EXPLORING THE DIFFERENCE OF STUDENTS' LEARNING OUTCOME EVALUATION BETWEEN INDUSTRIAL PROFESSIONALS AND STUDENTS

Edward Clay, College of Engineering, California State Polytechnic University, Pomona, CA 91768
909-749-7358, erclay@cpp.edu
Mankirat Singh, College of Engineering, California State Polytechnic University, Pomona, CA 91768
909-282-9025, mankirats@cpp.edu
Maninder Kaur, College of Engineering, California State Polytechnic University, Pomona, CA 91768
626-554-6182, mkaur@cpp.edu
Wen Cheng, College of Engineering, California State Polytechnic University, Pomona, CA 91768
909-869-2957, wcheng@cpp.edu
Yasser Salem, College of Engineering, California State Polytechnic University, Pomona, CA 91768
909-869-4312, ysalem@cpp.edu

ABSTRACT

Both professionals' assessment and students' self-assessment have been used to evaluate student's learning outcomes. Thus far, very few studies have been dedicated to the examination of the difference between them. This study aims to fill the gap by investigating the discrepancy in between. Various statistical tests were employed to compare the two types of evaluation responses collected from senior project assessments over the span of seven years. The study revealed that the industry professionals generally rated students higher than the students' self-evaluation of their own capabilities, and most students' performance and achievements were up to the industry professionals' expectations.

Keywords: Industrial professionals; Engineering students; Self-evaluation; two-sample t-test; student learning outcomes.

INTRODUCTION

Nowadays, most higher education institutions tend to regard student assessment as a learning strategy rather than a method of measuring learning [1]. Evaluation and the application of rubrics enable students to be well encouraged and motivated. The assessment is also an effective modus to evaluate student performance and main issues at the learning stage [2-4]. Assessment is an effective method to detect learning progress, achievements, and other information about students [5]. Student assessment surveys provide teachers with feedback on development opportunities, and they form the basis for tenure decisions and achieving the desired goals [6,7]. Based on educational psychology research, self-assessment is related to motivation, and students' motivation to learn is essential for lifelong learning [8,9]. Similarly, peer assessment is also a common means of assessing students' competencies. In peer assessment, students evaluate the abilities of other students but not their own, which is typically applied when evaluating projects and practical presentations [10]. Most students consider peer assessment to be a beneficial activity that helps their learning and motivation [1]. Likewise, studies have shown that when using global judgment or scoring academic products and processes rather than professional practice, peer assessment is very similar to teacher assessment [11]. A study done by [12] reveals the significant correlation between teacher evaluation and self-assessment scores [12]. The findings shown that teacher evaluation was more accurate than self-assessment.

In addition, there are also different types of traditional evaluation strategies, such as alternative assessment. Instead of selecting from the given specific options, alternative assessment prompts students to provide their original and innovative responses, potentially guiding students to express their feelings and emotions [13]. The assessment keeps students in a positive state throughout teaching, learning, and evaluation process, thereby helping them develop and evaluate their abilities. However, due to the responses' characteristics, the open-ended responses from students are difficult to be classified and used for statistical analysis. Therefore, the form of assessment used in this study will allow the students to select an assessment that accurately describes their abilities from a specific list of options.

Due to the lack of attention to learners’ sense of responsibility, many traditional assessment methods have become less popular [14,15]. This study emphasized professionals’ assessment and student self-evaluation among different students' abilities evaluation methods. Senior project is the best term project to inspect the students’ learning achievements in the entire undergraduate program. The authors believe that for these students who will soon obtain their undergraduate degree certificate, their self-evaluations and the evaluations from industry professionals are more important than other assessment methods.

Self-assessment provides students with valuable feedback about themselves and helps them with personal and professional development [16]. Students' self-assessment of their work and processes helps improve academic performance and self-regulation skills. It is suggested that students' self-assessment should no longer be regarded as an assessment but as the basic ability of self-regulation [17]. Furthermore, self-assessment can be defined as a “descriptive and evaluative act carried out by the student concerning his or her work and academic abilities [18]”.

Professionals' views on the effective assessment process involve their knowledge, skills, and abilities as the main targets of experienced educators and education leaders [19, 20]. It is one of the most critical aspects of the job for a teacher to assess student performance [21]. Assessment includes selecting the relevant elements of the education system, evaluate its quality, and giving weight to them. The choice of relevant features depends on the intended purpose of education, including social goals and personal goals [22]. Compared with non-trained evaluators, a higher percentage of trained evaluators consider themselves competent [23].

Self-assessment is sometimes considered to be the most difficult because the students feel it is impossible to assess their work [24]. Self-assessment has certain qualities, so its use as a formal assessment is limited [25]. Self-evaluation and lecturer-assessment are intended to better engage students in the learning process through evaluation [26,27]. Students in Hong Kong believe that lecturers are more knowledgeable in assessment, and lecturers also regard assessment as their duty [1]. Another study [28] has indicated that there is resistance to this shift from lecturer assessment to peer assessment by both staff and students for reasons such as the reliability and fairness of peer assessment and the increasing the workload of lecturers as the peer marks would have to be collated. Therefore, based on student self-evaluation, this study adds professionals to compare with to increase assessment rigor. Although many articles have researched professionals’ assessment and students’ self-assessment separately, very few articles examined their relationship. The primary purpose of this study is to discover the difference in student outcome evaluation between industry professionals and students.

In summary, this paper contributes to the field of higher education by providing a reference for educators and some industry professionals with ratings and evaluations on the graduation projects of civil engineering students, by comparing the significance of professionals' scoring of students' various engineering professional skills and the self-evaluation by students themselves. Explicitly, industry professionals review and evaluate 11 different student abilities in the survey data. The assessment includes students' application ability of engineering knowledge and tools, presentation skills, understanding of the project objectives, and ability to solve problems.

To compare the responses from professionals' assessment and student's self-evaluation, three types of t-test models were used in this study to examine the corresponding student's ability in the two groups. For each model, the dependent variable is the evaluation score for a specific student's abilities. The independent variable is the rating from professionals (P) or students (S). In these models, the classic two-sample t-test and Welch's t-test models are used to judge whether the mean score of professionals' assessment and students' self-evaluation are significantly different. Subsequently, the Wilcoxon signed-rank test model can be used to determine the median in the two groups. Then we can compare the results of these models and discover the relationship of the evaluations from industry professionals and students themselves.

DATA DESCRIPTION

In this project, both Senior Exit Survey and Senior Project Assessment data were used in analyzing and expounding viewpoints. A Senior Exit Survey was conducted by the Civil Engineering Department of California State Polytechnic University, Pomona, on civil engineering students who have completed their senior project at the last of their undergraduate programs and Senior Project Assessment data are each group of students' competency evaluation scored by advisors after students finished their senior project. The student survey includes comprehensive questions such as student names, status upon admission, length of study in CPP, campus involvement, employment history, licensing qualifications, self-assessment of student outcomes, etc. Table 1 below shows eleven questions selected from this survey data. Senior Project Assessment consists of different questions about student ability assessments, such as applying knowledge, ability to function on an interdisciplinary team, quality of presentation and communication, understanding contemporary issues, lifelong learning, etc. Likewise, eleven questions matching the student survey were chosen from the assessment and are identified in Table 2.

TABLE 1. QUESTOINS SELECTED FROM THE SENIOR EXIT SURVEY

Question ID	Description of Student Learning Outcomes
a	Ability to apply knowledge of mathematics, science, and engineering.
b	Ability to design and conduct civil engineering experiments, as well as to analyze and interpret data.
c	Ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability.
d	Ability to function on multidisciplinary teams.
e	Ability to identify, formulate, and solve engineering problems.
f	Understanding of professional and ethical responsibility.
g	Ability to communicate effectively.
h	Understanding of the impact of engineering solutions in a global, economic, environmental, and societal context.
i	Recognition of the need for, and an ability to engage in life-long learning.

j	Knowledge of contemporary issues and their importance to engineering systems.
k	Ability to use the techniques, skills, and modern engineering tools necessary for engineering practice.

TABLE 2. QUESTIONS SELECTED FROM THE SENIOR PROJECT ASSESSMENT

Question ID	Description of Student Learning Outcomes
1(a)	Ability to apply knowledge.
2(k)	Use of engineering techniques and tools.
3(e)	Ability to gather data and solve engineering problems.
4(c)	Ability to design a system.
5(h)	Understanding of outside constraints & contemporary.
6(g)	Quality of visual presentation.
7(g)	Quality of oral communication.
8(d)	Ability to function on an interdisciplinary team.
9(c)	Level of design experience.
10(i)	Ability to recognize the need for and be able to pursue lifelong learning.
11(j)	Awareness and understanding of contemporary issues and their interactions.

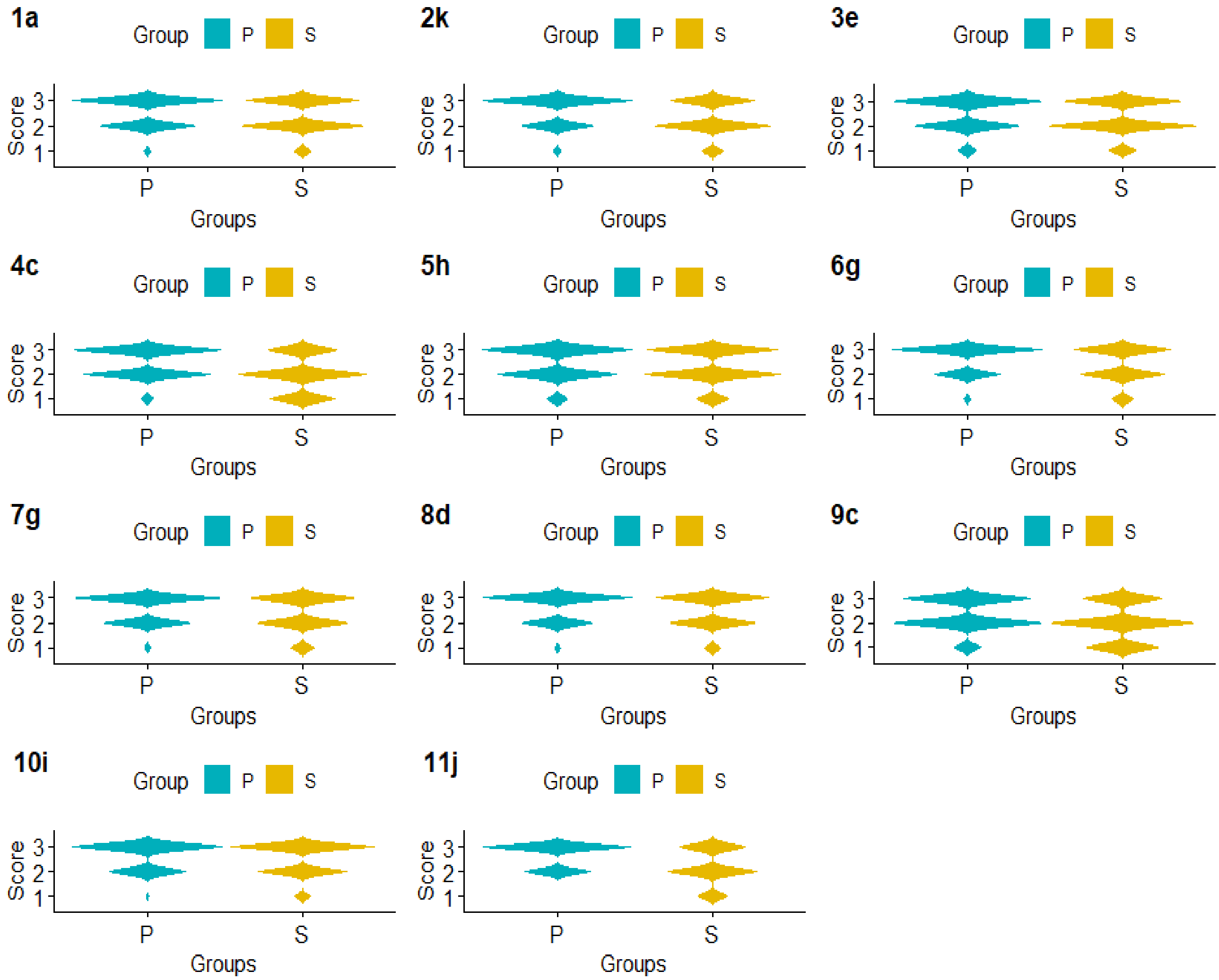
In the student survey, each result has four possible answers: (0) Not applicable; (1) Poor; (2) Moderate; (3) Excellent. After data-cleaning, which removed the missing or wrong data, 662 responses were finally retained for this research analysis. The survey aims to collect feedback from undergraduate students and allow students to make self-assessments on their 11 different professional competencies and skills. As for assessment, separate from the student survey, the result has only four possible answers: (0) Not applicable; (1) Poor; (2) Moderate; (3) Excellent. A total of 424 advisors evaluated groups of students over seven years, from 2013 to 2019. Table 3 lists the detailed categorization of responses. The preliminary investigation of this study is to match the capability evaluation questions from two surveys and find out the significant impact between them.

TABLE 3. DESCRTPTIVE STATISTICS OF ADJUSTED EVALUATION OUTCOMES

Question ID	Count and Percentage of Responses						
	Excellent (3)	Moderate (2)	Poor (1)	N/A (0)	Rating Average	Total Response Count	Valid Response Count
P (1a)	247 58.25%	162 38.21%	14 3.30%	1 0.24%	2.55	424	423
S (1a)	290 43.81%	322 48.64%	50 7.55%	0 0.00%	2.36	662	662
P (2k)	269 54.01%	133 31.37%	17 4.01%	5 1.18%	2.60	424	419
S (2k)	247 37.31%	342 51.66%	70 10.57%	3 0.45%	2.37	662	659
P (3e)	229 54.01%	162 38.21%	31 7.31%	2 0.47%	2.47	424	422
S (3e)	265 40.03%	332 50.15%	65 9.82%	0 0.00%	2.40	662	662
P (4c)	214 50.47%	182 42.92%	20 4.72%	8 1.89%	2.47	424	416
S (4c)	175 26.44%	321 48.19%	165 24.92%	1 0.15%	2.01	662	661

P (5h)	213 50.24%	169 39.86%	31 7.31%	11 2.59%	2.44	424	413
S (5h)	285 43.05%	297 44.86%	75 11.33%	5 0.76%	2.31	662	657
P (6g)	279 65.80%	125 29.48%	16 3.77%	4 0.94%	2.62	424	419
S (6g)	312 47.13%	272 41.09%	75 11.33%	3 0.45%	2.36	662	659
P (7g)	254 59.91%	151 35.61%	14 3.30%	5 1.18%	2.57	424	419
S (7g)	312 43.71%	272 41.09%	75 11.33%	3 0.45%	2.36	662	659
P (8d)	268 63.21%	130 30.66%	14 3.30%	12 2.83%	2.62	424	412
S (8d)	342 51.66%	266 40.18%	53 8.01%	1 0.15%	2.44	662	661
P (9c)	170 40.09%	200 47.17%	38 8.96%	16 3.77%	2.32	424	408
S (9c)	175 26.44%	321 48.49%	165 24.92%	1 0.15%	2.02	662	661
P (10i)	126 58.60%	65 30.23%	4 1.86%	20 9.30%	2.63	215	195
S (10i)	377 56.95%	235 35.50%	46 6.95%	4 0.60%	2.50	662	658
P (11j)	139 64.65%	65 30.23%	4 1.86%	20 9.30%	2.63	215	202
S (11j)	235 35.50%	311 46.98%	111 16.77%	5 0.76%	2.19	662	657

Notes: “P” represents the response from industry professionals; “S” denotes that the response came from the students’ self-evaluation.



Notes: "P" represents the response from industry professionals; "S" denotes that the response came from the students' self-evaluation.

FIGURE 1. DISTRIBUTION OF PROFESSIONAL AND STUDENT EVALUATION SCORING RESPONSES

The violin charts show the distributions of professionals' and students' responses to these eleven survey questions in Fig. 1. The replies from professionals are shown in blue, and student's responses are shown in yellow. In Fig. 1, the rectangle's width represents the proportion of people who selected the corresponding score. As a result, most professionals and students rated the abilities as moderate or excellent (2 or 3 points), and few people rated the abilities as poor (1 point). Professionals tend to give students good reviews. Excellent is the professionals' most popular option for ten of these eleven questions. Only the evaluation of the design experience level (9c) is different from the above situation, with moderate ratings as the choice of most professionals. Furthermore, the professionals evaluated the students immediately and three professionals evaluated the group of six to eight students. Overall, professionals generally give high marks to students' abilities. And for these eleven survey questions, the average ratings from professionals are higher than students' self-evaluation.

METHODS

This paper aims to find out the potential difference between industrial professionals' assessment and students' self-evaluation. Two separate evaluations against a series of students' engineering competencies who are about to finish their undergraduate program. For determining differences between the two groups, the t-test is a valuable method for estimating whether two random variables are independent. In this study, two different test methods (t-test and Wilcoxon test) will be used to determine the mean and median of the students' 11 different abilities assessment results. The unpaired two-sample t-test was used to determine the mean of the student ability score. The null hypothesis (H_0) is that there is no significant difference between the mean scores of professionals' assessment and the mean scores of students' self-evaluation. Meanwhile, the Wilcoxon test was used to determine the median. The significant difference between them can be determined by comparing the mean and median scores between professionals' assessment and students' self-evaluation. The null hypothesis (H_0) is that there is no significant difference between the median scores in professionals' evaluations and students' self-evaluation.

Fisher's F-test

Before doing the t-test to determine the mean scores between the two groups, whether the professionals' assessment and students' self-evaluation have the same variance need to be verified. Fisher's F-test for homogeneity of variances [29] was used in this study. The test statistic can be obtained by computing the ratio of the variances of students' self-evaluation (S_s^2) and professionals' assessment (S_p^2), the equation shows below [30].

$$F = \frac{S_s^2}{S_p^2} \tag{1}$$

By calculating the F-value, we can define whether the variance of students' self-evaluation scores was the same as the professionals' assessment scores. For each students' engineering professional ability to be tested, according to the variance is the same or not, the following two different t-test methods can be chosen.

Classic Two-Sample T-test

The classic two-sample t-test [31] was used to determine the difference of mean scores in the two groups if the variances of the two groups were equivalent. The t-test value can be calculated as follows [30].

$$t = \frac{m_s - m_p}{\sqrt{\frac{S^2}{n_s} + \frac{S^2}{n_p}}} \tag{2}$$

Where m_s and m_p represent the sample mean of students' self-evaluation scores and professionals' assessment scores, respectively. n_s and n_p represent the sample sizes of the two groups. S^2 is an estimator of the pooled variance of the two groups, which can be calculated as follow [30].

$$df = n_s + n_p - 2 \tag{3}$$

$$S^2 = \frac{\sum(x-m_s)^2 + \sum(x-m_p)^2}{n_s + n_p - 2} \tag{4}$$

Where df represents the degree of freedom. x represents the independent variable which is the evaluation score.

Welch’s Two-Sample T-test

Instead of using the pooled variance S^2 in the classic t-test. Welch’s t-test [32] involves the variance of each of the two groups S_s^2 and S_p^2, which are the standard deviation of the two groups of students’ self-evaluation scores and professionals’ assessment scores, respectively.

$$t = \frac{m_s - m_p}{\sqrt{\frac{S_s^2}{n_s} + \frac{S_p^2}{n_p}}} \tag{5}$$

The degree of freedom of the Welch t-test is estimated as follows.

$$df = \left(\frac{S_s^2}{n_s} + \frac{S_p^2}{n_p}\right) / \left(\frac{S_s^2}{n_s^2(n_s - 1)} + \frac{S_p^2}{n_p^2(n_p - 1)}\right) \tag{6}$$

After concluded the t-value and degree of freedom for each student's ability evaluation outcome, the t distribution reference table can be used to detect whether there is a statistically significant relationship between student self-evaluation and professionals’ assessment of the corresponding question in the survey. For the tested question with a p-value less than 0.05 obtained from the t distribution reference table, the result rejects the null hypothesis, which means at a 90% confidence level, the mean scores of professionals’ assessment and the mean scores of students' self-evaluation are statistically significantly different.

Wilcoxon Signed-rank Test

Wilcoxon test was used to determine the differences in the scores in professionals’ assessments and students’ self-evaluation by ranking method [33]. First, calculate and rank the distance between the two sets of sample values: the evaluation scores. To achieve this, the absolute value after subtracting the two sets of sample values can be calculated, and then sort these computed values. Secondly, exclude the pairs with a distance equal to 0, and record the remaining sample size. Thirdly, sort the remaining pairs from small to large and ranked the pairs. Finally, the test statistic W can be calculated using the following equation [34].

$$W = \sum_{i=1}^{N_r} [sgn(x_{p,i} - x_{s,i}) \cdot R_i] \tag{7}$$

Where N_r is the number of samples remaining after excluding the pairs with the calculated distance is zero. *sgn* represents the sign function which is a logical function used to determine the sign of a real number. R_i denotes the rank number.

Similarly, after calculated the W value, the corresponding p-value can be found against the reference table. The tested question with a p-value less than 0.05 indicates that the result rejects the null hypothesis. The median scores in professionals’ assessment and students’ self-evaluation are statistically significantly different.

RESULTS

A mentioned previously, potential difference between industrial professionals’ assessment and students’ self-evaluation. The statistical software ‘R’ was used for statistical modeling in this study. The results obtained by Fisher’s F-test, two-sample t-tests (Welch's two-sample t-test and classic two-sample t-test), and Wilcoxon signed-rank test are illustrated in Table4, Table 5, and Table 6, respectively.

Fisher's F-test for homogeneity of variances between the two groups includes the professionals' assessment score and students' self-evaluation score. The calculated F-value, p-value, and degree of freedom for each question were shown in Table 4. Based on the result of p-values, a total of 9 sets of questions rejected the null hypothesis, indicating that the variances of the two groups in these questions are not equal at the 90% confidence level. For these nine sets of questions, we used Welch's two-sample t-test the significance of professionals' assessment scores and students' self-evaluation. And for the remaining four sets of questions, the classic two-sample t-test was used to examine the significance of the two groups.

The outcomes of Welch's two-sample t-test and classic two-sample t-test were merged and displayed in Table 5. As a result, all 11 sets of questions have p-values less than 0.05. This means that for all student competence assessment questions in this survey, the professionals' assessments are significantly different from students' self-evaluation. Also, based on the response outcomes shown in Table 3 and Fig. 1, the sample mean of the professionals' assessments is greater than students' self-evaluation. Therefore, combining the p-value results and the average rating scores, we can safely conclude that the mean scores in professionals' assessment and students' self-evaluation are significantly different. The mean score of the former is higher than the latter.

Analogously, the outcomes of the Wilcoxon signed-rank test are shown in Table 6. All these 11 questions rejected the null hypothesis (H_0) with p-values less than 0.05, which demonstrates the relevant findings with two-sample t -test that the median scores in professionals' assessments and students' self-evaluation are significantly different.

TABLE 4. F-TEST FOR HOMOGENEITY IN VARIANCES

Question ID	1(a)	2(k)	3(e)	4(c)	5(h)	6(g)	7(g)	8(d)	9(c)	10(i)	11(j)
df (S)	661	658	661	660	656	658	658	6601	660	657	656
df (P)	422	418	421	415	412	419	418	411	407	194	201
F	1.218	1.273	1.027	1.489	1.123	1.475	1.469	1.334	1.268	1.411	2.162
p-value	**0.027**	**0.007**	0.764	**1.07e-05**	0.197	**1.63e-05**	**2.02e-05**	**1.41e-03**	**0.009**	**0.004**	**3.84e-10**

Notes: 1.Statistically significant variables with a p-value less than 0.05 are bolded. 2. "df (S)" and "df (P)" indicate the degree of freedom of students' responses and professionals' responses, respectively. 3. F values are derived from Equation (1).

TABLE 5. TWO-SAMPLE T-VALUE OUTCOMES

Question ID	1(a)	2(k)	3(e)	4(c)	5(h)	6(g)	7(g)	8(d)	9(c)	10(i)	11(j)
df	962.22	966.39	1082	1003.2	1068	1009.1	1005.9	963.71	938.75	371.03	490.59
t	-5.18	-8.93	-4.22	-11.24	-2.95	-7.03	-5.61	-4.87	-7.32	-2.73	-11.42
p-value	**2.70e-07**	**<2e-16**	**2.62e-05**	**<2e-16**	**3.27e-03**	**3.72e-12**	**2.5e-08**	**1.32e-06**	**5.28e-13**	**0.007**	**<2e-16**

Notes: 1. Statistically significant variables with a p-value less than 0.05 are bolded. 2. T values are derived from Equations (2) and (5). 3."df" stands for degrees of freedom.

TABLE 6. WILCOXN TEST P-VALUE OUTCOMES

Question ID	1(a)	2(k)	3(e)	4(c)	5(h)	6(g)	7(g)	8(d)	9(c)	10(i)	11(j)
W	162231	176698	159685	183620	148523	167312	160148	155871	165730	69877	91569

p-value	**6.8e-07**	**<2e-16**	**9.78e-06**	**<2e-16**	**0.0038**	**6.19e-11**	**6.57e-07**	**5.09e-06**	**7.05e-12**	**0.0289**	**<2e-16**

Notes: 1. Statistically significant variables with a p-value less than 0.05 are bolded. 2. W values are derived from Equation (7).

CONCLUSIONS

For a series of civil engineering students' abilities, such as applying engineering knowledge and tools, this study discusses the difference between the evaluation of industry professionals and students' self-evaluation. Based on the p-value outcome of the t-test models, it is evident that the mean score of professionals' assessments is statistically significantly different from the mean score of students' self-evaluation. Moreover, industry professionals generally rated students higher than students' assessment of their abilities. This shows that most students' presentation performance and achievements of the senior project satisfied the industry professionals. However, students have stricter requirements for their abilities. Compared to experts, students have insufficient work experience in the engineering field. Professionals tended to give students more encouragement to strengthen their confidence, which can help them cope with more complex tasks they may encounter in future work.

On the other hand, the higher evaluation from professionals may also be that they do not have enough time to participate in students' projects. Most of the professionals only focused their review on the final presentation of the project, which the students prepared for a long time. With the final speech alone, professionals could not understand the methods and time spent by students in solving problems. Hence, educators can appropriately adjust the project's progress so that professionals can be more involved in the entire project. This allows professionals to reach a consensus with students 'self-evaluation, which makes the program more profitable.

Therefore, if educational institutions or educators lack the professionals' assessment of the program or students' abilities, the evaluation outcomes of the professionals can be estimated by the students' self-evaluation according to the results of this study. Meanwhile, to make the estimated results more applicable, this study needs more future development. Increasing the sample size of the original data and collecting more evaluation data from multiple departments and universities can make the models more comprehensive. Additionally, when doing the assessment, providing some specific and detailed rubrics can also enable professionals and students to have a more accurate assessment of students' abilities.

REFERENCES

[1] Hassell, D., & Lee, K. Y. (2019). Evaluation of Self and Peer Assessments in a Second Year Engineering Module. IAFOR Journal of Education, 7(2).
[2] Falchikov, N. (2003). Involving students in assessment. Psychology Learning & Teaching, 3(2), 102–108.
[3] Falchikov, N. (2005). Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education. Oxon, UK: Routledge.
[4] Hanrahan, S. J., & Isaacs, G. (2001). Assessing self-and peer-assessment: The students' views. Higher Education Research & Development, 20(1), 53–70.

[5] Pandra, V., & Mardapi, D. (2017). Development of Mathematics Achievement Test for Third Grade Students at Elementary School in Indonesia. International Electronik Journal of Mathematics Education, 12(8), 769–776.
[6] Dolnicar, S., & Grün, B. (2009). Response Style Contamination of Student Evaluation Data. Journal of Marketing Education, 31(2), 160–172. https://doi.org/10.1177/0273475309335267
[7] Baird, J. A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: fields apart? Assessment in Education: Principles, Policy and Practice, 24(3), 317–350. https://doi.org/10.1080/0969594X.2017.1319337
[8] Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. Educational Research Review, 22, 74–98. https://doi.org/10.1016/j.edurev.2017.08.004.
[9] Rodríguez, María Constanza, et al. "TBL, Self-Assessment, and Use of ICT: A Methodological Option to Benefit the English Learning Process in a Colombian University." Journal of Language Teaching & Research, vol. 10, no. 6, Nov. 2019, pp. 1162–1172.
[10] Setiawan, Ari & Mardapi, Djemari & Supriyoko, Supriyoko & Andrian, Dedek. (2019). The Development of Instrument for Assessing Students' Affective Domain Using Self- and Peer-Assessment Models. International Journal of Instruction. 12. 425-438.
[11] Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. Review of educational research, 70(3), 287-322.
[12] Ünaldı, İ. (2016) 'Self and teacher assessment as predictors of proficiency levels of Turkish EFL learners', Assessment & Evaluation in Higher Education, 41(1), pp. 67–80.
[13] Petre, Aureliana-Loredana. "The impact of alternative assessment strategies on students." Scientific Research and Education in The Air Force–AFASES2017 (2017): 157-160.
[14] Jafarpur, A. (1991). Can naive EFL learners estimate their own proficiency? Evaluation & Research in Education, 5(3), 145-157.
[15] McNamara, T. (2001). Language assessment as social practice: Challenges for research. Language Testing, 18(4), 333-349.
[16] Taylor, S. N. (2014). Student Self-Assessment and Multisource Feedback Assessment: Exploring Benefits, Limitations, and Remedies. Journal of Management Education, 38(3), 359–383. https://doi.org/10.1177/1052562913488111
[17] Brown, G., & Harris, L. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. Frontline Learning Research. doi: 10.14786/flr.v2i1.24
[18] Brown, G. T., and Harris, L. R. (2013). "Student self-assessment," in Sage Handbook of Research on Classroom Assessment, ed J. H. McMillan (Los Angeles, CA: Sage), 367–393. doi: 10.4135/9781452218649.n21
[19] Silver, I., Campbell, C., Marlow, B., and Sargeant, J. (2008). Self-assessment and continuing professional development: the Canadian perspective. J. Contin. Educ. Health Prof. 28, 25–31. doi: 10.1002/chp.152
[20] Zimmerman, S., & Deckert-Pelton, M. (2003). Evaluating the Evaluators: Teachers' Perceptions of the Principal's Role in Professional Evaluation. NASSP Bulletin, 87(636), 28–37. https://doi.org/10.1177/019263650308763604
[21] Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. Improving Schools, 12(2), 101–113. https://doi.org/10.1177/1365480209105575
[22] Karmel, P. (1985). Quality and Equality in Education. Australian Journal of Education, 29(3), 279–293. https://doi.org/10.1177/000494418502900307
[23] Irons, D., Irons, T., & Maddux, C. D. (1984). A Survey of Perceived Competence among Psychologists Who Evaluate Students with Severe Handicaps. Journal of the Association for Persons with Severe Handicaps, 9(1), 55–60. https://doi.org/10.1177/154079698400900108

[24] Lindblom-ylänne, S., Pihlajamäki, H., & Kotkas, T. (2006). Self-, peer- and teacher-assessment of student essays. Active Learning in Higher Education, 7(1), 51–62. https://doi.org/10.1177/1469787406061148
[25] Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. Language Testing, 26(1), 075–100. https://doi.org/10.1177/0265532208097337
[26] Kearney, SP. (2013). 'Improving engagement: The use of "Authentic Self and Peer-assessment for Learning" to enhance the student learning experience', Assessment and Evaluation in Higher Education, vol. 38, no. 7, pp. 875-891.
[27] Kearney, Sean. Transforming the first-year experience through self and peer assessment, Journal of University Teaching & Learning Practice, 16(5), 2019.
[28] Liu, N.-F., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. Teaching in Higher Education, 11(3), 279–290.
[29] Box, G. (1953). Non-Normality and Tests on Variances. Biometrika, 40(3/4), 318-335.
[30] Armitage, P., Berry, G., & Matthews, J. N. S. (2001). Statistical methods in medical research.
[31] Snedecor, Cochran, & Cochran, William G. (1989). Statistical methods (8th ed.). Ames: Iowa State University Press.
[32] Welch, B. (1947). The Generalization of `Student's' Problem when Several Different Population Variances are Involved. Biometrika, 34(1/2), 28-35. doi:10.2307/2332510
[33] Wilcoxon, Frank (1945). "Individual comparisons by ranking methods". Biometrics Bulletin. 1 (6): 80–83.
[34] Pratt, J (1959). "Remarks on zeros and ties in the Wilcoxon signed rank procedures". Journal of the American Statistical Association. 54 (287): 655–667.