

University of Wollongong

Research Online

Faculty of Engineering and Information
Sciences - Papers: Part A

Faculty of Engineering and Information
Sciences

1-1-2017

Maximising resource allocation in the teaching laboratory: understanding student evaluations of teaching assistants in a team-based teaching format

Sasha Nikolic

University of Wollongong, sasha@uow.edu.au

Thomas F. Suesse

University of Wollongong, tsuesse@uow.edu.au

Timothy J. McCarthy

University of Wollongong, timmc@uow.edu.au

Thomas Goldfinch

University of Wollongong, tomgold@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/eispapers>



Part of the [Engineering Commons](#), and the [Science and Technology Studies Commons](#)

Recommended Citation

Nikolic, Sasha; Suesse, Thomas F.; McCarthy, Timothy J.; and Goldfinch, Thomas, "Maximising resource allocation in the teaching laboratory: understanding student evaluations of teaching assistants in a team-based teaching format" (2017). *Faculty of Engineering and Information Sciences - Papers: Part A*. 6531. <https://ro.uow.edu.au/eispapers/6531>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Maximising resource allocation in the teaching laboratory: understanding student evaluations of teaching assistants in a team-based teaching format

Abstract

Minimal research papers have investigated the use of student evaluations on the laboratory, a learning medium usually run by teaching assistants with little control of the content, delivery and equipment. Finding the right mix of teaching assistants for the laboratory can be an onerous task due to the many skills required including theoretical and practical know-how, troubleshooting, safety and class management. Using larger classes with multiple teaching assistants, a team-based teaching (TBT) format may be advantageous. A rigorous three-year study across twenty-five courses over repetitive laboratory classes is analysed using a multi-level statistical model considering students, laboratory classes and courses. The study is used to investigate the effectiveness of the TBT format, and quantify the influence each demonstrator has on the laboratory experience. The study found that TBT is effective and the lead demonstrator most influential, influencing up to 55% of the laboratory experience evaluation.

Disciplines

Engineering | Science and Technology Studies

Publication Details

S. Nikolic, T. F. Suesse, T. J. McCarthy & T. L. Goldfinch, "Maximising resource allocation in the teaching laboratory: understanding student evaluations of teaching assistants in a team-based teaching format," *European Journal of Engineering Education*, vol. 42, (6) pp. 1277-1295, 2017.

**Maximising Resource Allocation in the Teaching Laboratory:
Understanding Student Evaluations of Teaching Assistants in a Team
Based Teaching Format**

Sasha Nikolic (lead author)

Engineering and Information Sciences, University of Wollongong, Australia

Northfields Ave, University of Wollongong, NSW, Australia. +61242213418

sasha@uow.edu.au

Thomas Suesse

Engineering and Information Sciences, University of Wollongong, Australia

Northfields Ave, University of Wollongong, NSW, Australia. +61242214173

tsuesse@uow.edu.au

Timothy McCarthy

*Engineering and Information Sciences, University of Wollongong, Wollongong,
Australia*

Northfields Ave, University of Wollongong, NSW, Australia. +61242214591

timmc@uow.edu.au

Thomas Lachlan Goldfinch

*Engineering and Information Sciences, University of Wollongong, Wollongong,
Australia*

Northfields Ave, University of Wollongong, NSW, Australia. +61242213138

tomgold@uow.edu.au

Maximising Resource Allocation in the Teaching Laboratory: Understanding Student Evaluations of Teaching Assistants in a Team Based Teaching Format

Minimal research papers have investigated the use of student evaluations on the laboratory, a learning medium usually run by teaching assistants with little control of the content, delivery and equipment. Finding the right mix of teaching assistants for the laboratory can be an onerous task due to the many skills required including theoretical and practical know how, troubleshooting, safety and class management. Using larger classes with multiple teaching assistants, a team based teaching (TBT) format may be advantageous. A rigorous three year study across twenty-five courses over repetitive laboratory classes is analysed using a multi-level statistical model considering students, laboratory classes and courses. The study is used to investigate the effectiveness of the TBT format, and quantify the influence each demonstrator has on the laboratory experience. The study found that TBT is effective and the lead demonstrator most influential, influencing up to fifty-five percent of the laboratory experience evaluation.

Keywords: demonstrator; laboratory; student evaluation; teaching assistants; team based teaching

1. Introduction

In engineering and science, teaching laboratories play an important role in developing the skills and knowledge of students, as identified in numerous studies across many STEM disciplines (Khan, Al-Doussari, and Al-Kahtani 2002; Feisel and Rosa 2005; Deacon and Hajek 2010; O'Toole et al. 2012; Nikolic, Ritz, et al. 2015; Rathod and Kalbande 2016). Learning in the laboratory can be conducted by undertaking experiments on campus, remotely or via computer simulations (Abdulwahed and Nagy 2009; Koretsky, Kelly, and Gummer 2011; Lowe et al. 2015). Student development in the laboratory includes the development of conceptual thinking, enabling scientific

discovery, and inquiry based learning across the cognitive, psychomotor, and affective domains (O'Toole et al. 2012; Salim et al. 2013).

While the role of the laboratory is acknowledged as important in facilitating learning, merely having a laboratory does not in any way ensure that productive learning is taking place, that it is enjoyable, or a quality experience (Casas and del Hoyo 2009; Stamatelos and Stamatelos 2009). Laboratories can be expensive to run and maintain due to the cost of equipment, licencing software, and teaching load if run under the typical model of small student to staff ratios. Sometimes small student to staff ratios are necessary due to the limited availability of equipment. Due to the high cost of running a laboratory and the central role that it plays in psychomotor development (most lecture and tutorial arrangements focus on cognitive and affective skills), quality management systems should be used to ensure productive learning is taking place. In many cases quality is measured by receiving feedback from student questionnaires that investigate experiences and opinions (Ugursal and Cruickshank 2015; Nikolic, Ritz, et al. 2015).

One form of quality management universities use, are student evaluations of teaching (SETs). The SET is a widely used tool to evaluate the quality of instruction. While the original intention of SETs was to provide feedback to improve teaching, the tool is now commonly used to help make decisions on retention, tenure, and promotion (Walker and Palmer 2009). For example, for promotion at the authors' university four SET surveys or peer reviews on teaching must be presented as evidence of good teaching. However, like every other tool, the SET must be carefully understood to be used effectively. The controversial issues in regard to the validity, benefits, and biases of SETs has resulted in several thousand research studies (Spooren, Brockx, and Mortelmans 2013). An analysis of 154 research articles between 1924 and 1998 by

Aleamoni (1999) found that the use of SETs can be extremely beneficial for students, staff and institutions when the SET is well-designed and analysed correctly. On the other hand, the study found that the SET can be easily misinterpreted and misused, undermining their credibility.

The great volume of SET research indicates the complexity and limitations of such studies. The reason for this is that there are so many variables that need to be considered (such as class size, bias, appearance, time of day, weather and gender) and the correlation to learning can always be easily questioned. Most SET research is focussed on the evaluation of the main lecturer of the subject. For example, the laboratory in recent times is predominantly taught by teaching assistants that do not create the teaching content but simply help deliver it, and more information is needed on how they influence a SET (O'Toole et al. 2012). In addition, teaching in the laboratory is significantly more complex than teaching in a lecture or tutorial because the teaching assistant not only needs to know the theory but also needs to know about the hardware, software, troubleshooting, safety, and more. Research has also suggested that the quality of experiment design and equipment can significantly influence evaluations (Nikolic, Ritz, et al. 2015). While it is commonly assumed that the teacher is highly influential in student evaluations, more is needed to understand the relationship between teaching ability and the role the laboratory experiment and facilities play in the student experience. For example, how much influence does a great teacher have in making a poorly designed and equipped laboratory become a great learning experience?

To complicate matters further, a laboratory may be run by more than one teaching assistant. The question then needs to be raised as to what influence the different teaching assistants play in the student experience of the laboratory. Having

multiple teachers in the classroom is known as team based teaching (TBT). Advantages of TBT include; greater support in the classroom due to additional teachers; students get to interact with a variety of personalities and teaching styles; and, students can seek assistance from the teacher they find most effective (Buckley 1999). It is also used to keep a constant student to teacher ratio in the classroom. The use of TBT has been found successful in a number of teaching areas including languages (Carless and Walker 2006), marketing (Yanamandram and Noble 2006), and engineering (Nikolic, Vial, et al. 2015). To increase knowledge in this area it is important to understand the most effective way of allocating teaching assistants in a TBT format. For example, is having two highly experienced teachers any different from having one experienced and one in training? The cost of using two highly experienced teachers could be more expensive so it is important to know the cost benefit. In this study TBT is used in the context of running a laboratory class with multiple teaching assistants. The course coordinator supports the teaching assistants but spends minimal time in the laboratory.

2. Purpose of Study

The purpose of this study is to gain a greater understanding of how student evaluations can be used to improve learning and the student experience within the teaching laboratory. This knowledge will make learning and allocating resources to the laboratory more efficient and targeted to learning objectives. This work expands upon a number of related studies by the author. The first study (Nikolic, Vial, et al. 2015) investigated how training of the sessional/casual teaching assistants (called laboratory demonstrators) correlated to increases in teacher evaluation scores. The study found that evaluation scores increased by improving communication, classroom management and teaching technique. A few other studies (Nikolic, Ritz, et al. 2015; Nikolic 2015; Vial et al. 2015; Nikolic 2014) explored how improvements to the laboratory experience,

defined by the experiments they undertook and the facilities they used, were correlated to the student evaluation scores. These studies found an intertwined connection between activity, clarity, resources and facilities. The last study (Nikolic, Suesse, et al. 2015) discovered that the evaluation scores can be linked to the student's perceived learning across the cognitive and psychomotor domains. This was also linked to student's cognitive performance in a laboratory exam, but shortcomings in the assessment process limited the evaluation of the psychomotor skills.

Unclear from the previous studies is an understanding on how improvements from the teaching evaluations improved the laboratory experience evaluations or vice versa. In particular, when using a TBT laboratory format little information is known as to how influential each of the demonstrators are to the evaluations. This understanding would lead to greater efficiency in resourcing. The study will commence by examining common biasing factors such as: class format (does TBT help alleviate any negativity of large class sizes?); course level (do evaluations improve with each course level, especially when electives are chosen?); and, gender (the influence of male and female teachers). These factors will be compared against student evaluations of the laboratory demonstrators, student evaluations of the laboratory experience (the experiment and facilities), and the number of demonstrators used. A relationship will then be determined to understand how these three factors influence each other, leading to improved resource management.

This study commences by providing context, followed by investigating the literature behind the various biasing factors. This literature is used to define the relevant hypotheses. The research design is explained, followed by the results, discussion and conclusion.

3. Background Information

This study is conducted within the School of Electrical, Computer and Telecommunications Engineering (SECTE) at the University of Wollongong, Australia. The students undertook undergraduate or postgraduate coursework study. The university uses two official evaluation instruments. The first is a Teacher Evaluation and is generally targeted at teaching staff involved in lecturing, usually the course coordinator. This survey is generally not suited to laboratory demonstrators due to the fact that they have no major say in how the laboratory is run or delivered. The second instrument is a Subject Evaluation Survey that takes a holistic view of the entire course. Due to the importance the school placed on laboratory learning, laboratory tailored evaluations were deemed necessary. As a result, the two official university survey instruments are only used to measure course quality in its entirety and not used to specifically measure quality in the teaching laboratories. This led to the formation of laboratory specific evaluations (Nikolic, Ritz, et al. 2015; Nikolic, Vial, et al. 2015). It is only the laboratory specific evaluation instruments that are the focus of this study.

The survey instruments were developed by the SECTE School Committee and commenced officially in 2009. The developed evaluation questions were not guided by research, rather through metrics that the committee valued as important to both students and the school. While the construction of the survey could have been improved, a number of tests for validity have been performed. This includes testing the face validity, for example the demonstrator questions aligning with characteristics that students value the most from teachers being knowledge, communication and competence (Pozo-Munoz, Reboloso-Pacheco, and Fernandez-Ramirez 2000). Convergent validity was confirmed in Nikolic, Suesse, et al. (2015) with the relationship to student achievement and perception examined. In section 5.1 structural validity is tested to confirm if the

evaluation questions can be grouped into scores. Discriminant validity is tested within the research questions of this study to determine any possible effect of bias.

The student evaluation of teaching performance of the laboratory demonstrators is based on five survey questions on a Likert scale that are calculated into a weighted average score. The five questions are:

- Question 1: At the start of each laboratory does the casual demonstrator give you a satisfactory introduction to the laboratory?
- Question 2: Is the casual demonstrator well prepared for the subject?
- Question 3: Does the casual demonstrator communicate the subject matter clearly?
- Question 4: Did the casual demonstrator appear interested in helping me to learn?
- Question 5: Is the casual demonstrator helpful in responding to questions or problems?

Similarly, three statements are used to measure student perception of the laboratory experiment (S1-3), and the equipment and facilities (S4-6). In this study statements one to three are grouped into a score referred as LAB1 and statements four to six as LAB2. The statements are:

- Statement 1: I have a great overall impression of the laboratory component for this course.
- Statement 2: The contents of the laboratory notes provided me with enough information to successfully complete the required exercises.
- Statement 3: The experiments undertaken in this laboratory are worthwhile learning experiences.
- Statement 4: The computers in the laboratory are suitable for the work required.
- Statement 5: The electronic equipment in the lab, other than the computers, is suitable for the work required.
- Statement 6: The laboratory is in a good condition

The responsibility of running the laboratory for any given course is directed at the course coordinator. They are responsible for the overall learning conducted within the laboratory and supporting the teaching assistants. Most of the teaching in the laboratory is conducted by one or more teaching assistants' dependant on the size of the class. As a result the focus of this study is only on the laboratory demonstrators. The term TBT in this study is used to describe the use of one, two or three demonstrators within the laboratory. Further detail is provided in section five, the research design section.

4. Biasing Factors and Hypothesis

In an ideal world, it would be expected that the SET would be free of bias and be a reflection on student learning. The thousands of SET research studies suggest otherwise. Possible biases that can be found to impact a SET when used in the laboratory include class format, course level, and gender of the teacher.

4.1 Class Format

The class format refers to the number of demonstrators in the laboratory. It is used to compare the difference of having a small laboratory class with only one laboratory demonstrator, or a larger class with multiple demonstrators. While in many instances the class size will be dependent on the equipment, in some situations such as in a computer laboratory large classes are achievable and generally more efficient, especially in terms of timetabling. The assignment of these parameters are out of the control of the laboratory demonstrator. Therefore, it is important to understand the role class size plays, and the impact TBT has on the student evaluation. The class size is defined as the number of students participating in a single laboratory session.

Literature studying the effect class size has on SETs, is skewed towards showing some form of impact (Shapiro 1990; Watkins 1990; El Ansari and Oskrochi 2006; Johnson, Narayanan, and Sawaya 2013). A small number of studies have shown that class size has little impact on SET (Lin 1992; Zabaleta 2007). However, a number of studies involving large datasets found class size has a significant negative correlation with SET scores. This includes a study by Narayanan, Sawaya, and Johnson (2014) covering 983 business and engineering courses, Johnson, Narayanan, and Sawaya (2013) covering 3938 courses and 549 unique engineering instructors, and Watkins (1990) with 20,000 ratings from over 200 courses. In terms of student learning Bandiera, Larcinese, and Rasul (2010) found that on average class size did not have much of an effect, except however, for smallest and largest classes. A large negative effect was found with classes from 1-19 students compared to 20-33 students, the most commonly sized laboratory classes conducted in this study. The relationship between the SET and student learning was also investigated by (Galbraith, Merrill, and Kline 2012). Investigating the relationship between 116 business related courses, the study found that the larger the class, the less probable that the SET could predict achievement in student learning.

In this study, the student to staff ratio was designed to be constant with approximately one demonstrator for every fifteen students. When team teaching was required, an experienced demonstrator would be partnered up with a less experienced demonstrator. Considering the constant student to teacher ratio, irrespective of the class size, and that an experienced and inexperienced demonstrator would work together the following hypotheses are tested:

H1a: Class size does not have an effect on demonstrator teaching scores in a team based teaching format

H1b: Class size does not have an effect on laboratory experience scores in a team based teaching format

4.2 Course Level

When students start an undergraduate degree in Australia they have little understanding about the learning experience at university. With age and experience, it is often observed that the students (excluding mature age entry students) mature throughout the degree. That is, the students do not fully understand what constitutes good teaching until late into their degree, or years after completion. Secondly, the first few years of the degree are heavily prescribed core courses that are general across many disciplines. In this study, some core courses to specific disciplines commence in the third year, with the fourth year comprising of many discipline specific electives. As a result of this, student motivation may increase with course level. This has led to numerous studies that have investigated if course level has any effect on SET. The complexity of such analysis was highlighted by Cranton and Smith (1986) who investigated five different departments, and found that the relationship differed across departments.

Studies outside engineering, such as psychology courses (Blackhart et al. 2006) and business courses (Scherr and Scherr 1990) have shown that course level can have little effect on SET. Major engineering based studies (Johnson, Narayanan, and Sawaya 2013; Narayanan, Sawaya, and Johnson 2014), indicate the opposite. In particular, the study by Narayanan, Sawaya, and Johnson (2014) comparing engineering and business, found that the effect on course level was greatest for engineering. A sample of 3,185 business, economics, accounting, and statistics students by Badri et al. (2006) also

indicated that a relationship between course level and SET existed. In addition, it is important to see if and how TBT would influence the SET. Considering the literature in terms of engineering the following hypotheses are tested:

H2a: Higher level courses receive higher demonstrator teaching scores in the laboratory, regardless of team based teaching format

H2b: Higher level courses receive higher laboratory experience scores, regardless of team based teaching format

4.3 Gender

The gender of the instructor and/or student, and the relationship with SET has also been widely investigated. The findings from the research provide very mixed conclusions. A recent investigation by (MacNell, Driscoll, and Hunt 2014) into student ratings of online teaching staff, where the real gender of the staff was not known, found that the SET was biased against females. This was determined by using a male and female instructor participating in multiple online discussion forums, using both a male or female name in different sessions. However, due to the small sample (N=72) and design of the experiment, researchers have questioned the validity of the findings (Benton and Li 2014). Other gender bias studies have also found; that males give significantly lower SET scores to females (Basow and Silberg 1987; Centra and Gaubatz 2000); male instructors receive higher SET scores regardless of the gender of the student (Basow and Silberg 1987; McPherson, Jewell, and Kim 2009); and, male instructors received equal ratings from males and females (Centra and Gaubatz 2000).

A wide ranging study of literature by Aleamoni (1999) found that the majority of studies have established that gender does not play a biasing role in SETs. Some of the studies that have found such a conclusion include: a study of four universities in

Pakistan (Hameed et al. 2014), a controlled social experiment (Feldman 1993), and a large scale study of a Spanish program (Zabaleta 2007).

The large scale engineering studies found that; male instructors do receive higher SET ratings, and that female instructors were not being held to a higher standard (Johnson, Narayanan, and Sawaya 2013); and, that for engineering, male instructors received higher ratings, but this is not the case in the business college (Narayanan, Sawaya, and Johnson 2014). This study investigates this bias further by comparing male and female instructors in a mixed TBT format. Considering the mixed literature and engineering studies, the following hypotheses are tested:

H3a: Male instructors receive higher demonstrator teaching scores, regardless of team based teaching format

H3b: Laboratories with male instructors receive higher laboratory experience scores, regardless of team based teaching format

4.4 Relationship: Teaching vs Laboratory Experience

The final hypothesis and main goal of this study is to investigate the relationship between student evaluations of teaching and the laboratory experience (quality of the experiments and facilities). As evidenced in the literature reviewed earlier, substantial research has been undertaken to investigate the SET. Separately, research has been conducted to investigate student's perception of the quality of university facilities and laboratories (Douglas, Douglas, and Barnes 2006; Deshwal, Mahajan, and Choudhary 2012; Gonsai et al. 2013; Stanisavljevic et al. 2013; Nikolic, Ritz, et al. 2015). What needs more attention, is understanding how student perception of teaching quality relates to their perception of the quality of the laboratory experience. This is important to understand because the *"quality of the classroom life is significant in shaping*

students' feelings and attitudes to their classmates and teachers" (pg. 368) (Che Ahmad, Osman, and Halim 2013). By being able to quantify this relationship universities will be able to more effectively decide how they allocate resources to maximise the student experience, leading to efficiency and monetary gains. Therefore, the following hypothesis is tested, and used to quantify the relationship amongst the teaching team.

H4a: There is a positive relationship between demonstrator teaching scores and laboratory experience scores

4.5 Bias Limitations

This study does not attempt to investigate all forms of bias that may occur in a SET. Other forms of biases that have been investigated in literature include bias against minority and non-native English speakers (Plank and Chiagouris 1997; Hamermesh and Parker 2005; Reid 2010; Bavishi, Madera, and Hebl 2010); age discrimination (Stolte 1996; Arbuckle and Williams 2003); physical attractiveness (Langlois et al. 2000; Hamermesh and Parker 2005); and, difficulty in achieving high grades (Braga, Paccagnella, and Pellizzari 2014). These biases together with many other possible biases like weather and time of day were not in the scope of this study.

5. Research Design

The research design for this study was based on data collected and used in previous studies by the first author. The first is based on student evaluations of sessional laboratory demonstrators, the second based on the laboratory experience (experiment and facilities). The survey instrument design and findings on how training can be used to improve teaching effectiveness can be found in Nikolic, Vial, et al. (2015). The survey instrument design and findings on how to improve the laboratory experience can

be found in Nikolic, Ritz, et al. (2015). The correlation between the two survey instruments and learning can be found in Nikolic, Suesse, et al. (2015).

The data used for this study consisted of student evaluations run between 2012 and 2014. This consisted of 2519 survey responses across six teaching semesters. Evaluation surveys are only conducted in laboratories that are run by sessional laboratory demonstrators. The same laboratory demonstrator is used in a laboratory throughout the semester. The results of this study may be different for permanent faculty. This is due to the differences in teaching experience, additional biases, and knowledge about the course and structure. Laboratories that may have had any influence from other individuals, such as permanent faculty, have resulted in evaluations being removed from this study. This resulted in 2,161 survey responses being evaluated in this analysis.

The typical engineering course at the university consists of a lecture, tutorial, and laboratory per week. For many courses multiple, repetitive laboratory classes are scheduled. This is because the maximum laboratory class size possible due to equipment constraints is 45 students. Again, due to equipment constraints some laboratory class sizes could be as small as 10-15 students. The demonstrators usually have no input to the delivery, material and learning objectives associated with the experiments.

Having repeated laboratories with consistent factors such as assessment, structure, experiment, and facilities while having different laboratory demonstrators provides the framework for a rigorous multi-level analysis. Most SET based research studies are not multi-level as the comparisons are simply made between different courses, and/or the same courses over a number of years. This method has the limitation of many more variables, due to uncontrollable differences between subjects or years. As

the multiple laboratory classes have constant conditions, and use multiple demonstrators the analysis can be conducted between laboratory classes of the same course as well as between courses.

This scenario of having students within a laboratory class, and laboratory classes within a course, is a typical example of hierarchal data, for which typically a multi-level model, with levels students, laboratory classes and courses, is used for the statistical analysis (Berkhof and Kampen 2004; Goldstein 2003). Another, typical example in the educational literature is pupils within classes and classes within schools (Moerbeek 2004). Not accounting for the hierarchal structure of the data might lead to incorrect statistical results, as standard errors would be either under or overestimated (Moerbeek 2004). The statistical platform R (R Core Team 2013) and the R package lme4 (Bates et al. 2014) are used for the statistical analysis. In the analysis fixed effects (usually differences in the means) and p-values of the multilevel model will be presented. We apply a multivariate Wald test based on the multi-level approach. In some cases when multiple hypotheses are tested at once, we also apply the Bonferroni method or correction by multiplying the p-value with the number of hypotheses and comparing this adjusted p-value with the significance level (Abdi 2007).

We have also compared our data via a non-hierarchal analysis. To assess the effect of a factor with more than 2 levels, One-way ANOVA is applied, as well as p-values from Welch's t-test. However, the latter should be considered as unreliable, as it does not account for the hierarchal structure of the data collected in this study. The results of the Welch's t-test and ANOVA are presented for comparison purposes due to the fact that all studies outlined in section four did not use a multi-level model, highlighting the significance of this research. Our conclusions will always be based on the multi-level approach and based on a 5% significance level.

The data for the statistical analysis was obtained via a paper based survey instrument, conducted during a laboratory session that was administered to students towards the end of each semester. For each course a survey was only undertaken in multiple laboratory classes if different laboratory demonstrators were used. However, for some laboratories in a TBT format some demonstrators may be evaluated more than once. Across multiple teaching semesters, demonstrators involved in teaching a course may have repeated or changed. This is accounted for in the multi-level analysis due to the fact that in any given teaching semester for any given course the experiment, resources and facilities were constant. With the laboratory component being constant this research design was an effective approach to compare teaching evaluations between demonstrators.

Laboratory courses that run at least one and a half hours per week were surveyed. Twenty-five laboratory courses have been analysed. Experiments run for an entire semester either on a weekly or fortnightly basis. The experiments are generally based on three models: hardware experiment; combination of simulation/programming and hardware; or simulation/programming only. The twenty five courses that were analysed, including the number of multiple evaluations is shown in Table 1. Included in this data is the number of laboratory classes, laboratory courses and student sample evaluated in each teaching semester across the three years. Table 2 outlines the gender and ethnicity of the laboratory demonstrators, showing that male and international demonstrators were most heavily used across the three year period.

TABLE 1: Summary of the 25 laboratory courses investigated

Course Code	Course Name	2012		2013		2014	
		Sem 1	Sem 2	Sem 1	Sem 2	Sem 1	Sem 2
ECTE222	Power Engineering 1	2		3		3	
ECTE233	Digital Hardware	2		1		4	
ECTE301	Digital Signal Processing	1		1		1	
ECTE333 A	Microcontroller Architecture and Applications Part A	1		0		0	
ECTE344	Control Theory	2		2		3	
ECTE363	Communication Systems	3		3		2	
ECTE401/901	Multimedia Signal Processing	1		2		1	
ECTE412/912	Power Electronics and Drives	1		2		3	
ECTE423/923	Power System Analysis	1		2		3	
ECTE433/933	Embedded Systems	2		1		1	
ECTE170/172	Introduction to Circuits and Devices		3		3		3
ECTE182	Internet Technology 1		1		1		1
ECTE203	Signals and Systems		1		2		1
ECTE212	Electronics		2		2		1
ECTE290	Fundamentals of Electrical Engineering		4		2		4
ECTE323	Power Engineering 2		2		2		3
ECTE324	Foundations in Electrical Energy Utilisation		2		2		1
ECTE333 B	Microcontroller Architecture and Applications Part B		1		1		1
ECTE364	Data Communications		1		2		1
ECTE432/932	Computer Architecture		1		0		0
ECTE465/965	Wireless Communication Systems		1		1		1
ECTE469/962	Queuing Theory and Optimization		0		3		2
ECTE903	Image and Video Processing		1		1		0
ECTE906	Advanced Signals and Systems	1	1	1	1	2	0
ECTE955	Advanced Laboratory	2	1	1	1	1	1
	Total Laboratory Classes Surveyed	19	22	19	24	24	20
	Total Courses Surveyed	12	14	11	14	11	12
	Total Survey Responses Analyzed	316	432	245	419	368	381

TABLE 2: Summary of the course and demonstrator characteristics

Semester	No. of Courses	Male	Female	Australian Citizen	International	Survey Responses Analyzed
Semester 1 2012	12	17	7	1	24	316
Semester 2 2012	14	24	8	2	30	432
Semester 1 2013	11	14	6	1	19	245
Semester 2 2013	14	19	10	3	26	419
Semester 1 2014	11	19	9	4	24	368
Semester 2 2014	12	16	7	3	20	381
Number of unique demonstrators evaluated		41	15	5	51	

The operational management of the laboratory is allocated to the designated course coordinator (permanent faculty). The course coordinator is responsible for the design, assessment, and running of the laboratory. The allocation of laboratory demonstrators was conducted centrally, by a staff member responsible for managing, training and mentoring teaching assistants (the first author). The purpose of this structure was to ensure quality in the laboratory, with detailed reasoning available in Nikolic, Vial, et al. (2015). Course coordinators have the right to recommend laboratory demonstrators, but the final allocation was completed centrally by the first author. Considerations in allocation include experience, work load, and skills required. In larger classes, when a teaching team was needed, inexperienced laboratory demonstrators were partnered with experienced ones (a master/apprentice model). The only exception was for first year courses, due to the importance that the first year has on student retention (Pendergrass et al. 2001; Daempfle 2003; Karataş, Bodner, and Unal 2016).

Due to the assumption that two of the best demonstrators are needed in the laboratory to try and ensure a great first year experience only the most highly valued teaching assistants were used. This assumption highlights the importance of this study in resource allocation. That is, does only one demonstrator need to be exceptional, with the other used more effectively in another class?

Across all laboratory courses the ratio of teaching assistants to students was aimed to be one to 15. Therefore on average if a laboratory consists of 15 students, one laboratory demonstrator was used, if 35 students' two laboratory demonstrators were used and, if 45 students' three demonstrators were used. A breakdown of the number of courses that used the three different class formats is shown in Table 3. The type of format used was dependent on student numbers as well as timetable and equipment constraints. The number of courses sampled using three demonstrators is low, limiting the significance of data under this TBT format.

TABLE 3: Summary of courses using the three different class formats

Semester	One Demonstrator	Two Demonstrators	Three Demonstrators
Semester 1 2012	7	4	1
Semester 2 2012	5	6	3
Semester 1 2013	5	6	0
Semester 2 2013	6	7	1
Semester 1 2014	7	3	1
Semester 2 2014	3	8	1
Number of unique courses (<i>a course can be represented in multiple columns</i>)	16	16	5

5.1 Student Evaluation Instrument

The teaching allocations were assigned in that the most experienced demonstrator would lead the class and assume overall responsibility. This demonstrator would also be

responsible for providing an introduction to the students. In this study the lead demonstrator is referred to as DEM1 (demonstrator 1). When a TBT format is required the second and third demonstrators are used to assist, and follow the instructions of DEM1. The second and third demonstrators are referred to as DEM2 and DEM3, and are listed in the evaluation instrument in terms of experience. The six survey questions outlined in section three are used to evaluate DEM1 and five questions (Q2-3) for both DEM2 and DEM3. The difference of one question relates to the introduction required to be completed by DEM1. The survey of the laboratory experience was also outlined in section three. This consisted of three statements (S1-3) that evaluated the student's perception of the experiment, referred to as LAB1. Three statements (S4-6) were used to evaluate the equipment and facilities, referred to as LAB2. Full details of the instrument is available in Nikolic, Ritz, et al. (2015). The survey responses for the demonstrator and laboratory experience were converted to a weighted average score for comparison purposes.

Analysis of the survey questions was conducted to understand how the data could be cross compared. For the demonstrator survey questions, the smallest correlation is 0.3324, but which is still highly significant ($p < 0.0001$), similarly for DEM1, DEM2 and DEM3 the smallest correlations are 0.9053 ($p\text{-value} < 0.0001$), 0.9794 ($p\text{-value} < 0.0001$), and 0.9778 ($p\text{-value} < 0.0001$).

The next step was to confirm the number of components/factors within each learning domain. This is used to determine if and how the questions and statements could be grouped to produce a score. That is, for any questions/statements that could be grouped the average score can be considered. The default method of determining factors is via Kaiser Criterion by observing if the eigenvalues are greater than one. However, literature suggests that it should not be the only criterion as it tends to over extract

factors (Lance and Vandenberg 2009). Therefore, four different checks were used; Kaiser Rule, parallel analysis, optimal coordinates and acceleration factor.

For DEM1, DEM2 and DEM3, the methods suggest only one underlying factor meaning that all demonstrator questions can be grouped together to form one score. The largest and 2nd largest eigenvalues were 4.73 and 0.108 for DEM1, 3.94 and 0.02 for DEM2 and 3.94 and 0.023 for DEM3. Based on the Kaiser rule (Child 1990), these findings all suggest one underlying factor. Other methods such as parallel analysis and optimal coordinates also suggest only one factor using the R package “psych” (Revelle 2015). Similar to Johnson, Narayanan, and Sawaya (2013) we use the average score of all demonstrator questions to maintain interpretability.

The six laboratory experience questions had the eigenvalues 3.43, 1.11, 0.497, 0.39, 0.28 and 0.28. Using the R package “psych” (Revelle 2015), the Kaiser rule, parallel analysis and optimal coordinates all suggest two underlying factors (Child 1990). Factor loadings and “varimax” rotation was used to assess the groupings of the factors. Factor 1 has loadings 0.87 (Q1), 0.86 (Q2) and 0.83 (Q3), whereas factor 2 has loadings 0.77, 0.85 and 0.86 with all other loadings below the cut-off value of 0.3. A Biplot of a principal component analysis supported that Q1, Q2 and Q3 are clustered, as are Q4, Q5 and Q6. As the factor loadings are approximately equal (around 0.85) we use the average of Q1, Q2, Q3 scores and the average of Q4, Q5, Q6 scores as the two variables of interest, denoted by LAB1 and LAB2 in order to maintain interpretability, similar to Johnson, Narayanan, and Sawaya (2013) for one factor. To ensure reliability the standardized Cronbach’s α (R package “psych”) was calculated for all scores of interest. The values are: 0.85 (EXP), 0.82 (FACIL), 0.99 (DEM1), 1.00 (DEM2), 1.00 (DEM3) and are all above 0.70, a common cut-off value for validity.

5.2 Limitations

Each laboratory class has either one, two or three demonstrators. This means 2,161 student responses are available for the evaluation of the laboratory experience, but fewer responses are available for a particular demonstrator. Due to the different sample sizes, it is more likely to obtain significant results for laboratory scores compared to demonstrator scores. In addition, only a small sample is available for the case of three demonstrators. The study was conducted in a school of electrical, computer and telecommunications engineering and different disciplines may have different outcomes. Similarly, it is important to note that different approaches or pedagogies to laboratory learning may be prevalent across disciplines, universities or countries. The school has a large percentage of international students and international teaching assistants. Different combinations of student cohort, student quality, communication level, as well as different social attitudes towards race and gender could also alter the findings. All demonstrators had received on the job training and were recognised as capable. If untrained demonstrators were evaluated this could also have had an impact on the findings. These limitations are in addition to the possible forms of bias mentioned throughout section four.

6. Results

6.1 Class format: H1a and H1b was to confirm if class size had no effect on teaching or the laboratory experience scores in a team based teaching format, with constant student to staff ratio. Each laboratory class has a size of approximately 15 (1 demonstrator), 35 (2 demonstrators) or 45 students (3 demonstrators). Table 4 shows that H1a and H1b are supported by the analysis based on a multi-level model, as the mean differences between scores are not significant at the 5% significance level for any of the comparisons.

TABLE 4: A TBT comparison of class format using a multi-level model and Welch t-test

Score	Multi-level Model effect / p-value	Welsh t-test effect /p-value	Conclusion
LAB1			
15 vs 35	- 0.0565 / 0.9723	- 0.2963 / 0.6684	H1a supported
15 vs 45	3.3665 / 0.2466	3.0232 / 0.0137	
35 vs 45	- 3.4230 / 0.2150	3.3195 / 0.0042	
LAB2			
15 vs 35	-0.2088/0.8161	-0.1072 / 0.8490	H1a supported
15 vs 45	2.2555/0.1868	2.3730 / 0.0216	
35 vs 45	-2.4643/0.1417	2.4801 /0.0122	
DEM1			
15 vs 35	-0.5482 / 0.6749	-1.0867 /0.0780	H1b supported
15 vs 45	0.5578 / 0.8114	0.4022 / 0.7057	
35 vs 45	1.1061 /0.6332	1.4889 /0.1474	
DEM2			
15 vs 35	--	--	H1b supported
15 vs 45	--	--	
35 vs 45	1.445 / 0.4764	2.677 /0.0037	
DEM3			
15 vs 35	--	--	H1b supported
15 vs 45	--	--	
35 vs 45	1.033 / 0.6984	1.1961 / 0.3573	

6.2 Course level: H2a and H2b was to confirm if the higher level courses had a positive effect on teaching or the laboratory experience scores in a team based teaching format. The effect of course level (1st year, 2nd year, 3rd year and 4th year courses) was investigated. The results of a possible bias of course level (COURSE) are shown in Table 5a. The makeup of the courses was represented previously in Table 1 with the first digit of each course representing the course level. All 4th year courses surveyed contained postgraduate coursework students (900 level), so 400 and 900 level courses are treated as the same level.

TABLE 5a: Possible bias of course level compared to the laboratory demonstrator and experience

Score	Multi-level Model Wald statistic, p-value	One way ANOVA F statistic, p-value
COURSE on LAB1	W=1.625, p=0.6537	F=6.893, p=0.0001
COURSE on LAB2	W=2.821, p=0.4200	F=2.530, p=0.0556
COURSE on DEM1	W=14.67, p=0.0021	F=19.625, p<0.0001
COURSE on DEM2	W=8.765, p=0.0325	F=28.429, p<0.0001
COURSE on DEM3	W=6.232, p=0.1008	F=10.499, p<0.0001

The results from Table 5a show that there is an effect on DEM1 and DEM2 at the 5% significance level, but no effect on LAB1, LAB2 or DEM3 (using multi-level approach). Interestingly, the single level approach shows an effect on all tests, highlighting the importance of selecting the appropriate statistical model. To investigate how levels compare to other levels, Table 5b shows the effects (differences in mean) and p-values for every pair of course levels. The p-values are based on the multi-level approach. In general, H2b must be rejected, as there is no clear direction. For example

for DEM1 4th year courses score lower than 1st and 3rd year, but higher than 2nd year. Third year courses score significantly higher than 2nd year courses (even after a Bonferroni correction for multiple testing is applied, i.e. p-values are multiplied by 6 and compared with the significance level of 0.05). For DEM2, after a Bonferroni correction, there are no significant differences, except third year courses score higher than 2nd year.

Overall, as there is no clear pattern, H2a and H2b are rejected

TABLE 5b: *A comparison of course level between each year of the engineering degree*

	DEM1 effect	DEM1 p-value	DEM2 effect	DEM2 p-value
2nd vs 1st	-4.5215	0.0154	-3.931	0.1203
3rd vs 1st	0.5487	0.7651	2.9851	0.2827
4th vs 1st	-2.1288	0.2447	0.1821	0.9401
3rd vs 2nd	5.0702	0.0003*	6.9160	0.0045*
4th vs 2nd	2.3924	0.0837	4.1129	0.0443
4th vs 3rd	-2.6775	0.0460	-2.8031	0.2322

6.3 Gender: H3a and H3b was to confirm if the male demonstrators had a positive effect on teaching or the laboratory experience scores regardless of the class format. To test whether a male instructor has an effect on LAB scores, we calculate the average proportion of male instructors (PROPMALE). This proportion has values of 0, 1/3, 1/2, 2/3 and 1. For example 1/3 indicates 1 out of 3 demonstrators are male. Then we test whether the coefficient associated with PROPMALE is significant. We also test whether the gender of DEM1, DEM2 and DEM3 have an effect on LAB1 and LAB2.

To test H3b, we test whether ‘male’ demonstrators receive higher demonstrator scores.

Table 6 shows that none of the tests provide evidence at the 5% significance level.

Therefore, H3a and H3b are rejected.

TABLE 6: Comparison of gender and the laboratory demonstrator and experience

Score	Multi-level Model effect / p-value	Welch t- test/LM effect /p-value	Conclusion
PROPMale on LAB1	-0.1384 / 0.7605	0.0200 / 0.4102	H3a not supported
MALE DEM1 on LAB1	-1.3522 / 0.8183	-0.9861 / 0.9236	H3a not supported
MALE DEM2 on LAB1	0.0925 / 0.4789	0.7659 / 0.2087	H3a not supported
MALE DEM3 on LAB1	1.3848 / 0.3578	-0.1032 / 0.5272	H3a not supported
PROPMale on LAB2	-0.1139 / 0.8363	-0.0892 / 0.8887	H3a not supported
MALE DEM1 on LAB2	-0.4585 / 0.6888	-0.4401 / 0.7689	H3a not supported
MALE DEM2 on LAB2	1.4558 / 0.1138	1.4954 / 0.0353	H3a not supported
MALE DEM3 on LAB2	0.0774 / 0.4838	-0.0416 / 0.5137	H3a not supported
MALE DEM1 on DEM1	0.4200 / 0.3665	0.6556 / 0.1534	H3b not supported
MALE DEM2 on DEM2	-1.0651 / 0.7425	-0.4239 / 0.6972	H3b not supported
MALE DEM3 on DEM3	-1.0292 / 0.6480	-2.6075 / 0.9774	H3b not supported

6.4 Relationship: H4a was to confirm if a positive relationship existed between teaching and laboratory experience scores regardless of the class format. In order to investigate the relationships of LAB1, LAB2 and demonstrator scores (DEM1, DEM2 and DEM3), we need to take into account that only some classes have only i) DEM1, some have ii) DEM1 and DEM2, and some iii) DEM2 and DEM3 (when an

introduction is not required in the class) and some iv) DEM1, DEM2 and DEM3. We consider these cases separately using the multi-level model analysis.

Table 7 shows the results under i), Table 8 under ii), Table 9 under iii) and Table 10 under scenario iv). All results show that there is a strong positive relationship between LAB1 scores and demonstrator scores. For example Table 7 shows that for an increase in one unit of the DEM1 score results in an increase of 0.5451 of the LAB score. Tables 8, 9 and 10 show a similar pattern, but they also show that DEM1 scores are more important than DEM2 scores and DEM2 scores are more important than DEM3 scores, except for scenario iv) which indicates that DEM3 scores could be more important than DEM2 scores. Table 7 shows that DEM3 scores are not significant, which could be due to low sample size for scenario iii). The same analysis was repeated between LAB2 scores and demonstrator scores. A very similar outcome was found for LAB2 scores across all four scenarios. Table 11 shows the relationship for scenario iv) between LAB2 scores and the three demonstrator scores. These results confirm that DEM1 has the largest influence on laboratory experience scores and is therefore most important for students to evaluate the teaching of the laboratory classes.

TABLE 7: Relationship between DEM1 and the laboratory experience

	Estimate	Standard Error	t-value	p-value (2-sided)
Intercept	35.3687	2.1368	16.55	<0.0001
DEM1	0.5451	0.0225	24.17	<0.0001
	Level 1 (students)	Level 2 (class)	Level 3 (Courses)	
Variance- Estimates	132.13	20.691	6.061	$R^2=0.3616$

TABLE 8: Relationship between DEM1 and DEM2 with the laboratory experience

	estimate	Standard Error	t-value	p-value (2-sided)
Intercept	28.1943	2.7814	10.137	<0.0001
DEM1	0.3790	0.0384	9.859	<0.0001
DEM2	0.2486	0.0378	6.562	<0.0001
	Level 1 (students)	Level 2 (class)	Level 3 (Courses)	
Variance- Estimates	114.217	4.667	7.844	$R^2=0.4908$

TABLE 9: Relationship between DEM2 and DEM3 (when no laboratory introduction needed with two demonstrators) with the laboratory experience.

	estimate	Standard Error	t-value	p-value (2-sided)
Intercept	26.7086	4.6322	5.766	<0.0001
DEM2	0.4087	0.0767	5.328	<0.0001
DEM3	0.2547	0.0701	3.595	0.0003
	Level 1 (students)	Level 2 (class)	Level 3 (Courses)	
Variance- Estimates	114.62	12.01	6.45	$R^2=0.4653$

TABLE 10: Relationship between DEM1, DEM2 and DEM3 with the laboratory experience

	estimate	Standard Error	t-value	p-value (2-sided)
Intercept	15.3720	7.1941	2.137	0.0326
DEM1	0.2931	0.1085	2.700	0.0069
DEM2	0.2783	0.1192	2.334	0.0196
DEM3	0.2177	0.1324	1.644	0.1002
	Level 1 (students)	Level 2 (class)	Level 3 (Courses)	
Variance- Estimates	114.94	0.00	12.68	$R^2=0.4873$

TABLE11: Relationship between LAB2 scores and the three demonstrator scores

	estimate	Standard Error	t-value	p-value (2-sided)
DEM1				
Intercept	57.3383	1.9690	29.12	<0.0001
LAB2	0.3594	0.0213	16.86	<0.0001
DEM2				
Intercept	58.5487	2.4239	24.15	<0.0001
LAB2	0.3539	0.0254	13.92	<0.0001
DEM3				
Intercept	47.8763	4.5387	10.548	<0.0001
LAB2	0.45720	0.0471	9.702	<0.0001

However the results also show that DEM-ratings are not fully explaining the LAB1 ratings, as R^2 the proportion of the variance explained ranges from 0.36 to 0.49 showing that 36%-49% of the variance of LAB1 scores is explained by demonstrator scores. The highest values are obtained when more than two demonstrator scores are available, showing that all three demonstrators ratings jointly contribute to an accurate prediction of the laboratory scores.

7. Discussion

The first hypothesis was to investigate the effect of class size on student evaluations scores when using a team based teaching format. Literature had suggested that class size does have an impact on student evaluations (Shapiro 1990; Watkins 1990; El Ansari and Oskrochi 2006; Johnson, Narayanan, and Sawaya 2013). As the ratio of teachers to students in the laboratory was relatively constant using team based teaching, no impact was suspected. The results show that based on a multi-level model, as the mean differences between scores for different classes are not significant at the 5% significance level, class size had no impact on evaluation scores of teaching staff and the laboratory experience. This can provide weight to the benefits of teaching teams in the laboratory, and the use of a master/apprentice model for demonstrator training.

The literature suggested course level as a potential bias in influencing student ratings, especially in engineering (Badri et al. 2006; Johnson, Narayanan, and Sawaya 2013; Narayanan, Sawaya, and Johnson 2014). Analysis of the results suggest that this bias was not present regardless of the class format, as the differences in mean and corresponding p-values for every pair of course levels showed no clear pattern in any direction. Therefore, student ratings of the demonstrators and laboratory experience did not increase with course level. However, the results did show that the 3rd and 4th year results were greater than the 2nd year results. This does tend to suggest that in the years

the students participate in more discipline specific courses, the evaluation is more favourable.

Findings on the effects of teacher gender on student evaluations from the literature has been mixed. Some studies suggested no effect (Aleamoni 1999; Feldman 1993; Hameed et al. 2014), while others, including large engineering studies did (Basow and Silberg 1987; McPherson, Jewell, and Kim 2009; Johnson, Narayanan, and Sawaya 2013; Narayanan, Sawaya, and Johnson 2014). The results from the study show that none of the tests indicated a significant correlation regardless of class format. The gender of the demonstrator did not provide a significant influence on student evaluations.

The final hypothesis was to investigate the relationship between student evaluation scores on teaching and the laboratory experience. This understanding is needed to help quantify the importance of the laboratory demonstrator as well as to understand how to apply this most effectively in a TBT class format. The laboratory experience was broken into two components as a consequence of the factor analysis. The two components, one focussed on the experiments and the other on the facilities had a consistent finding. The results show that the lead demonstrator has the greatest influence on the laboratory experience score, regardless if one, two or three demonstrators are used. As expected, in a laboratory with just one demonstrator the influence is the highest, with an increase in one unit of the DEM1 score resulting in an increase of 0.5451 of the LAB1 score. In a team based teaching format, a lower but still important influence was shown for the other two demonstrators. This shows the importance on selecting high quality demonstrators, and providing laboratory specific training, due to the influence they have on the laboratory experience. In particular, it shows the significance of using the best lead demonstrator possible. The finding also

shows that using a master/apprentice model in a TBT format will not have a major impact on the student laboratory experience as long as the lead demonstrator is of high calibre. This should help provide allocators of teaching staff the confidence to use TBT to help train less experienced demonstrators. In addition, researchers evaluating the success of changes to laboratory experiment's or facilities can now quantify the impact the teaching staff have in their research design.

The results also show that while the laboratory demonstrators play a large role in influencing the laboratory experience only 36% to 49% of the variance is explained by demonstrator scores. Therefore, from this research we could suggest that such other factors contribute to at least 50% of the student evaluations of the laboratory experience. The remaining unexplained variation could be further explained by factors collected in the study, factors that could be observed but were not collected, and other factors that are not observable. For example, other demonstrator factors such as age and ethnicity that were not included, could further reduce the unexplained variance. Factors associated with the laboratory experience are also very important. Previous research by the Nikolic, Ritz, et al. (2015), found that the laboratory notes (activity and clarity), quality of equipment, and workload within the allocated laboratory timeslot influenced laboratory experience scores.

The findings, together with the findings from related studies (Nikolic, Ritz, et al. 2015; Nikolic, Vial, et al. 2015) point to the fact that the teaching staff add significant value to the laboratory learning experience be it computer based, hardware based or a combination of both. The authors support all forms of laboratory learning including in a physical laboratory, online, virtual and remote under the premise that they follow the learning objectives required of the course and the thirteen fundamental objectives of engineering instructional laboratories (Feisel and Rosa 2005) across the entire degree.

Each approach has a pedagogical strength and weakness and therefore using a mixture of approaches is recommended. In particular, as outlined in Nikolic (2015) student learning can be compromised if laboratory learning is targeted predominately towards cognitive development only, particularly online forms of learning. For example, there is a substantial difference between remotely connecting two components together than doing it physically.

Laboratory learning is about the development of cognitive, psychomotor and affective skills (Nikolic, Suesse, et al. 2015; Salim et al. 2013) and the teaching staff play a key role in that development. At the minimum, students need some exposure to teachers that can assist students to work with their hands; question their understanding of measurements; appreciate the importance of smell, touch and sound in troubleshooting; guide them to learn independently through help files, technical datasheets, Google and YouTube; and many other skills covered in the fundamental laboratory learning objectives outlined in Feisel and Rosa (2005). The less exposure students have to teaching staff in the laboratory, the greater the importance in having high quality teaching staff to cover the development of such skills in minimum time.

While this study has shown the impact teaching staff has on the laboratory experience it is focussed only in one school, one university, one country. The teaching capability of the demonstrators in the study vary, but all meet a quality threshold (Nikolic, Vial, et al. 2015). The findings from this research can be strengthened by testing this relationship with permanent faculty; a mixture of high and low quality teaching assistants; and different disciplines, universities and countries.

8. Conclusion

This paper investigated the use of student evaluations to gain a greater understanding of team based teaching in an engineering teaching laboratory. Evaluations were conducted

on laboratory teaching assistants, called laboratory demonstrators, and analysed using a multi-level model. The paper also showcased the importance of using multi-level statistical models in analysing student evaluations due to the contrasting findings of single level models. The TBT approach found no bias in terms of class format, course level and demonstrator gender.

A key conclusion is that team based teaching is a valued way of enhancing the laboratory learning experience. As expected, the lead demonstrator provides the most influence to the laboratory experience. With this influence quantified, staff conducting teaching allocations should be encouraged to use the master/apprentice model by partnering up very experienced demonstrators with less experienced ones. The experience is extremely valuable and the effect of a negative influence on the laboratory experience is small. This quantification will also assist researchers that are evaluating improvements in the laboratory to understand the possible influence the demonstrators have on the research design.

References

- Abdi, H. 2007. "Bonferroni and Šidák corrections for multiple comparisons. In NJ Salkind (ed.). *Encyclopedia of Measurement and Statistics*." Review of. *Encyclopedia of measurement and statistics*.
- Abdulwahed, Mahmoud, and Zoltan K. Nagy. 2009. "Applying Kolb's Experiential Learning Cycle for Laboratory Education." Review of. *Journal of Engineering Education* 98 (3):283-94. doi: 10.1002/j.2168-9830.2009.tb01025.x.
- Aleamoni, Lawrence M. 1999. "Student Rating Myths Versus Research Facts from 1924 to 1998." Review of. *Journal of Personnel Evaluation in Education* 13 (2):153-66.
- Arbuckle, Julianne, and Benne D Williams. 2003. "Students' perceptions of expressiveness: Age and gender effects on teacher evaluations." Review of. *Sex Roles* 49 (9-10):507-16.
- Badri, Masood A, Mohamed Abdulla, Mohammed A Kamali, and Hamzeh Dodeen. 2006. "Identifying potential biasing variables in student evaluation of teaching in a newly accredited business program in the UAE." Review of. *International Journal of Educational Management* 20 (1):43-59.
- Bandiera, Oriana, Valentino Larcinese, and Imran Rasul. 2010. "Heterogeneous Class Size Effects: New Evidence from a Panel of University Students*." Review of.

The Economic Journal 120 (549):1365-98. doi: 10.1111/j.1468-0297.2010.02364.x.

- Basow, Susan A, and Nancy T Silberg. 1987. "Student evaluations of college professors: are female and male professors rated differently?" Review of. *Journal of Educational Psychology* 79 (3):308.
- Bates, D, M Maechler, B Bolker, and S Walker. 2014. "lme4: Linear mixed-effects models using Eigen and S4 (Version 1.0-6)[R package] Available from <http://CRAN.R-project.org/package=lme4>." Review of.
- Bavishi, Anish, Juan M Madera, and Michelle R Hebl. 2010. "The effect of professor ethnicity and gender on student evaluations: Judged before met." Review of. *Journal of Diversity in Higher Education* 3 (4):245.
- Benton, Steve, and Dan Li. 2014. "What's in the Study: Exposing Validity Threats In the MacNell, Driscoll, and Hunt Study of Gender Bias." In *What's in the Study: Exposing Validity Threats In the MacNell, Driscoll, and Hunt Study of Gender Bias*.
- Berkhof, Johannes, and Jarl Kennard Kampen. 2004. "Asymptotic effect of misspecification in the random part of the multilevel model." Review of. *Journal of Educational and Behavioral Statistics* 29 (2):201-18.
- Blackhart, Ginette C, B Michelle Peruche, C Nathan DeWall, Jr Joiner, Thomas E, K Laurie Dickson, Michelle D Miller, K Laurie Dickson, Michael S Devoley, Michelle D Miller, and Andrew N Christopher. 2006. "Factors Influencing Teaching Evaluations in Higher Education." Review of. *Teaching of Psychology* 33 (1):37-63.
- Braga, Michela, Marco Paccagnella, and Michele Pellizzari. 2014. "Evaluating students' evaluations of professors." Review of. *Economics of Education Review* 41 (0):71-88. doi: <http://dx.doi.org/10.1016/j.econedurev.2014.04.002>.
- Buckley, Francis J. 1999. *Team teaching: what, why, and how?* : Sage.
- Carless, David, and Elizabeth Walker. 2006. "Effective Team Teaching between Local and Native-speaking English Teachers." Review of. *Language and Education* 20 (6):463-77. doi: 10.2167/le627.0.
- Casas, Jordi de la Hoz i, and Alfredo de Blas del Hoyo. 2009. "'Learning by doing' methodology applied to the practical teaching of electrical machines." Review of. *International Journal of Electrical Engineering Education* 46 (2):133-49.
- Centra, John A., and Noreen B. Gaubatz. 2000. "Is There Gender Bias in Student Evaluations of Teaching?" Review of. *The Journal of Higher Education* 71 (1):17-33. doi: 10.2307/2649280.
- Che Ahmad, CheNidzam, Kamisah Osman, and Lilia Halim. 2013. "Physical and psychosocial aspects of the learning environment in the science laboratory and their relationship to teacher satisfaction." Review of. *Learning Environments Research* 16 (3):367-85. doi: 10.1007/s10984-013-9136-8.
- Child, Dennis. 1990. *The essentials of factor analysis*: Cassell Educational.
- Cranton, Patricia A., and Ronald A. Smith. 1986. "A New Look at the Effect of Course Characteristics on Student Ratings of Instruction." Review of. *American Educational Research Journal* 23 (1):117-28. doi: 10.2307/1163047.
- Daempfle, Peter A. 2003. "An analysis of the high attrition rates among first year college science, math, and engineering majors." Review of. *Journal of College Student Retention: Research, Theory and Practice* 5 (1):37-52.
- Deacon, Christopher, and Allyson Hajek. 2010. "Student Perceptions of the Value of Physics Laboratories." Review of. *International Journal of Science Education* 33 (7):943-77. doi: 10.1080/09500693.2010.481682.

- Deshwal, Pankaj, Ashima Mahajan, and Gaurav Choudhary. 2012. "A Study of Satisfaction Among Engineering Students in Delhi." Review of. *International Journal of Management Research and Reviews* 2 (7):1280-4.
- Douglas, Jacqueline, Alex Douglas, and Barry Barnes. 2006. "Measuring student satisfaction at a UK university." Review of. *Quality Assurance in Education* 14 (3):251 - 67.
- El Ansari, W, and R Oskrochi. 2006. "What matters most? Predictors of student satisfaction in public health educational courses." Review of. *Public health* 120 (5):462-73.
- Feisel, Lyle D., and Albert J. Rosa. 2005. "The Role of the Laboratory in Undergraduate Engineering Education." Review of. *Journal of Engineering Education* 94 (1):121-30.
- Feldman, Kenneth A. 1993. "College students' views of male and female college teachers: Part II—Evidence from students' evaluations of their classroom teachers." Review of. *Research in Higher Education* 34 (2):151-211. doi: 10.1007/BF00992161.
- Galbraith, Craig S, Gregory B Merrill, and Doug M Kline. 2012. "Are Student Evaluations of Teaching Effectiveness Valid for Measuring Student Learning Outcomes in Business Related Classes? A Neural Network and Bayesian Analyses." Review of. *Research in Higher Education* 53 (3):353-74.
- Goldstein, H. 2003. "Multilevel statistical models." Review of. *Edward Arnold, London*.
- Gonsai, Atul M, Anil Ambasana, Bhargavi H Goswami, Manasi A Antani, and Vidhdhi J Rughani. 2013. "A Study of Lab Satisfaction of MCA and MSc. (IT) Students of Department Computer Science of Saurashtra University, Rajkot." Review of. *Academia.edu*:4.
- Hameed, Fatima, Amjad Ali, Amina Hameed, Zohra Saleem, and Yasir Javed. 2014. "Teacher evaluation: the role of gender." Review of. *Quality & Quantity*:1-11. doi: 10.1007/s11135-014-0054-3.
- Hamermesh, Daniel, and Amy M. Parker. 2005. "Beauty in the Classroom: Professors' Pulchritude and Putative Pedagogical Productivity." Review of. *Economics of Education Review* 24 (4):369-76.
- Johnson, Michael D., Arunachalam Narayanan, and William J. Sawaya. 2013. "Effects of Course and Instructor Characteristics on Student Evaluation of Teaching across a College of Engineering." Review of. *Journal of Engineering Education* 102 (2):289-318. doi: 10.1002/jee.20013.
- Karataş, F. Ö, G. M. Bodner, and Suat Unal. 2016. "First-year engineering students' views of the nature of engineering: implications for engineering programmes." Review of. *European Journal of Engineering Education* 41 (1):1-22. doi: 10.1080/03043797.2014.1001821.
- Khan, Wasim Ahmed, Saad Muhammad Rasheed Al-Doussari, and A. H. M. Al-Kahtani. 2002. "Establishment of engineering laboratories for undergraduate and postgraduate studies." Review of. *European Journal of Engineering Education* 27 (4):425-35. doi: <http://dx.doi.org/10.1080/03043790210166684>.
- Koretsky, Milo, Christine Kelly, and Edith Gummer. 2011. "Student Perceptions of Learning in the Laboratory: Comparison of Industrially Situated Virtual Laboratories to Capstone Physical Laboratories." Review of. *Journal of Engineering Education* 100 (3):540-73. doi: 10.1002/j.2168-9830.2011.tb00026.x.

- Lance, Charles E, and Robert J Vandenberg. 2009. *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*: Taylor & Francis.
- Langlois, Judith H, Lisa Kalakanis, Adam J Rubenstein, Andrea Larson, Monica Hallam, and Monica Smoot. 2000. "Maxims or myths of beauty? A meta-analytic and theoretical review." Review of. *Psychological bulletin* 126 (3):390.
- Lin, Wen-ying. 1992. "Is Class Size a Bias to Student Evaluations of Faculty? A Review." Review of. *Chinese University Education Journal* 20 (1):49-53.
- Lowe, David, Bridgette Dang, Keith Daniel, Stephen Murray, and Euan Lindsay. 2015. "On the viability of supporting institutional sharing of remote laboratory facilities." Review of. *European Journal of Engineering Education* 40 (6):611-22. doi: <http://dx.doi.org/10.1080/03043797.2014.1001815>.
- MacNell, Lillian, Adam Driscoll, and Andrea N Hunt. 2014. "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching." Review of. *Innovative Higher Education*:1-13.
- McPherson, Michael A, R Todd Jewell, and Myungsup Kim. 2009. "What determines student evaluation scores? A random effects analysis of undergraduate economics classes." Review of. *Eastern Economic Journal* 35 (1):37-51.
- Moerbeek, Mirjam. 2004. "The consequence of ignoring a level of nesting in multilevel analysis." Review of. *Multivariate Behavioral Research* 39 (1):129-49.
- Narayanan, Arunachalam, William J. Sawaya, and Michael D. Johnson. 2014. "Analysis of Differences in Nonteaching Factors Influencing Student Evaluation of Teaching between Engineering and Business Classrooms." Review of. *Decision Sciences Journal of Innovative Education* 12 (3):233-65. doi: <http://dx.doi.org/10.1111/dsji.12035>.
- Nikolic, Sasha. 2014. Training laboratory: Using online resources to enhance the laboratory learning experience. Paper presented at the Teaching, Assessment and Learning (TALE), 2014 International Conference on.
- . 2015. "Understanding How Students Use and Appreciate Online Resources in the Teaching Laboratory." Review of. *International Journal of Online Engineering* 11 (4):8-13. doi: <http://dx.doi.org/10.3991/ijoe.v11i4.4562>.
- Nikolic, Sasha, Christian Ritz, Peter J. Vial, Montserrat Ros, and David Stirling. 2015. "Decoding Student Satisfaction: How to Manage and Improve the Laboratory Experience." Review of. *IEEE Transactions on Education* 58 (3):151-8. doi: <http://dx.doi.org/10.1109/TE.2014.2346474>.
- Nikolic, Sasha, Thomas Suesse, Thomas Goldfinch, and Timothy McCarthy. 2015. "Relationship between Learning in the Engineering Laboratory and Student Evaluations." In *Australasian Association for Engineering Education Annual Conference*. Geelong, Australia.
- Nikolic, Sasha, Peter J. Vial, Montserrat Ros, David Stirling, and Christian Ritz. 2015. "Improving the Laboratory Learning Experience: A Process to Train & Manage Teaching Assistants." Review of. *IEEE Transactions on Education* 58 (2):130-9. doi: <http://dx.doi.org/10.1109/TE.2014.2335712>.
- O'Toole, Paddy, Andrea Crampton, Gerry Rayner, Alison Beavis, Stephanie Beames, Peter Meier, and Julian Cox. 2012. "Demonstrator development: Preparing for the Learning Lab." In *The Australian Council of Deans of Science*.
- Pendergrass, N. A., Robert E. Kowalczyk, John P. Dowd, Raymond N. Laoulache, William Nelles, James A. Golen, and Emily Fowler. 2001. "Improving First-Year Engineering Education." Review of. *Journal of Engineering Education* 90 (1):33-41. doi: 10.1002/j.2168-9830.2001.tb00564.x.

- Plank, Richard E., and Larry Chiagouris. 1997. "Perceptions of Quality of Higher Education: An Exploratory Study of High School Guidance Counselors." Review of. *Journal of Marketing for Higher Education* 8 (1):55-67. doi: 10.1300/J050v08n01_05.
- Pozo-Munoz, Carmen, Enrique Reboloso-Pacheco, and Baltasar Fernandez-Ramirez. 2000. "The 'Ideal Teacher'. Implications for student evaluation of teacher effectiveness." Review of. *Assessment & Evaluation in Higher Education* 25 (3):253-63.
- R Core Team. 2013. "A Language and Environment for Statistical Computing. Vienna, Austria: R foundation for statistical computing." Review of. URL <http://www.R-project.org>.
- Rathod, SS, and DR Kalbande. 2016. "Improving Laboratory Experiences in Engineering Education." Review of. *Journal of Engineering Education Transformations*. doi: <http://dx.doi.org/10.16920/jeet/2016/v0i0/85576>.
- Reid, Landon D. 2010. "The role of perceived race and gender in the evaluation of college teaching on RateMyProfessors. Com." Review of. *Journal of Diversity in Higher Education* 3 (3):137.
- Revelle, William. 2015. "Procedures for psychological, psychometric, and personality research." Review of. *Computer software manual*]. Evanston, IL: Northwestern University. Retrieved from <http://www.personality-project.net/r/psychmanual.pdf> (R package version 1.5.4).
- Salim, Kamilah Radin, ALI Rosmah, Noor Hamizah Hussain, and Habibah Norehan Haron. 2013. "An Instrument for Measuring the Learning Outcomes of Laboratory Work." In *International Engineering and Technology Education Conference*. Ho Chi Minh City, Vietnam.
- Scherr, Frederick C, and Susan S Scherr. 1990. "Bias in Student Evaluations of Teacher Effectiveness." Review of. *Journal of Education for Business* 65 (8):356-58.
- Shapiro, E. Gary. 1990. "Effect of instructor and class characteristics on students' class evaluations." Review of. *Research in Higher Education* 31 (2):135-48. doi: 10.1007/BF00992258.
- Spooren, Pieter, Bert Brockx, and Dimitri Mortelmans. 2013. "On the Validity of Student Evaluation of Teaching The State of the Art." Review of. *Review of Educational Research* 83 (4):598-642.
- Stamatelos, G. M., and A. M. Stamatelos. 2009. "The role of engineering laboratories in the establishment of a quality culture in higher education in Greece." Review of. *European Journal of Engineering Education* 34 (1):1-13. doi: <http://dx.doi.org/10.1080/03043790802699115>.
- Stanisavljevic, Z., V. Pavlovic, B. Nikolic, and J. Djordjevic. 2013. "SDLDS - System for Digital Logic Design and Simulation." Review of. *IEEE Transactions on Education* 56 (2):235-45. doi: 10.1109/TE.2012.2211598.
- Stolte, John F. 1996. "Evaluation of persons of varying ages." Review of. *The Journal of Social Psychology* 136 (3):305.
- Ugursal, V Ismet, and Cynthia A Cruickshank. 2015. "Student opinions and perceptions of undergraduate thermodynamics courses in engineering." Review of. *European Journal of Engineering Education* 40 (6):593-610. doi: <http://dx.doi.org/10.1080/03043797.2014.987646>.
- Vial, Peter J., Sasha Nikolic, Montserrat Ros, David Stirling, and Parviz Doulai. 2015. "Using Online and Multimedia Resources to Enhance the Student Learning Experience in a Telecommunications Laboratory within an Australian

- University." Review of. *Australasian Journal of Engineering Education* 20 (1):71-80. doi: <http://dx.doi.org/10.7158/D13-006.2015.20.1>.
- Walker, D. J., and E. Palmer. 2009. "The relationship between student understanding, satisfaction and performance in an Australian engineering programme." Review of. *Assessment & Evaluation in Higher Education* 36 (2):157-70. doi: 10.1080/02602930903221451.
- Watkins, David. 1990. "Student ratings of tertiary courses for 'alternative calendar' purposes." Review of. *Assessment and Evaluation in Higher Education* 15 (1):12-21.
- Yanamandram, V, and G Noble. 2006. "Student Experiences and Perceptions of Team-Teaching in a Large Undergraduate Class." Review of. *Journal of University Teaching & Learning Practice* 3 (1):6.
- Zabaleta, Francisco. 2007. "The use and misuse of student evaluations of teaching." Review of. *Teaching in Higher Education* 12 (1):55-76. doi: 10.1080/13562510601102131.