Laboratory Learning Objectives Measurement: Relationships Between Student Evaluation Scores and Perceived Learning

Sasha Nikolic
*University of Wollongong, sasha@uow.edu.au*

Thomas Suesse
*University of Wollongong, tsuesse@uow.edu.au*

Kosta Jovanovic
*University of Belgrade*

Zarko Stanisavljevic
*University of Belgrade*

Follow this and additional works at: [https://ro.uow.edu.au/eispapers1](https://ro.uow.edu.au/eispapers1)

Part of the Engineering Commons, and the Science and Technology Studies Commons

**Recommended Citation**

Nikolic, Sasha; Suesse, Thomas; Jovanovic, Kosta; and Stanisavljevic, Zarko, "Laboratory Learning Objectives Measurement: Relationships Between Student Evaluation Scores and Perceived Learning" (2021). *Faculty of Engineering and Information Sciences - Papers: Part B. 4557.*


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
Laboratory Learning Objectives Measurement: Relationships Between Student Evaluation Scores and Perceived Learning

Abstract
Contribution: This article provides evidence that perceived learning has a relationship and influences the way students evaluate laboratory experiments, facilities, and demonstrators. Background: Debate continues on the capability and/or reliability of students to evaluate teaching and/or learning. Understanding such relationships can help educators decode evaluation data to develop more effective teaching experiences. Research Question: Does a relationship exist between student evaluation scores and perceived learning? Methodology: Perceived learning across the cognitive, psychomotor, and affective domains was measured using the Laboratory Learning Objectives Measurement (LLOM) tool at an Australian (344 students) and Serbian (181 students) university. A multilevel statistical analysis was conducted. Findings: Statistically significant relationships were found between student evaluation scores and perceived learning across the cognitive, psychomotor, and affective domains with some differences found between the two universities. This provides evidence that perceived learning plays a role in influencing student evaluation scores. Students perceived an improvement of learning across all three domains confirming the multifaceted benefits of the laboratory for engineering education.

Disciplines
Engineering | Science and Technology Studies

Publication Details

This journal article is available at Research Online: https://ro.uow.edu.au/eispapers1/4557
Laboratory Learning Objectives Measurement: Relationships between Student Evaluation Scores and Perceived Learning

Sasha Nikolic, Senior Member, IEEE, Thomas F. Suesse, Kosta Jovanovic, Member, IEEE, and Zarko Stanisavljevic

Abstract—
Contribution: This paper provides evidence that perceived learning has a relationship and influences the way students evaluate laboratory experiments, facilities and demonstrators.

Background: Debate continues on the capability and/or reliability of students to evaluate teaching and/or learning. Understanding such relationships can help educators decode evaluation data to develop more effective teaching experiences.

Research Question: Does a relationship exist between student evaluation scores and perceived learning?

Methodology: Perceived learning across the cognitive, psychomotor and affective domains was measured using the Laboratory Learning Objectives Measurement (LLOM) tool at an Australian (344 students) and Serbian (181 students) university. A multi-level statistical analysis was conducted.

Findings: Statistically significant relationships were found between student evaluation scores and perceived learning across the cognitive, psychomotor and affective domains with some differences found between the two universities. This provides evidence that perceived learning plays a role in influencing student evaluation scores. Students perceived an improvement of learning across all three domains confirming the multifaceted benefits of the laboratory for engineering education.

Index Terms— Blooms Taxonomy, Laboratory, Learning Objectives, Program evaluation, Student Experience

I. INTRODUCTION

Student evaluations of teaching were first used in 1920 and have now become mainstream across most of the world, transitioning from solely being a tool to provide feedback on teaching to a measure of quality and linked to funding, promotion and other metrics [1, 2]. Thousands of research studies have participated in the debate [3], as to the capability and/or reliability of students to evaluate teaching and/or learning. Student evaluation data can be beneficial if used appropriately and dangerous if applied without understanding the instrument [4]. Evaluations scores can be impacted by many possible influences ranging from the gender of the teaching staff to the course level, time of day and grade leniency [5-7].

One pocket of research of significance is trying to determine if relationships exist between student evaluation scores and learning with some studies showing links [8, 9] and others not [6, 10]. Understanding this relationship can help provide greater clarity as to the message being expressed by the students. For example, is the message a confirmation of a good learning experience or confirmation that the student expects an enjoyable customer experience? [11]. What learning is occurring? The inconclusiveness showcases the need to develop a clearer picture of why students evaluate the way they do.

While this learning relationship is important, it is also very difficult to analyze, simply because learning is very difficult to measure. Learning is more than just differences in test results (a student can get a high-test score or grade and not have necessarily learned anything) or analysis of course defined learning objectives. With many student evaluations conducted during session, before final examination and release of final results, how aware are students of how much they have learned at the time of completing the evaluation? If any relationships held at the time of completing the evaluation it would most probably be centered on student’s perception of their learning at the time [12, 13]. Understanding if perceived learning influences the way students evaluate would provide some evidence of the weight and value of their opinion. Therefore, this study contributes to existing research by exploring if relationships exist between student evaluation scores and perceived learning. The aim is to develop a better understanding if and how perceived learning influences students in their evaluation decisions; helping teaching staff better understand what the student voice is telling them and how to incorporate this feedback into facilitating better laboratory learning experiences. A holistic tool for measuring laboratory learning outcomes is used in an engineering laboratory context at an Australian and Serbian university.

K. Jovanovic is with the School of Electrical Engineering, University of Belgrade, Belgrade, Serbia (e-mail: kostaj@etf.bg.ac.rs)
Z. Stanisavljevic is with the School of Electrical Engineering, University of Belgrade, Belgrade, Serbia (e-mail: zarko.stanisavljevic@etf.bg.ac.rs)
Colour versions of one or more of the figures in this paper are available online at http://ieeexplore.org.au
Digital Object Identifier XXXXXXXXXXXXXXXXXXXXX
II. Related Literature

The teaching laboratory is acknowledged as playing an important role in developing engineering students [14]. Laboratory learning activities can encompass many forms, including and not limited to, interacting with physical equipment, simulations, virtual and remote experiments; each approach having different learning and logistical advantages and limitations. For example, research by Lindsay and Good [15] found that different modes (such as hands-on and remote labs) have different advantages and limitations in which a particular mode may improve some learning outcomes at the expense of degradation in others. Measuring success of such various forms of laboratory activities can be difficult.

Research focused on exploring the impact of changes to a laboratory generally consist of a quantitative and qualitative component. Most commonly this consists of an analysis of multiple years of scores or grades supplemented with some form of student evaluation. Examples include: investigating the impact of changes to a control laboratory by looking at differences in grades across multiple years and a satisfaction survey [16]; and an attempt to improve an introductory programming course by comparing seven years of grades with a student questionnaire [17]. While such studies contribute to the field, the use of grades provides very limited insight to the learning achieved, and little is known about the relationship between the student questionnaires and learning. What factors contributed to the way students evaluated their experience? If a link could be made between the ways students evaluated the laboratory experience with the way they perceived an improvement in learning, this would result in an appreciation of the value of such feedback. Therefore, there is much still to be learnt about the way students evaluate their experience.

The difficulties in measuring learning is further complicated in the laboratory. This is because experiments generally provide learning experiences that easily enrich multiple learning domains. Bloom’s Taxonomy considers learning across three overlapping domains: cognitive (reflect students’ knowledge and thinking skills); psychomotor (focus on manual tasks that require the manipulation of objects or apparatus which involves the coordination between the brain and body in performing the tasks) and affective (changes in attitude, beliefs, emotions and feelings) [18, 19]. Early attempts to measure and compare different modes of laboratory learning focused on cognitive learning [20] and over time more holistic approaches have been put forward to increase understanding [14]. However, more work on appreciating the types of learning occurring in the laboratory is needed.

In 2002, in response to advances in technology that could reshape learning, a three-day colloquy was held to form a consensus of learning objectives associated with the laboratory. This led to the formation of thirteen laboratory learning objectives covering instrumentation, modelling, experimentation, data analysis, design, learning from failure, creativity, psychomotor, safety, communication, teamwork, ethics, and sensory awareness [14]. The learning objectives sit above the specific course learning outcomes and allow for a holistic comparison of different modes of learning. The interconnection between laboratory objectives and learning outcomes is well explained in Reck [21].

Salim, et al. [19] categorized the thirteen laboratory learning objectives [22] across the cognitive, psychomotor and affective domains creating an instrument called Measuring the Learning Outcomes of Laboratory Work (MeLOLW). The MeLOLW instrument is important because it takes a holistic approach to evaluate the success of a laboratory experience. The instrument works by students self-evaluating their level of learning on a given scale across twenty-three statements. Limitations of the instrument include its reliance on perceived learning and moulding to an electronics laboratory, but it showcases the benefit of not only measuring knowledge, but also the growth in learning across the psychomotor and affective domains. This is valuable because employers seek graduates with a holistic set of skills [23, 24] and active learning, such as in the laboratory, has been shown to develop skills such as communication, learning from error, motivation, reflection, retention and teamwork [25, 26]; skills that written assessments are unable to effectively measure alone.

 Simply relying on course learning outcomes can limit the types of learning being measured and can limit the effectiveness of comparing two different modes of learning. For example, Lucke, et al. [27] compared a traditional and flipped classroom and found that the motivation and engagement increased but assessment scores remained the same. This data applied to a typical statistical analysis would show no relation between the student evaluation and teaching effectiveness. However, the improvement in the way students engaged and were motivated with the material can be argued as learning in the affective domain. Similarly, mismatches can be found between stated learning outcomes and student assessment [28]. Learning may be occurring just not necessarily as defined by the learning outcomes or as measured by assessment, supporting the need for a holistic instrument.

The literature above showcased the complexity of measuring learning and why a holistic laboratory learning approach is beneficial to recognize the diverse forms of learning that may occur in the laboratory, intentionally or unintentionally. More research is needed to understand why students evaluate the way they do, especially any relationship to learning.

III. Background

This work builds upon several studies such as [29-34] by the authors examining how changes and improvements to experiments, quality processes and teaching can influence student evaluation scores. The studies combined provided some evidence that students can measure a quality laboratory experience and provide a solid foundation of recommendations for educators to develop laboratory experiences with high student satisfaction. However, a major limitation within these works has been acknowledged by the authors and others in the community [35]. This major limitation is the lack of understanding of the relationships between the ways the students evaluate and how much they learn. While highly enjoyable learning experiences are beneficial, as educators a key focus remains on facilitating learning. A key risk is using
student evaluation data to guide changes to experiments which may increase the customer driven experience but have unseen long-term impact on learning. The authors own work [31, 32, 36] and that of others [16, 17, 37] provide evidence that improvements in student satisfaction can be correlated with an improvement in final grades. Such work provides a probable, but weak and inconclusive link to improvement in learning.

In an attempt to gain stronger evidence, a pilot study [38] was undertaken that involved exploring the relationships with both actual (via a laboratory exam) and perceived learning (by comparing before and after survey data based on their perceived knowledge at the time). The preliminary analysis showed that relationships could be found for both types of learning, but several limitations were identified that made measuring actual learning across the three domains problematic and unreliable (e.g. exam only measures achievement) and will be reconsidered for a following study. The authors are considering a different and more accurate approach to measure actual learning, with more details provided in section VI. Alternative methods of laboratory assessment may need to be investigated [39].

The pilot provided justification that it was possible to move forward with exploring the student evaluation relationships with perceived learning across the three learning domains. Knowing if students’ factor in their own thoughts on how they perceive their learning growth, with the way they evaluate the laboratory experience, is still very important. If such relationships can be found it provides greater certainty into the messages they are providing about their studies, especially when placed into the context of the author’s previous work.

IV. CONTEXT AND METHOD

This study was conducted across two universities across two continents. The study initiated within the School of Electrical, Computer and Telecommunications Engineering (SECTE) at the University of Wollongong (UOW), Australia. The study was then replicated at the School of Electrical Engineering (SEE) at the University of Belgrade (UB), Serbia. The UB was selected due to differences in culture, students, teaching and delivery providing an opportunity to compare findings.

The first author spent three months at UB before data collection commenced to observe teaching methods and assess feasibility. The greatest differences were observed in student gender balance (UB was balanced while UOW was predominately male); available resources (UB had great resources but limited in number and diversity of equipment compared to UOW); and in policy (formal policies drove direction and quality to a greater degree at UOW). Delivery of content was similar, but UB staff had a greater tendency towards more traditional teaching approaches including greater use of the white/black board. The importance of laboratory learning was equal at both universities, but due to resources, simulation played a more important role at UB. More focussed information is provided in sub-sections A and B below.

The analysis of teaching in this study concentrates only on teaching assistants called laboratory demonstrators, as they are the main source of face to face contact in the laboratory. For the Australian data, substantial analysis was conducted in previous work [38, 40] to understand possible bias and influences, providing confidence in the analysis being undertaken.

A. Overview SECTE UOW

The majority of SECTE courses follow a two-hour lecture and one-hour weekly tutorial followed by a three-hour fortnightly lab structure. Students are engaged with a diverse range of engineering laboratory setups; equipment, software and pedagogical approaches. Students regularly engage in a range of student evaluations, both formal and informal.

Students self-enrolled into laboratory classes for each course. Class sizes were determined by availability of equipment or workspaces. This study is limited to laboratory classes with either one or two laboratory demonstrators. One demonstrator is used for class sizes up to 20 students (average 15) and two demonstrators are used for class sizes up to 39 students (average 35). All laboratory demonstrators are Masters or PhD research students that undergo competitive recruitment and undertake in-depth training and received feedback as outlined in [36]. For each laboratory class, students had the same demonstrator for every experiment (between 5 and 12 experiments in each course). A total of 19 different demonstrators were used in the study with some teaching across multiple courses. Cases included one demonstrator teaching multiple classes in one course and/or laboratory classes across several courses.

B. SEE UB

Most SEE-UB courses follow a two-hour lecture, two-hour classroom exercise and 10 to 40 hours of lab exercises per semester organized weekly or in block session depending on the course. Laboratory sessions can last from 1 to 4 hours, depending on the course, and are usually organized several times per semester.

Students are automatically enrolled into laboratory classes. Class sizes were determined by availability of equipment or workspaces. This study targeted laboratory classes with two laboratory demonstrators. Two demonstrators are used for class sizes up to 40 students (average 25). All laboratory demonstrators are PhD, masters or undergraduate students that are at least one year ahead with their studies compared to the year in which the course is taught, they have high average grade during their studies and they have the highest grade in a particular course and related courses they are teaching. They do not undergo any formal training program. For each laboratory class, students had the same demonstrator(s) for every experiment (between 3 and 8 experiments in each course). A total of 10 different demonstrators were used in the study with some teaching across multiple courses. Cases include one demonstrator teaching multiple classes in one course and/or laboratory classes across several courses.

C. Student Evaluations

The laboratory student evaluations focussed on measuring the quality of the experiments (EXP), facilities (FACIL) and demonstrators (DEM). The evaluation questions were provided to students during the last experiment. The makeup of the student evaluations provided to the students are outlined below.
Full details of the validation of the survey instruments can be found in the following papers [29, 30, 40]. The Cronbach’s alpha coefficients were 0.85 (EXP), 0.82 (FACIL), 0.99 (DEM) and are all above 0.70, a common cut-off value for validity.

1) Teaching Quality: The teaching quality evaluation questions are based on the research conducted by Nikolic, et al. [30] on graduate teaching assistants to improve the laboratory learning experience. The questions focus on communication, preparation, interest and helpfulness. An example question included “is the casual demonstrator helpful in responding to questions or problems?”.

In this study a laboratory can have one or two demonstrators depending on class size as outlined earlier. The evaluation of teaching is comprised by grouping the responses and converting them to a score, confirmed via a factor analysis following the process detailed in [40]:

2) Experiment & Facilities: The evaluation questions focused on measuring the quality of the experiment and facilities are based on the research and process conducted by Nikolic, et al. [29] to improve the quality of the laboratory experience. The statements used to evaluate the quality of the experiments were also grouped and converted to a score referred to as EXP. The statements used to evaluate the quality of the facilities were also grouped and converted to a score referred to as FACIL. The groupings were also confirmed via a factor analysis. An example question included “the experiments undertaken in this laboratory are worthwhile learning experiences”.

D. Laboratory Learning Outcomes Measurement (LLOM)

The MeLOLW instrument [19] was modified and tested in a previous study [38] to enable universal application in any laboratory, used to develop a holistic understanding of the laboratory learning objectives across the cognitive, psychomotor and affective domains. This modified instrument LLOM, allowed measures to be reframed according to course content (by adjusting the italics text) and could be applied to any laboratory environment including hardware, simulation, programming, mixed, virtual and remote. While the measures were reframed (validated by staff and from the pilot), the subtext remained consistent with the MeLOLW instrument. Key words are changed in the template to represent learning content for each individual course, compared to the static statements in MeLOLW; making it easier to distinguish the differences in learning across a greater range of courses. Additionally, some of the questions have a hardware (H) and software (S) option to provide relevance across different courses. Therefore, to provide flexibility across different courses, the template must be adjusted. The universal design of this instrument will allow other education researchers the ability to explore the impact of changes between different laboratory modes and experiments in conjunction with measuring changes to course learning outcomes.

As outlined in the original work of the MeLOLW instrument and others [19, 41], the cognitive, psychomotor and affective learning domains cannot be isolated from each other because almost all learning activities involve more than one domain. The domain groupings were made to be consistent with the original work used to develop MeLOLW. Further analysis was undertaken to confirm the reliability and factors to determine how the statements should be analysed with this information presented in sub-section F.

The LLOM instrument was provided to students at the start of the first experiment and at the end of the last experiment. The initial data collection was to allow for a benchmarking of students perceived capability. The second data collection was used to determine how the student perceived those same capabilities after completing all the experiments. The difference between the two datasets was used to determine if perceived learning had occurred.

The LLOM instrument required students to think about the laboratory skills listed below. They needed to think about their current ability (in terms of the skill set required to undertake the learning activities and use the facilities associated with a particular course) and rate themselves according to the following scale: 5 – I am extremely confident; 4 – I am confident; 3 – I believe that I could, but might need to refer to my notes or other help; 2 - I could attempt it, but I most probably would get it wrong and not know what to do; 1 - I have heard about this, but I don’t know how to apply this; and 0 – I have no idea at all. The LLOM attributes listed was as follows with italics text indicating where course specific adjustments are made:

1. Cognitive
   1 - Understand the operation of equipment/software used within the laboratory?
   2 - Design experiments/models (physical or simulation) to verify course concepts?
   3 - Use engineering tools (e.g. [name of hardware/software used]) to solve problems?
   4 - Read and understand datasheets/circuit-diagrams/procedures/user-manuals/help-menus?
   5 - Draw & interpret relevant charts, graphs, tables & signals?
   6 - Recognise safety issues associated with laboratory experimentation?
   7 - Analyse the results from an experiment?
   8 - Write a conclusion summarising your findings from an experiment?
   9 - Write a laboratory report/entry into a logbook in a professional manner?

2. Psychomotor
   1 - Correctly conduct an experiment on [course equipment/software name - e.g. power systems]?
   2H - Select and connect appropriate instruments for the input, output and measurement of your circuit?
   2S - Select appropriate commands and navigate interface to simulate/program a model?
   3 - Plan and execute experimental work related to this course?
   4 - Construct/Code a working circuit/simulation/program?
   5 - Interpret sounds, temperature, smells and visual cues to diagnose faults/errors?
   6H - Operate electrical instruments (e.g. [equipment name]) required for experimentation?
   6S - Operate software packages (e.g. [software name]) required for coding/simulation?
7 - Take the reading of the output from circuits/ instruments /simulations/ programs?

3. Affective

1 - Work in a team to conduct experiments, diagnose problems and analyse results?
2 - Communicate laboratory setup, fault diagnosis, readings and findings with others?
3 - Work independently to conduct experiments, diagnose problems and analyse results?
4 - Consider ethical issues in laboratory experimentation and communication of discoveries?
5 - Creatively use software/hardware to design or modify an experiment to solve a problem?
6 - Learn from failure (when experiment/simulation/code fails or results are unexpected)?
7 - Motivate yourself to complete experiments and learn from the laboratory activities

E. Data Collection

The LLOM instrument was distributed to UOW students across 2015 and 2017 and UB students between 2017 and 2018. At the start of the first laboratory session for each course the LLOM self-assessment was undertaken. Students that consented to engage with the research after being provided with ethics guidelines were invited to include their student number for matching purposes. During the last laboratory session, the laboratory surveys were conducted and the LLOM self-assessment activity repeated.

Participation in the study was voluntary. At UOW 520 students were provided the opportunity to participate across 11 different courses (25 laboratory classes) with a total of 344 students completing all components. At UB 295 students were provided the opportunity to participate across 6 different courses (15 laboratory classes) with a total of 181 students completing all components. Course details are found in Table 1.

The data for the self-assessments, student evaluations and laboratory exam were matched using the student number at UOW and de-identified for analysis, while a unique anonymous ID was used at the UB and then the responses were matched for analysis. This was all done via an assistant external to the study to ensure all responses remained anonymous.

The platform R [42] and the R package lme4 [43] was used for the statistical analysis using a multi-level model using the hierarchical data with levels students, laboratory classes and courses. The estimated effects of the learning domains on teaching evaluations and p-values of the multilevel model will be presented.

F. LLOM Analysis

The data collected was crosschecked against the initial findings [38] of the development of the LLOM instrument. The new UOW data and the UB data reconfirmed reliability of the instrument with Cronbach’s alpha coefficients above 0.83, greater than the required 0.70, as shown in Table II.

An exploratory factor analysis was conducted to assess the underlying number of factors, and a principal component analysis to determine the underlying components. The number of factors within each learning domain was examined to determine how the statements were needed to be grouped in order to be correctly analyzed.

Four different checks were used to confirm the number of factors: Kaiser rule, parallel analysis, optimal coordinates, and acceleration factor. As per the initial findings [38] the results of underlying factors indicate that the psychomotor and affective domains only had one factor and the cognitive domain had two. Therefore, measures under the psychomotor domain are grouped together and are referred to as PSYCH and the affective measures under the affective domain are grouped and referred to as AFFECT. A principle component analysis was undertaken on the cognitive domain to determine the components indicating that measures eight and nine (concentrating on writing) for the cognitive domain are separate to measures one to seven (analytical skills). Therefore, measures one to seven are grouped and referred to as COG-A and measures eight and nine are grouped and referred to as COG-W. Based on these grouping, the final variables are defined as the average scores within each group. For the final analysis linear regression was used to investigate the relationship between perceived learning and student evaluation scores, using the linear mixed model approach in the R package lme4 [43] to account for dependence within classes and courses.

---

**TABLE I**

<table>
<thead>
<tr>
<th>Course</th>
<th>Content Area</th>
<th>Laboratory</th>
</tr>
</thead>
<tbody>
<tr>
<td>UOW ENGG104</td>
<td>Electrical Systems</td>
<td>Circuits + Multisim</td>
</tr>
<tr>
<td>UOW ECTE233</td>
<td>Digital Hardware</td>
<td>Circuits + Multisim</td>
</tr>
<tr>
<td>UOW ECTE301</td>
<td>Signal Processing</td>
<td>Matlab</td>
</tr>
<tr>
<td>UOW ECTE344</td>
<td>Control</td>
<td>HiLink + Matlab</td>
</tr>
<tr>
<td>UOW ECTE363</td>
<td>Communications</td>
<td>TIMS</td>
</tr>
<tr>
<td>UOW ECTE423</td>
<td>Power Systems</td>
<td>Labvolt</td>
</tr>
<tr>
<td>UOW ECTE465</td>
<td>Wireless Comms</td>
<td>Multiple Simulators</td>
</tr>
<tr>
<td>UOW ECTE823</td>
<td>Power Systems</td>
<td>Labvolt</td>
</tr>
<tr>
<td>UOW ECTE865</td>
<td>Wireless Comms</td>
<td>Multiple Simulators</td>
</tr>
<tr>
<td>UOW ECTE923</td>
<td>Power Systems</td>
<td>Labvolt</td>
</tr>
<tr>
<td>UOW ECTE965</td>
<td>Wireless Comms</td>
<td>Multiple Simulators</td>
</tr>
<tr>
<td>UB 13E053HP</td>
<td>Hydraulic Control Systems</td>
<td>Matlab/Simulink</td>
</tr>
<tr>
<td>UB 13E053RA</td>
<td>Robotics &amp; Automation</td>
<td>Robot Simulators (ABB &amp; Denso) + Matlab</td>
</tr>
<tr>
<td>UB 13E054CNC</td>
<td>Computer Numerical Control Machines &amp; Flexible Automation</td>
<td>Industrial robots and CNC machines; SolidCam</td>
</tr>
<tr>
<td>UB 13E112OR</td>
<td>Computer Organization</td>
<td>SPECS Simulator</td>
</tr>
<tr>
<td>UB 13E113OR</td>
<td>Computer Architecture and Organization</td>
<td>SPECS Simulator</td>
</tr>
<tr>
<td>UB 13E054RS</td>
<td>Theory of Robotic Systems</td>
<td>Robot TurtleBot III + Gazebo &amp; ROS</td>
</tr>
</tbody>
</table>

---

**TABLE II**

<table>
<thead>
<tr>
<th>Learning Domain</th>
<th>MelOLW [19]</th>
<th>LLOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
<td>0.90</td>
<td>0.83</td>
</tr>
<tr>
<td>Psychomotor</td>
<td>0.85</td>
<td>0.87</td>
</tr>
<tr>
<td>Affective</td>
<td>0.77</td>
<td>0.88</td>
</tr>
</tbody>
</table>
TABLE III
Relationship between student evaluations and perceived learning showing the effect of one score increase of each learning domain compared to EXP, FACIL and DEM. Shading shows commonality

<table>
<thead>
<tr>
<th>Wollongong</th>
<th>COG-A</th>
<th>COG-W</th>
<th>PSYCH</th>
<th>AFFECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXP</td>
<td>3.5590 ***</td>
<td>4.3689 ***</td>
<td>3.2158 ***</td>
<td>3.2313 **</td>
</tr>
<tr>
<td>FACIL</td>
<td>3.0830 ***</td>
<td>3.4615 ***</td>
<td>1.8855 *</td>
<td>3.0977 ***</td>
</tr>
<tr>
<td>DEM</td>
<td>2.9810 **</td>
<td>3.1915 ***</td>
<td>2.6391 **</td>
<td>1.5838</td>
</tr>
<tr>
<td>Belgrade</td>
<td>COG-A</td>
<td>COG-W</td>
<td>PSYCH</td>
<td>AFFECT</td>
</tr>
<tr>
<td>EXP</td>
<td>6.1150 ***</td>
<td>4.6370 ***</td>
<td>3.7840 ***</td>
<td>2.7730</td>
</tr>
<tr>
<td>FACIL</td>
<td>3.4478 **</td>
<td>1.8194</td>
<td>1.6275</td>
<td>0.7522</td>
</tr>
<tr>
<td>DEM</td>
<td>1.8850 *</td>
<td>1.5436</td>
<td>0.9722</td>
<td>1.2580</td>
</tr>
</tbody>
</table>

*p-value < 5%  **p-value < 1%  ***p-value < .01%

TABLE IV
Multi-level analysis to estimate mean before and after LLOM scores and perceived improvement in learning

<table>
<thead>
<tr>
<th>Wollongong</th>
<th>First LLOM Survey</th>
<th>Second LLOM Survey</th>
<th>Improvement (standard error) significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COG-A</td>
<td>COG-W</td>
<td>PSYCH</td>
</tr>
<tr>
<td>First LLOM Survey</td>
<td>3.172</td>
<td>3.432</td>
<td>3.119</td>
</tr>
<tr>
<td>Second LLOM Survey</td>
<td>3.893</td>
<td>3.843</td>
<td>3.823</td>
</tr>
<tr>
<td>Improvement (standard error) significance</td>
<td>0.721 (0.045) ***</td>
<td>0.411 (0.053) ***</td>
<td>0.704 (0.052) ***</td>
</tr>
<tr>
<td>Belgrade</td>
<td>COG-A</td>
<td>COG-W</td>
<td>PSYCH</td>
</tr>
<tr>
<td>First LLOM Survey</td>
<td>3.362</td>
<td>3.345</td>
<td>3.139</td>
</tr>
<tr>
<td>Second LLOM Survey</td>
<td>3.971</td>
<td>3.931</td>
<td>3.916</td>
</tr>
<tr>
<td>Improvement (standard error) significance</td>
<td>0.609 (0.083) ***</td>
<td>0.586 (0.102) ***</td>
<td>0.777 (0.101) ***</td>
</tr>
</tbody>
</table>

*p-value < 5%  **p-value < 1%  ***p-value < .01%

V. RESULTS

The student evaluation scores from the teaching (DEM), experiment (EXP) and facilities (FACIL) questions were compared against the differences in self-assessment scores (before and after) using the LLOM instrument across the cognitive (COG-A and COG-W), psychomotor (PSYCH) and affective (AFFECT) domains. Perceived learning is the difference in value for each statement between the two surveys.

The data is presented in Table III as the relationship between variables (the effect of one score increase of each learning domain compared to EXP, FACIL and DEM) with indicator of significance. For example, in Table III the relationship between EXP (student evaluation scores of the laboratory experiment) and COG-A (measures 1-7 in the cognitive domain) for UOW is represented by a slope of 3.559 meaning that for every one score increase in the COG-A domain the EXP score is expected to increase by 3.559. A negative value would mean that the score decreases. Shaded squares represent commonality across universities. Table IV provides a multi-level analysis using the hierarchical data with levels students, laboratory classes and courses to estimate the mean for each survey (first and second LLOM) and the improvement (difference between), including standard errors and significance. Significant is shown in

indicative ranges of p-value less than 5%, 1% and .01%.

VI. DISCUSSION

The UOW data indicates statistically significant relationships exist in all cases except for the relationship between the evaluation of the demonstrators and perceived learning in the affective domain. This suggests that there is a strong multi-domain perceived learning relationship influencing student evaluation behavior. In contrast, the UB data indicates that the strong relationship is not universal. Significant relationships only hold true across the cognitive-analytic domain for all evaluations and only the cognitive-writing and psychomotor domain for evaluating the quality of the experiments.

Across the common elements shown with shading, the coefficients for each relationship is very similar except at UB with EXP and COG-A showing a coefficient almost twice in size. The authors hypothesize that this increase may be an offset to the lack of significance in other measures based on the differences in laboratory experiences. At UOW the laboratory facilities are substantial and diverse while at UB they are limited and maximized to the most efficient use in terms of space and learning with a greater focus on simulation [32, 44], reflecting the funding available at each university. This limitation of equipment, setups, experimentation and hands on experimentation at UB might explain the lack of significance and lower coefficient across the other domains. For example, there is less comparison (range of laboratories for students to compare against) to be made across UB laboratories compared to at UOW. The differences in evaluating the demonstrators may also be linked to cultural differences, experience, or exposure in evaluating teachers.

Table IV attempts to quantify the amount of perceived learning taking place within the laboratory across the three domains by looking at the multi-level estimated mean of improvement across the two surveys. The data indicates that a statistically significant perceived learning improvement is occurring across all domains, across both universities. This showcases the multi-faceted learning that is occurring in the laboratory, highlighting the important role that it can play in engineering education. This highlights the need for more holistic measures to appreciate laboratory learning beyond the cognitive [20]. Across Tables III and IV for UOW, the data shows that the students think they are learning across all domains, and this is indeed a factor in their evaluation choices. For UB on the other hand, while the students think they are learning across all three domains, the students are not demonstrating this across all domains in the evaluations. For example, there was no statistically significant evaluation relationship in the affective domain, while Table IV shows that students are perceiving an improvement in affective skills is taking place.

As outlined in Section IV substantial differences can be found between running a teaching laboratory between UOW and UB. Beyond those differences the results show that across universities some strong statistically significant relationships exist between perceived learning and how students evaluate their laboratory experience. This is important as it reveals
evidence that shows students are evaluating beyond a customer-centric mindset [11]; and perceived learning influences their decisions, supporting previous findings in other studies [12, 13]. This provides the community interested in improving laboratory experiences a guideline as to how such evaluation data can be interpreted. If a change is made and an improvement in evaluation scores is found, we can now assume that in part the students at least perceived learning had improved. This would strengthen the evidence of the educational benefits associated with improvements made in many studies such as in [16, 17, 45] and links between evaluations and learning [46]. However, the data indicates that these relationships can differ across institutions, countries and cultures meaning institution-based analysis would be needed to best understand their student cohort.

Most studies exploring SET and learning relationships focus on correlations across an entire course, exploring the relationship between the instructor and final exam [10]. Many parameters can influence such a scenario such as teaching assistants, demonstrators and support networks. This also assumes that the final exam covers all learning being achieved. By focusing this study on only one component, the laboratory, the number of unaccountable possible influences are reduced because each course had consistent factors such as assessment, structure, experiment, and facilities. Secondly, by undertaking a multi-level statistical approach, analysis could be undertaken across laboratory classes of the same course as well as between courses and factor in variables such as a demonstrator teaching repeated classes in the same course and across different courses. Thirdly, by using LLOM perceived learning was not only measuring learning as expected by the course coordinator, but across a comprehensive list covering attributes across the cognitive, psychomotor and affective domains. The data collected provides an opportunity for studies as outlined in [10] to change direction and reconsider other possible learning relationships.

The findings intertwine and build upon multiple studies conducted by the authors. Over a decade of surveys, ethnography, data analysis and reflective practice has been conducted with the aim to maximize the laboratory learning experience. Each contribution has unlocked a greater understanding of the student experience, how they evaluate, and what they learn. In [30] it was found that training and professionally managing laboratory demonstrators lifted evaluation scores. In [29] the key ingredients of how to lift evaluation scores through changes to experiments, instructions, equipment and facilities was discovered. In [40] the statistical relationship between laboratory demonstrator teaching performance with the experiments, equipment and facilities was established; allowing the impact of teaching staff to be understood. In [33, 34, 36] the importance of scaffolding and developing supporting resources was realized, helping students move forward with independence. Collectively, the evidence strongly indicates that students do recognize and respond positively to quality improvements. However, across these studies a key missing element was understanding the impact on learning, the key motivator beyond student satisfaction (making students feel good about their experience in the laboratory).

Several studies [31, 32] and most particularly [36] (included recorded observations by the laboratory demonstrator) tried to address this by correlating the improvements with actual learning via final grades. Links to improvement in actual learning performance were found, but inconclusive as final grades don’t tell a complete picture about laboratory learning. In a pursuit to find a stronger connection, work continued in [38] which correlated the student evaluations with laboratory exam performance. Correlations were made between student evaluations and laboratory exam performance, but exam performance did not tell much about how students actual learning improved. Additionally, logistical challenges and difficulty in collecting a holistic snapshot of learning required the authors to rethink their approach to measuring actual learning. Together the findings across these studies did provide some evidence that students on average do reflect actual learning into their student evaluation decisions. However, more work was needed to strengthen this evidence as it was still not conclusive. This study has taken the next step forward by showing that students are perceiving increases in their learning (and across multiple domains). This is an important discovery within student evaluation research. By strengthening the links between student satisfaction and perceived learning, laboratory-based studies such as by the authors [31, 32, 36] and others [16, 17, 37], can appreciate a greater contribution between the improvements made and the impact that it made to learning.

The final step to close the loop is to correlate the student evaluations with improvements in actual learning by undertaking some form of laboratory assessment before the first experiment and comparing learning improvements by repeating the assessment after the final experiment. This would be one of the most accurate methods to understand student learning across all laboratory objectives and learning domains and explore correlations between evaluations and learning. This will provide insights as to how actual learning can be measured more holistically, improving laboratory assessment methods. As reported in [38] there are substantial hurdles to achieve this, but the authors are making plans.

This work also has its limitations. Firstly, this work was only conducted at two institutions and with limited courses, influenced by the complexity of data collection, student self-selection in participation, the time involved to collect the data and meeting ethics requirements. The authors hope that other educational researchers can contribute by expanding on this study to help increase diversity and sample size. Secondly, while it is encouraging to know that perceived learning is influential, greater benefit could be gained by understanding actual learning by overcoming the barriers faced in [38].

VII. CONCLUSION

Understanding the relationships between evaluations and learning is difficult, hence the plethora of studies. With no perfect approach found, this study has contributed to research by exploring the relationships between student evaluations in the laboratory and perceived learning, expanding the concept of
learning across the cognitive, psychomotor and affective domains using the LLOM instrument. Significant relationships between the variables were found, but not uniformly across universities. The joint relationship was strongest with perceived learning in the cognitive-analytical domain for all evaluations, and for experiment quality across the cognitive and psychomotor domains. This shows that perceived learning is influential in student evaluations building upon the authors [29, 30, 34, 36, 40] and others [12, 13, 19] work in this area.

The relationships between evaluation scores and learning was stronger at UOW compared to UB. The author’s hypothesized that limitations in laboratory facilities, activities and exposure to evaluations at UB may have been a contributing factor to a lower number of significant relationships and is an area of further research. These findings may explain why there is so much conflicting evidence in relation to student evaluations in the literature, in that groups of students separated by location, experiences, opportunities and cultures may simply value things differently.

The universal design of the LLOM instrument will allow staff to develop a greater understanding of the learning occurring within their laboratories beyond course objectives. It will also allow education researchers the ability to more accurately explore the impact of changes between different laboratory modes and experiments. For example, a more holistic viewpoint on learning can be measured when comparing physical, virtual or remote laboratories. LLOM was not designed to create a requirement that every course must enforce every self-assessment question asked in the instrument; this can happen (and confirmed with LLOM) across the entire degree. Instead it should be used to reflect on how changes made to courses experiments influence changes to students perceived learning. That is, what changes in learning do students think are occurring, both intentional or not, caused by changing the leaning mode or experiment. From experience it has been observed that teaching staff may not consider the impact of change across all possible learning objectives, with their focus being just on course objectives.

This study expands research on the relationship between student evaluations and teaching effectiveness by emphasizing the need to broaden the concept of measuring learning. It was evident that perceived learning increased across all domains highlighting the need to look beyond valuing laboratory learning only at the cognitive domain, including both psychomotor and affective learning into the equation. That is, while course learning outcomes may be measured correctly there could be other changes to learning occurring that are not considered. With the assessment of most engineering courses focused on large final written exams, quizzes and tutorials focused on cognitive learning, this work suggests that learning in a holistic form may not be measured in its entirety, questioning the finding of many statistical studies. The LLOM instrument also open the opportunity for other researchers to investigate new instruments to measure learning outcomes across the three domains for other teaching experiences such as the lecture, tutorial and workshop. This would continue to help our understanding of the types of learning occurring in the classroom and the impact of innovation. Considering this information, developing laboratory assessment methods that attempt to measure learning beyond the cognitive domain would be helpful. The authors are making plans to complete their loop of research by measuring actual learning across the three domains, possibly leading to new or enhancing existing laboratory assessment methods.

REFERENCES


Sasha Nikolic (M’14–SM’15) received the B.E. degree in telecommunications and a PhD in engineering education both from the University of Wollongong, Wollongong, Australia, in 2001 and 2017 respectively. He is a Senior Lecturer in Engineering and Information Sciences Education at the University of Wollongong, where he is also a Fellow of the Wollongong Academy of Tertiary Teaching and Learning Excellence. He previously worked as a Laboratory Manager at the same university and spent several years in industry. His interest is in developing career ready graduates involving research in teaching laboratories, industry engagement, work integrated learning, knowledge management, communication and reflection.

Dr. Nikolic is a member of the executive committee of AAE and an Associate Editor for AJEE. He has been recognized with many awards including an Australian Award for University Teaching citation in 2012 and 2019. He is currently Past Chair of the IEEE NSW Section and Past Chair of the IEEE Education Society’s NSW Chapter. He also served on the IEEE Australia Council. He was Co-Chair of IEEE TALE2018 held in Wollongong.

Thomas Suesse completed his MSc (Dipl-Math) degree in mathematics at the Friedrich-Schiller-University (FSU) of Jena, Germany, in 2003. Dr Suesse then worked as a research fellow at the Institute of Medical Statistics, Informatics and Documentation (IMSID) and FSU. In 2005 he went to Victoria University of Wellington (VUW), New Zealand, to start his PhD in statistics and his degree was conferred with his thesis titled, ‘Analysis and Diagnostics of Categorical Variables with Multiple Outcomes’ in 2008. In 2009 he started working as a research fellow at the Centre for Statistical and Survey Methodology (CSSM) at the University of Wollongong.

He was appointed as a lecturer at UOW in 2011 and promoted to senior lecturer in 2015. His research interests comprise categorical data analysis, spatial statistics, social networks, survey methods, medicine, education, and early childhood education.

Kosta Jovanovic (M’18) received a PhD in electrical and computer engineering with focus on robotics from the University of Belgrade in 2016. He is an Assistant Professor and Head of ETF Robotics laboratory at the University of Belgrade – School of Electrical Engineering. He was guest researcher at Technical University of Munich (2010) and German Aerospace Center - DLR Institute of Robotics and Mechatronics (2013). His interest are physical human-robot collaboration, robot control, and education in robotics.

Asst. Prof. Jovanovic has served as a topic editor in Frontiers in Neuro Robotics and as a member of International Scientific Committee of Int. Conf. on Robotics in Alpe-Adria-Danube region (RAAD). He has been recognized with many awards including Award of American Chamber of Commerce in Serbia for Excellence in Serbian Society, Nikola Tesla Foundation, City of Belgrade Award, and the University of Belgrade Award.