher education. Given the richness and complexities of measuring and unpacking learning gains, we would encourage researchers to conduct a similar systematic literature review of qualitative and mixed method learning gains studies, as well as studies that use similar notions of learning gains but who do not label these as learning gains. In particular in the UK context there is currently an emergence of conceptual and critical qualitative research (e.g., Kinchin, 2016; Ashwin, 2017; Boud, 2018) that provide some alternative perspectives on the primarily quantitative approaches to learning gains, and how policy makers in particular should start to make sense of these methodologically challenging concepts.

Secondly, the results of our analysis can only be as good and robust as methodological rigour of selected studies. All studies in our review were either using single sample of students or conducting a quasi-experimental research. There were no studies that used randomised control trials and it is not clear whether observed learning gains would be the same if more rigorous methodologies were used. Finally, research on learning gains is varied in methodology and measures and as such comparisons between studies are hard to make. The approach employed in this research is just one way of possible synthesis of research. However, the developed coding scheme was extensively validated and provided a useful perspective of the different employed learning gains methodologies and ABC focus.

Conclusion and recommendations to enhance learning gains research

As interest in the notion of learning gains has increased both within higher education providers and their policy makers and funding bodies, it is important to develop a greater understanding of how to interpret reported/observed learning gains. As evident from our fine-grained coding of the 52 studies, there were significant variations in the approaches taken and the reported impact of these. Given the wide disparity in reported learning gains studies, we encourage policy makers and researchers to be extremely cautious when interpreting reported learning gains, as our findings clearly highlight a lack of standardisation of used approaches and how learning gains results are reported.

Our first recommendation is that universities should consider to use pre-post measurements of learning gains in order to strengthen the methodological rigor of learning gains research. We encourage researchers to use appropriate pre-post tests, ideally those developed and validated across a range of contexts. Given the inevitable loss of data in longitudinal data collection, appropriate strategies will need to be put into place in terms of appropriate sampling techniques and non-response bias checks.

Our second recommendation is for researchers and policy makers to examine learning gains longitudinally throughout the entire student journey. As evidenced by our review, most of the studies examined learning gains within the limited time of a semester or two semesters, and mostly looked at introductory level courses. As such, it is not possible to infer that the same magnitude of learning gains is linear throughout a whole degree experience, as Roohr, Liu, and Liu (2017) found that learning gains may change over time. While we appreciate that this might be an expensive, complex, and a long-term endeavour, our systematic review highlights substantial variations in reported learning gains that may be the result of the respective time-scales used, the instruments employed.

Finally, our third recommendation is an urgent need for researchers and policy makers to start to develop, test, implement, and evaluate pre-post objective measurements of affective and behavioural learning gains. Given the importance of developing graduate skills and attitudes of our students, it is essential to develop effective and reliable measurements of affective and behavioural learning gains. To conclude, given the myriad ways of how researchers, managers and policy makers define and construct the concept of learning gains, we encourage that all stakeholders work together to construct a clear, inclusive definition of what learning gains are (and what they are not), and how one might be able to measure them to identify whether students' have actually improved in their ABCs, and whether this is a result of teaching excellence (or not)...

Acknowledgements

This literature review was conducted as part of a "A longitudinal mixed-method study of learning gains: applying ABC framework at three institutions" project funded by Higher Education Funding Council for England (HEFCE)/Office for Students (OFS) as part of its wider work on learning gain (to find out more see the OFS website: https://www.officeforstudents.org.uk/advice-and-guidance/teaching/learning-gain/)

References

Allen, P. E., and J. E. Cockman. 2009. "The 'RIPL' Effect on Learning Gains in Lecture." In *2009 Physics Education Research Conference*, edited by M. Sabella, C. Henderson, and C. Singh, 1179:73–76. Melville: American Institute of Physics. http://dx.doi.org/10.1063/1.3266757.

Anderson, A. 2006. "An Assessment of the Perception of Learning Gains of Freshmen Students in an Introductory Course in Nutrition and Food Science." *Journal of Food Science Education* 5 (March): 25–30. https://doi.org/10.1111/j.1541-4329.2006.tb00078.x.

Andrews, T. M., M. J. Leonard, C. A. Colgrove, and S. T. Kalinowski. 2011. "Active Learning Not Associated with Student Learning in a Random Sample of College Biology Courses." *Cbe-Life Sciences Education* 10 (4): 394–405. https://doi.org/10.1187/cbe.11-07-0061.

Arum, R, and J Roksa. 2014. *Aspiring Adults Adrift: Tentative Transitions Of College Graduates*. University of Chicago Press.

Ashwin, P. W. H. (2017). Making sense of the Teaching Excellence Framework (TEF) results *Centre for Global Higher Education policy briefings* (Vol. 1). Lancaster.

Bao, L. 2006. "Theoretical Comparisons of Average Normalized Gain Calculations." *American Journal of Physics* 74 (10): 917–22. https://doi.org/10.1119/1.2213632.

Beck, C. W., and L. S. Blumer. 2012. "Inquiry-Based Ecology Laboratory Courses Improve Student Confidence and Scientific Reasoning Skills." *Ecosphere* 3 (12): 1–11. https://doi.org/10.1890/ES12-00280.1.

Beetham, H., & R. Sharpe. 2013. *Rethinking pedagogy for a digital age: Designing for 21st century learning*. London: Routledge.

Buriak, S. E., and J. Potter. 2014. "Impact of an Online Survivorship Primer on Clinician Knowledge and Intended Practice Changes." *Journal of Cancer Education* 29 (1): 114–21. https://doi.org/10.1007/s13187-013-0556-z.

Boud, D. (2018). Assessment could demonstrate learning gains, but what is required for it to do so? *Higher Education Pedagogies, 3*(1), 1-3. doi: 10.1080/23752696.2017.1413671

Cabrera, A. F., C. L. Colbeck, and P. T. Terenzini. 2001. "Developing Performance Indicators for Assessing Classroom Teaching Practices and Student Learning: The Case of Engineering." *Research in Higher Education* 42 (3): 327–52. https://doi.org/10.1023/A:1018874023323.

Cahill, M. J., K. Mairin Hynes, R. Trousil, L. A. Brooks, M. A. McDaniel, M. Repice, J. Zhao, and R. F. Frey. 2014. "Multiyear, Multi-Instructor Evaluation of a Large-Class Interactive-Engagement Curriculum." *Physical Review Special Topics-Physics Education Research* 10 (2): 020101–19. https://doi.org/10.1103/PhysRevSTPER.10.020101.

Campbell, A. M., T. Eckdahl, B. Cronk, C. Andresen, P. Frederick, S. Huckuntod, C. Shinneman, A. Wacker, and J. Yuan. 2014. "PClone: Synthetic Biology Tool Makes Promoter Research Accessible to Beginning Biology Students." *Cbe-Life Sciences Education* 13 (2): 285–96. https://doi.org/10.1187/cbe.13-09-0189.

Casem, M. L. 2006. "Student Perspectives on Curricular Change: Lessons from an Undergraduate Lower-Division Biology Core." *CBE - Life Sciences Education* 5 (1): 65–75. https://doi.org/10.1187/cbe.05-06-0084.

Cheng, K-H, J-C Liang, and C-C Tsai. 2015. "Examining the Role of Feedback Messages in Undergraduate Students' Writing Performance during an Online

Peer Assessment Activity." *Internet and Higher Education* 25 (April): 78–84. https://doi.org/10.1016/j.iheduc.2015.02.001.

Coates, H. 2016. "Assessing Student Learning Outcomes Internationally: Insights and Frontiers." *Assessment & Evaluation in Higher Education* 41 (5): 662–76. https://doi.org/10.1080/02602938.2016.1160273.

Coil, D., M. Wenderoth, M. Cunningham, and C. Dirks. 2010. "Teaching the Process of Science: Faculty Perceptions and an Effective Methodology." *Cbe-Life Sciences Education* 9 (4): 524–35. https://doi.org/10.1187/cbe.10-01-0005.

Cook, T. D., & Campbell, D. T. (1979). The design and conduct of true experiments and quasi-experiments in field settings. In R. T. Mowday & R. M. Steers (Eds.), *Reproduced in part in Research in Organizations: Issues and Controversies*. Santa Monica, CA: Goodyear Publishing Company.

Cronbach, L. J., and L. Furby. 1970. "How We Should Measure 'Change': Or Should We?" *Psychological Bulletin* 74 (1): 68–80. https://doi.org/10.1037/h0029382.

Davies, R. S., D. L. Dean, and N. Ball. 2013. "Flipping the Classroom and Instructional Technology Integration in a College-Level Information Systems Spreadsheet Course." *Etr&d-Educational Technology Research and Development* 61 (4): 563–80. https://doi.org/10.1007/s11423-013-9305-6.

Department for Business, Innovation & Skills. 2016. "Higher Education: Success as a Knowledge Economy - White Paper." 28041602 2809873 05/16. UK. https://www.gov.uk/government/publications/higher-education-success-as-a-knowledge-economy-white-paper.

Dimitrov, D. M., and P. D. Rumrill Jr. 2003. "Pretest-Posttest Designs and Measurement of Change." *Work* 20 (2): 159–165.

Dollar, A., and P. S. Steif. 2008. "Web-Based Statics Course: Patterns in Use and the Relation to Learning Gains." In *Fie: 2008 Ieee Frontiers in Education Conference*, 1–3:1371–76. New York: Ieee.

Douglass, J., G. Thomson, and C-M Zhao. 2012. "The Learning Outcomes Race: The Value of Self-Reported Gains in Large Research Universities." *Higher Education* 64 (3): 317–35. https://doi.org/10.1007/s10734-011-9496-x.

Emke, A., A.C. Butler, and D. P. Larsen. 2016. "Effects of Team-Based Learning on Short-Term and Long-Term Retention of Factual Knowledge." *Medical Teacher* 38 (3): 306–11. https://doi.org/10.3109/0142159X.2015.1034663.

Erdmann, M. A., and J. L. March. 2014. "Video Reports as a Novel Alternate Assessment in the Undergraduate Chemistry Laboratory." *Chemistry Education Research and Practice* 15 (4): 650–57. https://doi.org/10.1039/c4rp00107a.

Everson, K. C. 2017. "Value-Added Modeling and Educational Accountability: Are We Answering the Real Questions?" *Review of Educational Research* 87 (1): 35–70. https://doi.org/10.3102/0034654316637199.

Georgiou, H., and M. D. Sharma. 2015. "Does Using Active Learning in Thermodynamics Lectures Improve Students' Conceptual Understanding and Learning Experiences?" *European Journal of Physics* 36 (1): 015020. https://doi.org/10.1088/0143-0807/36/1/015020.

Getha-Eby, T. J., T. Beery, B. O'Brien, and Y. Xu. 2015. "Student Learning Outcomes in Response to Concept-Based Teaching." *Journal of Nursing Education* 54 (4): 193–200. https://doi.org/10.3928/01484834-20150318-02.

Gill, T., and M. T. Mullarkey. 2015. "Taking a Case Method Capstone Course Online: A Comparative Case Study." *Journal of Information Technology Education-Research* 14: 189–218.

Gok, T. 2012. "The Impact of Peer Instruction on College Students' Beliefs About Physics and Conceptual Understanding of Electricity and Magnetism." *International Journal of Science and Mathematics Education* 10 (2): 417–36. https://doi.org/doi:10.1007/s10763-011-9316-x.

Hake, 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 (1): 64–74. https://doi.org/10.1119/1.18809.

Hatch, D. L., T. Zschau, A. Hays, K. McAllister, M. Harrison, K. L. Cate, R. A. Shanks, and S. A. Lloyd. 2014. "Of Mice and Meth: A New Media- Based Neuropsychopharmacology Lab to Teach Research Methods." *Teaching of Psychology* 41 (2): 167–74. https://doi.org/10.1177/0098628314530352.

Hei, M. de, J-W Strijbos, E. Sjoer, and W. Admiraal. 2016. "Thematic Review of Approaches to Design Group Learning Activities in Higher Education: The Development of a Comprehensive Framework." *Educational Research Review* 18 (May): 33–45. https://doi.org/10.1016/j.edurev.2016.01.001.

Higher Education Funding Council for England. 2015. "Learning Gain." Higher Education Funding Council for England. 2015. http://www.hefce.ac.uk/lt/lg/.

Hill, M., M. D. Sharma, and H. Johnston. 2015. “How Online Learning Modules Can Improve the Representational Fluency and Conceptual Understanding of University Physics Students.” *European Journal of Physics* 36 (4): 1–20. https://doi.org/10.1088/0143-0807/36/4/045019.

Jensen, J. L., T. A. Kummer, and P. D. D. M. Godoy. 2015. “Improvements from a Flipped Classroom May Simply Be the Fruits of Active Learning.” *Cbe-Life Sciences Education* 14 (1): 1–12. https://doi.org/10.1187/10.1187/cbe.14-08-0129.

Jindal-Snape, D., and B. Rienties. 2016. *Multi-Dimensional Transitions of International Students to Higher Education*. New York, NY: Routledge.

Kinchin, I. M. (2016). What do we mean by ‘learning gains’ within biological education? *Journal of Biological Education, 50*(4), 359-360. doi: 10.1080/00219266.2016.1207891

Lim, B., B. Hosack, and P. Vogt. 2012. “A Framework for Measuring Student Learning Gains and Engagement in an Introductory Computing Course: A Preliminary Report of Findings.” *Electronic Journal of E-Learning* 10 (4): 428–40.

Liu, H., J. Liu, and X. Chi. 2014. “Regulatory Mechanism of Self-Determination Involvement in Higher Education: Assessing Chinese Students’ Experiences.” *Higher Education: The International Journal of Higher Education and Educational Planning* 67 (1): 51–70. https://doi.org/10.1007/s10734-013-9640-x.

Mathabathe, K. C., and M. Potgieter. 2014. “Metacognitive Monitoring and Learning Gain in Foundation Chemistry.” *Chemistry Education Research and Practice* 15 (1): 94–104. https://doi.org/10.1039/c3rp00119a.

Matthews, K. E., P. Adams, and M. Goos. 2015. “The Influence of Undergraduate Science Curriculum Reform on Students’ Perceptions of Their Quantitative Skills.” *International Journal of Science Education* 37 (16): 2619–36. https://doi.org/10.1080/09500693.2015.1096427.

Moorer, C. 2009. “Service Learning and Business Education: Distinctions between Undergraduate and Graduate Business Students.” *American Journal of Business Education* 2 (3): 63–72. http://dx.doi.org/10.19030/ajbe.v2i3.4050.

Mortensen, C. J., and A. M. Nicholson. 2015. “The Flipped Classroom Stimulates Greater Learning and Is a Modern 21st Century Approach to Teaching Today’s

Undergraduates.” *Journal of Animal Science* 93 (7): 3722–31. https://doi.org/10.2527/jas2015-9087.

Nagel, J. K. S., O. Pierrakos, A. Zilberberg, and S. McVay. 2012. “Understanding Industry Experiences: From Problem Solving to Engineering Students’ Learning Gains.” In *2012 Frontiers in Education Conference (Fie)*. New York: Ieee.

Ojennus, D. D. 2016. “Assessment of Learning Gains in a Flipped Biochemistry Classroom.” *Biochemistry and Molecular Biology Education* 44 (1): 20–27. https://doi.org/10.1002/bmb.20926.

Ostrom, T. M. 1969. “The Relationship between the Affective, Behavioral, and Cognitive Components of Attitude.” *Journal of Experimental Social Psychology* 5 (1): 12–30. https://doi.org/10.1016/0022-1031(69)90003-1.

Panksepp, J. 2003. “At the Interface of the Affective, Behavioral, and Cognitive Neurosciences: Decoding the Emotional Feelings of the Brain.” *Brain and Cognition*, Affective Neuroscience, 52 (1): 4–14. https://doi.org/10.1016/S0278-2626(03)00003-4.

Pascarella, E. T., and C. Blaich. 2013. “Lessons from the Wabash National Study of Liberal Arts Education.” *Change: The Magazine of Higher Learning* 45 (2): 6–15. https://doi.org/10.1080/00091383.2013.764257.

Pentecost, T. C., and J. Barbera. 2013. “Measuring Learning Gains in Chemical Education: A Comparison of Two Methods.” *Journal of Chemical Education* 90 (7): 839–45. https://doi.org/10.1021/ed400018v.

Pollock, S. J. 2006. “Transferring Transformations: Learning Gains, Student Attitudes, and the Impacts of Multiple Instructors in Large Lecture Courses.” In *2005 Physics Education Research Conference,* edited by P. Heron, L. McCullough, and J. Marx, 818:141–44. Melville: Amer Inst Physics.

Radu, M. E., C. Cole, M. A. Dabacan, J. Harris, and S. Sexton. 2011. “The Impact of Providing Unlimited Access to Programmable Boards in Digital Design Education.” *IEEE Transactions on Education* 54 (2): 174–83. https://doi.org/10.1109/TE.2009.2037735.

Rienties, B., Brouwer, N., and Lygo-Baker, S. 2013. "The effects of online professional development on teachers' beliefs and intentions towards learning facilitation and technology". *Teaching and Teacher Education*, 29, 122-131. DOI: 10.1016/j.tate.2012.09.002

Riskowski, J. L. 2015. “Teaching Undergraduate Biomechanics with Just-in-Time Teaching.” *Sports Biomechanics* 14 (2): 168–79. https://doi.org/10.1080/14763141.2015.1030686.

Rocconi, L. M., and C. A. Ethington. 2009. “Assessing Longitudinal Change: Adjustment for Regression to the Mean Effects.” *Research in Higher Education* 50 (4): 368–76. https://doi.org/10.1007/s11162-009-9119-x.

Rogaten, J., B. Rienties, and D. Whitelock. 2017. "Assessing Learning Gains". In D. Joosten-ten Brinke & M. Laanpere (Eds.), *Technology Enhanced Assessment. TEA 2016. Communications in Computer and Information Science* (Vol. 653, pp. 117-132). Cham: Springer.

Roohr, K. C., H. Liu, and O. L. Liu. 2017. “Investigating Student Learning Gains in College: A Longitudinal Study.” *Studies in Higher Education* 42 (12): 2284–2300. https://doi.org/10.1080/03075079.2016.1143925.

Seymour, E., S. M. Daffinrud, D. J. Wiese, and A. B. G. Hunter. 2000. “Development, Testing, and Adaptation for Wider Faculty Use of the on Line Student Assessment of Their Learning Gains (SALG) Instrument, Originally Designed for Modular Chemistry Teachers.” *Abstracts of Papers of the American Chemical Society* 219 (March): U358–U358.

Shuster, M., and K. Peterson. 2009. “Development, Implementation, and Assessment of a Lecture Course on Cancer for Undergraduates.” *Cbe-Life Sciences Education* 8 (3): 193–202. https://doi.org/10.1187/cbe.09-03-0014.

Stolk, J. D., and R. Martello. 2015. “Can Disciplinary Integration Promote Students’ Lifelong Learning Attitudes and Skills in Project-Based Engineering Courses?” *International Journal of Engineering Education* 31 (1): 434–49.

Strayhorn, T. L. 2010. “Work in Progress - Diversifying STEM: Underrepresented Students’ Experiences in Undergraduate Research Programs.” In *2010 Ieee Frontiers in Education Conference (Fie)*. New York: Ieee.

Strijbos, J, N. Engels, and K. Struyven. 2015. “Criteria and Standards of Generic Competences at Bachelor Degree Level: A Review Study.” *Educational Research Review* 14: 18–32. https://doi.org/10.1016/j.edurev.2015.01.001.

Tomasik, J. H., K. E. Cottone, M. T. Heethuis, and A. Mueller. 2013. “Development and Preliminary Impacts of the Implementation of an Authentic Research-Based

Experiment in General Chemistry." *Journal of Chemical Education* 90 (9): 1155–61. https://doi.org/10.1021/ed300328p.

Torgerson, D. J., & Torgerson, C. (2008). *Designing randomised trials in health, education and the social sciences: an introduction*. London: Palgrave Macmillan.

U.S. Department of Education. 2006. "A Test of Leadership: Charting the Future of U.S. Higher Education." ED-06-C0-0013. Washington D.C. http://edf.stanford.edu/readings/test-leadership-charting-future-higher-education-report-commission-appointed-secretary.

Varsavsky, C., K. E. Matthews, and Y. Hodgson. 2014. "Perceptions of Science Graduating Students on Their Learning Gains." *International Journal of Science Education* 36 (6): 929–51. https://doi.org/10.1080/09500693.2013.830795.

Villafane, S. M., J. Loertscher, V. Minderhout, and J. E. Lewis. 2011. "Uncovering Students' Incorrect Ideas about Foundational Concepts for Biochemistry." *Chemistry Education Research and Practice* 12 (2): 210–18. https://doi.org/10.1039/C1RP90026A.

Wabash National Study of Liberal Arts Education. 2012. 2012. http://www.liberalarts.wabash.edu.

Wattiaux, M. A., and P. Crump. 2006. "Students' Perception of a Discussion-Driven Classroom Environment in an Upper-Level Ruminant Nutrition Course with Small Enrollment." *Journal of Dairy Science* 89 (1): 343–52. https://doi.org/10.3168/jds.S0022-0302(06)72100-2.

Willoughby, S. D., and A. Metz. 2009. "Exploring Gender Differences with Different Gain Calculations in Astronomy and Biology." *American Journal of Physics* 77 (7): 651–57. https://doi.org/10.1119/1.3133087.

Woltering, V., A. Herrler, K. Spitzer, and C. Spreckelsen. 2009. "Blended Learning Positively Affects Students' Satisfaction and the Role of the Tutor in the Problem-Based Learning Process: Results of a Mixed-Method Evaluation." *Advances in Health Sciences Education* 14 (5): 725–38. https://doi.org/10.1007/s10459-009-9154-6.

Woodall, T., A. Hiller, and S. Resnick. 2014. "Making Sense of Higher Education: Students as Consumers and the Value of the University Experience." *Studies in Higher Education* 39 (1): 48–67. https://doi.org/10.1080/03075079.2011.648373.

Yalaki, Y. 2010. “Simple Formative Assessment, High Learning Gains in College General Chemistry.” *Egitim Arastirmalari-Eurasian Journal of Educational Research* 10 (40): 223–43.

Table 1: *Total number of studies and sample size (in parenthesis) for each method employed to capture affective, behavioural and cognitive learning gains.*

	Affective	Behaviour	Cognitive
Pre-post objective	-	-	32 (22,004) (e.g., Andrews et al., 2011; Emke et al., 2016; Georgiou & Sharma, 2015)
Pre-post subjective	9 (1,561) (e.g., Beck & Blumer, 2012; Cheng, Liang, & Tsai, 2015; Mortensen & Nicholson, 2015)	1 (114) (Stolk and Martello 2015)	6 (12,942) (e.g., Hatch et al., 2014; Lim, Hosack, & Vogt, 2012; Stolk & Martello, 2015)
Cross-sectional subjective	10 (1,772) (e.g., Gok, 2012; Liu et al., 2014; Moorer, 2009)	12 (4,154) (e.g., Casem, 2006; Gill & Mullarkey, 2015; Gok, 2012)	16 (5,082) (Casem 2006; Douglass, Thomson, and Zhao 2012; Gok 2012)

Note: 47% off all studies assessed more than one type of learning gains and as such, one sample can fall into more than one category and number of samples in the table do not strictly add up to the total number of samples examined in this review.

Table 2: Studies coded by method employed, number of subsamples associated with each study and type of learning gain: affective, behavioural and cognitive.

Reference	*Number of groups*	*Design*	*Measure*	*A*	*B*	*C*
Anderson (2006)	4	Cross-sectional	Self-reported	No	No	Yes
Cabrera, Colbeck & Terenzini (2001)	1	Cross-sectional	Self-reported	No	Yes	Yes
Cahill et al. (2014)	2	Pre-post test	Self-reported	Yes	No	Yes
Casem (2006)	2	Cross-sectional	Self-reported	No	No	Yes
Casem (2006)	6	Cross-sectional	Self-reported	No	Yes	Yes
Douglass, Thomson & Zhao (2012)	1	Cross-sectional	Self-reported	No	No	Yes
Gill & Mullarkey (2015)	2	Cross-sectional	Self-reported	No	Yes	Yes
Gok (2012)	2	Cross-sectional	Self-reported	Yes	Yes	Yes
Hatch et al. (2014)	2	Pre-post test	Self-reported	No	No	Yes
Hodgson, Varsavsky & Matthews (2014)	1	Cross-sectional	Self-reported	No	Yes	Yes

Lim, Hosack & Vogt (2012)	1	cross-sectional	Self-reported	No	Yes	No
Lim, Hosack & Vogt (2012)	1	Pre-post test	Self-reported	Yes	No	Yes
Liu, Liu & Chi (2014)	1	Cross-sectional	Self-reported	Yes	Yes	Yes
Matthews, Adams & Goos (2015)	2	Cross-sectional	Self-reported	No	No	Yes
Moorer (2009)	2	Cross-sectional	Self-reported	Yes	Yes	Yes
Nagel, Pierrakos, Zilberberg & McVay (2012)	1	Cross-sectional	Self-reported	Yes	No	Yes
Ojennus (2016)	2	Cross-sectional	Self-reported	Yes	Yes	Yes
Radu, Cole, Dabacan, Harris & Sexton (2011)	1	Cross-sectional	Self-reported	Yes	No	Yes
Seymour, Daffinrud, Wiese & Hunter (2000)	1	Cross-sectional	Self-reported	Yes	Yes	Yes
Stolk & Martello (2015)	2	Pre-post test	Self-reported	Yes	Yes	Yes
Strayhorn (2010)	1	Cross-sectional	Self-reported	Yes	Yes	Yes
Tomasik, Cottone, Heethuis & Mueller (2013)	2	Cross-sectional	Self-reported	Yes	No	Yes

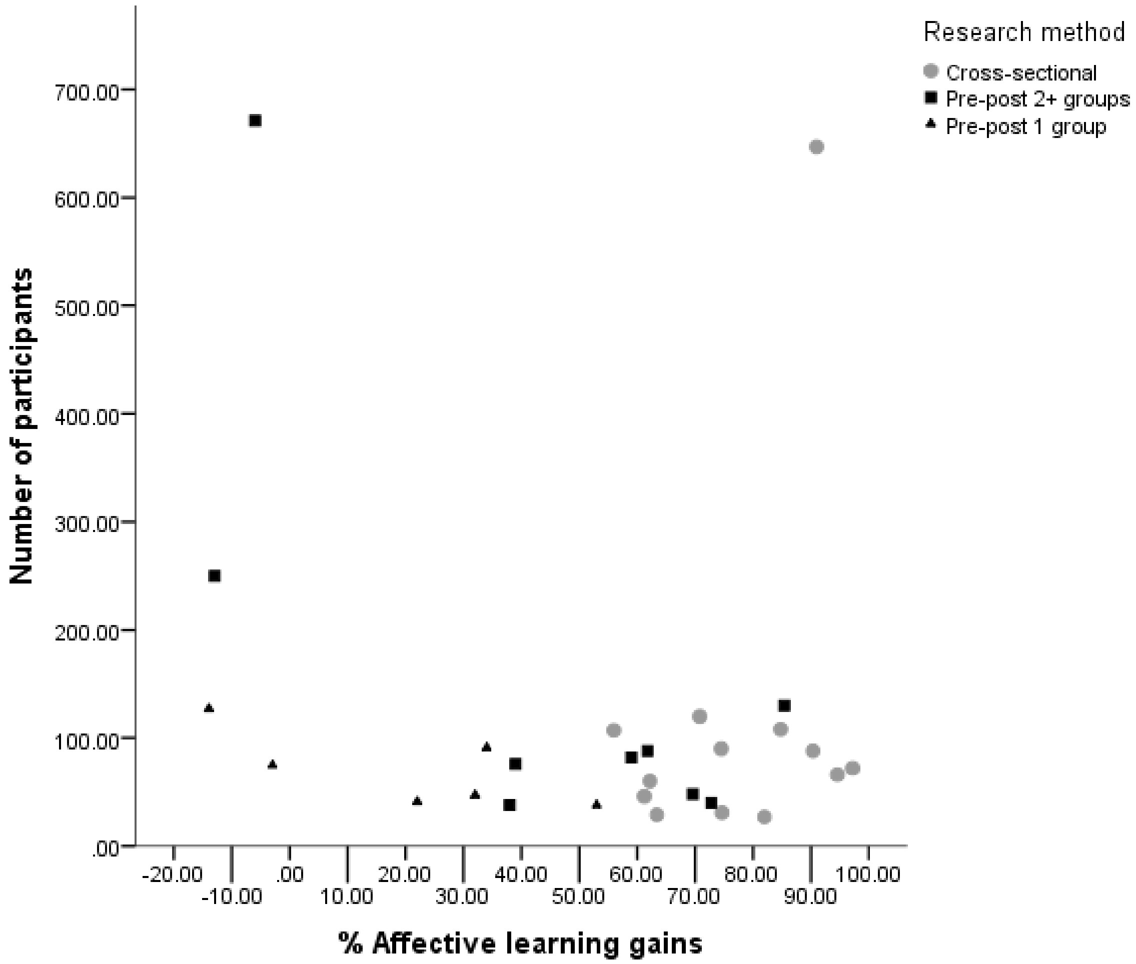
Varsavsky, Matthews & Hodgson (2014)	1	Cross-sectional	Self-reported	No	Yes	Yes
Wattiaux & Crump (2006)	1	Pre-post test	Self-reported	Yes	No	Yes
Woltering, Herrler, Spitzer & Spreckelsen (2009)	2	Cross-sectional	Self-reported	No	No	Yes
Allen & Cockman (2009)	2	Pre-post test	Objective	No	No	Yes
Andrews, Leonard, Colgrove & Kalinowski (2011)	1	Pre-post test	Objective	No	No	Yes
Boyas, Bryan & Lee (2012)	2	Pre-post test	Objective	No	No	Yes
Buriak & Potter (2014)	1	Pre-post test	Objective	No	No	Yes
Cahill et al. (2014)	4	Pre-post test	Objective	No	No	Yes
Campbell et al. (2014)	2	Pre-post test	Objective	No	No	Yes
Davies, Dean & Ball (2013)	3	Pre-post test	Objective	No	No	Yes
Dollar & Steif (2008)Dollar & Steif (2008)	1	Pre-post test	Objective	No	No	Yes
Emke, Butler & Larsen (2016)	2	Pre-post test	Objective	No	No	Yes
Erdmann & March (2014)	2	Pre-post test	Objective	No	No	Yes
Georgiou & Sharma (2015)	2	Pre-post test	Objective	No	No	Yes
Getha-Eby, Beery, O’Brien & Xu (2015)	2	Pre-post test	Objective	No	No	Yes
Gok (2012)	2	Pre-post test	Objective	No	No	Yes

Hill, Sharma & Johnston (2015)	3	Pre-post test	Objective	No	No	Yes
Jensen, Kummer & Godoy (2015)	2	Pre-post test	Objective	No	No	Yes
Margoniner (2014)	2	Pre-post test	Objective	No	No	Yes
Mortensen & Nicholson (2015)	1	Pre-post test	Objective	No	No	Yes
Ojennus (2016)	2	Pre-post test	Objective	No	No	Yes
Pentecost & Barbera (2013)	4	Pre-post test	Objective	No	No	Yes
Pollock (2006)	3	Pre-post test	Objective	No	No	Yes
Riskowski (2015)	1	Pre-post test	Objective	No	No	Yes
Roohr et al. (2017)	3	Pre-post test	Objective	No	No	Yes
Shi, Wood, Martin, Guild, Vicens & Knight (2010)	2	Pre-post test	Objective	No	No	Yes
Shuster & Peterson (2009)	1	Pre-post test	Objective	No	No	Yes
Villafane, Loertscher, Minderhout & Lewis (2011)	2	Pre-post test	Objective	No	No	Yes
Willoughby & Metz (2009)	4	Pre-post test	Objective	No	No	Yes
Wolkow, Durrenberger, Maynard, Harrall & Hines (2014)	4	Pre-post test	Objective	No	No	Yes
Yalaki (2010)	2	Pre-post test	Objective	No	No	Yes
Beck & Blumer (2012)	1	Pre-post test	Mixed	Yes	No	Yes
Cheng, Liang & Tsai (2015)	1	Pre-post test	Mixed	Yes	No	Yes

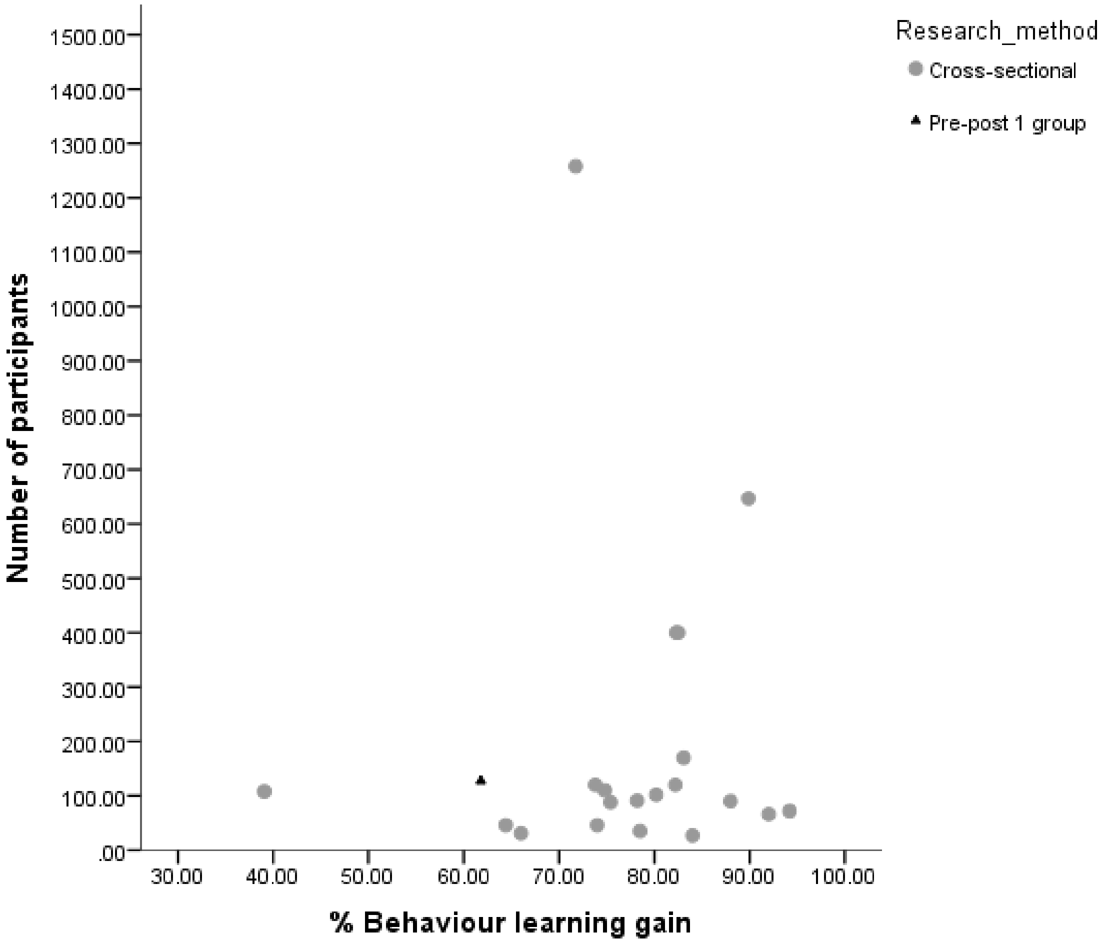
Mathabathe & Potgieter (2014)	1	Pre-post test	Mixed	Yes	No	Yes
Mortensen & Nicholson (2015)	1	Pre-post test	Mixed	Yes	No	Yes
O’Shea, Terry & Benenson (2013)	1	Pre-post test	Mixed	Yes	No	Yes
Wolkow, Durrenberger, Maynard, Harrall & Hines (2014)	4	Pre-post test	Mixed	Yes	No	Yes

Figure 1: Scatterplot of the relationship between study sample sizes and self-reported affective, behaviour and cognitive learning gains.

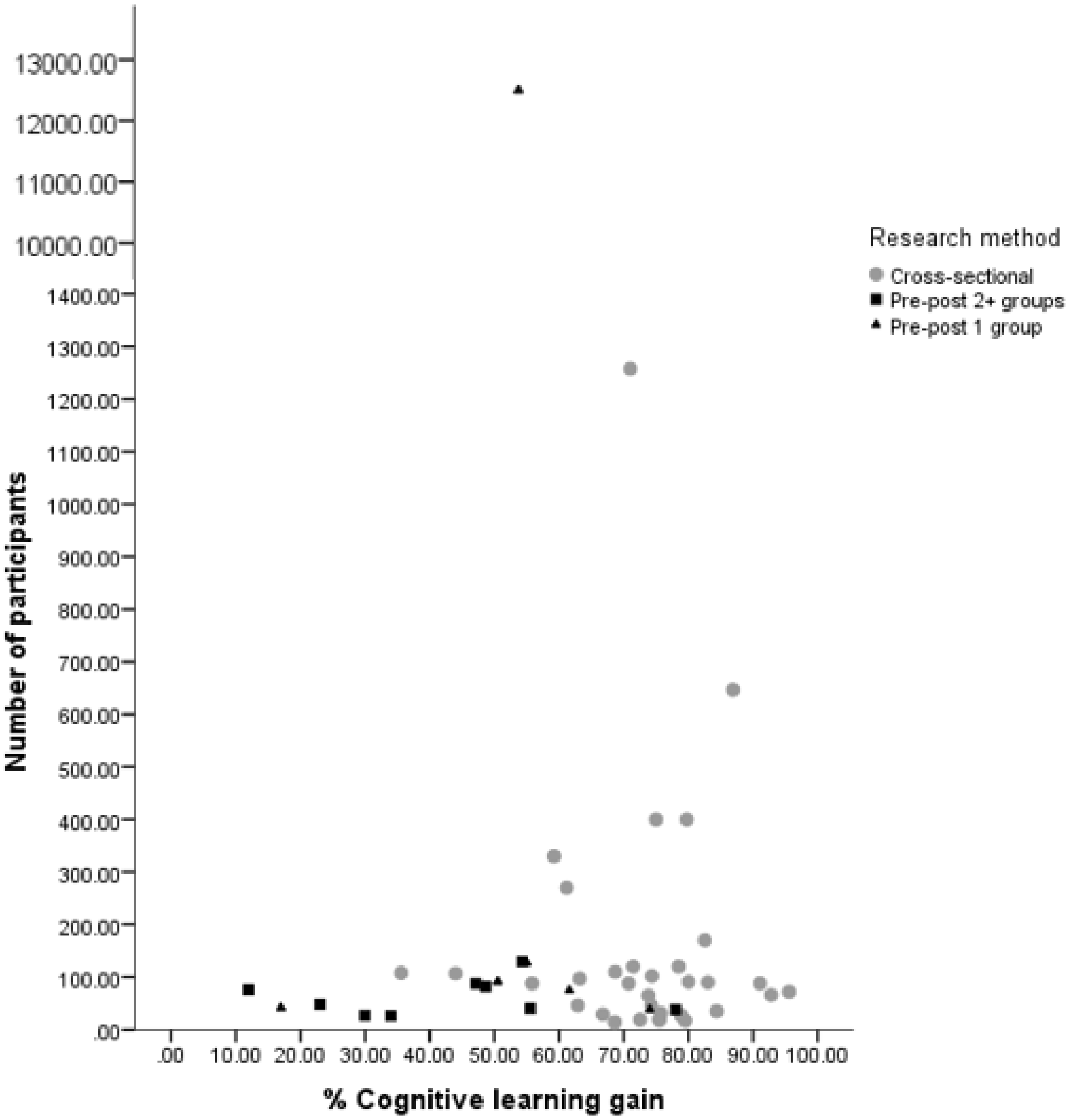
a) *Affective learning gains*



b) Behavioural learning gains



c) Cognitive learning gains



Outcome of studies using objective testing

Figure 2: *Scatterplot of the relationship between study sample sizes and normalised learning gains estimated from objective testing for different study conditions.*