Two University Courses and Three Assessment Methods: Multiple Comparisons

Cleginaldo Pereira de Carvalho[a,b], Eduarda Pinto Ferreira[c] and Maurício César Delamaro[b]

[a] *Universidade Estadual Paulista, Faculdade de Engenharia de Guaratingueta, UNESP-FEG, São Paulo, Brazil*

[b] *Faculdade de Ciências Humanas do Estado de São Paulo, FACIC, São Paulo, Brazil*

[c] *Instituto de Ensino Superior do Porto, ISEP-P.PORTO, ISRC - Interdisciplinary Studies Research Center, Porto, Portugal*

*Corresponding author: Universidade Estadual Paulista, Faculdade de Engenharia de Guaratingueta, UNESP-FEG, Faculdade de Ciências Humanas do Estado de São Paulo, FACIC, São Paulo, Brazil
Email: cleginaldopcarvalho@hotmail.com

Notes on contributors

Cleginaldo Pereira de Carvalho, Post PhD in Industrial Engineering by University of Sao Paulo, Post PhD in Industrial Engineering by University of the State of São Paulo, Ph.D. in Mechanical Engineering by University of the State of São Paulo. Professor and Researcher in the Industrial Engineering Department of the University of the State of São Paulo. Professor in the Faculty of Human Sciences of Cruzeiro and in the Faculty of Technology of Pindamonhangaba. With years of experience as an Industrial Director in the corporate world, he incorporates his background onto his current research, which focuses on Learning Innovation, Lean Manufacturing, Knowledge Management and Industry 4.0. For more information, access: http://lattes.cnpq.br/0632897063038377 and https://orcid.org/0000-0001-7364-2096.

Eduarda Pinto Ferreira, has a PhD in Engineering Science (2006), a Master in Electrical Engineering – Telecommunication profile (1995) and Bachelor in Applied Mathematics and Computer Science (1991). She is Adjunct Professor in ISEP's Mathematics Department where she lectures since 1992 in several programmes and courses. Has co-supervised 2 PhD theses. She was President of the Pedagogical Council between 2010 and 2014. Since 2018 she is Director of the Master in Development Practice at ISEP and an active member of the Student Learning Assessment group of the CDIO She is member of Academic Program Committee of CSD&M, Paris, 2018 She is Coordinator of Euclides network. She is Coordinator of the Engineering Education Research line at the Interdisciplinary Studies Research Center - ISRC She was awarded with the prize for Pedagogical Innovation in Distance Education She has several publications in education research area. ORCID 0000-0002-6270-0382.

Maurício César Delamaro, has a Bachelor's Degree in Mechanical Engineering from Paulista State University (UNESP) (1985), a Master's Degree in Production Engineering from Federal University of Rio de Janeiro (1989) and a Ph.D. in Production Engineering from Federal University of Rio de Janeiro (1997). He is currently an assistant professor at State University of São Paulo (UNESP). For more information, access: http://lattes.cnpq.br/3727732364475835 and https://orcid.org/0000-0002-8453-7299.

Abstract: Assessment in high education is an important issue, especially when the focus is knowledge acquire measurement. Currently, different innovative learning methodologies are applied in order to develop the student's transversal skills and attributes as well as to improve their chances of success in the academic environment and professional career. The purpose of this study was to compare the assessment outcomes when three different innovative learning methodologies were used. The data sample consisted of student's from field engineers' courses, Industrial Engineering and Business Administration. ANOVA and t-Student were used for the statistical analysis. Peer Assessment had the best outcomes, however, more data will be presented and discussed.

Keywords: assessment; high education; team-based learning; peer assessment; project-based learning

1. Introduction

The development of the project learning methodology started in 1897, when the American philosopher John Dewey (1859-1952) proved that learning based on doing was a revolutionary way of study (Dewey, 1897). Using experimental methods, he studied the student's ability to think in a gradual way and, the acquisition of knowledge related to the ability to solve real-life projects, adding to the area of study emotional and intellectual development.

Constructivism explains what human beings learn through interaction with the environment and this experience is perceived differently for each student, although the student learns with his current knowledge about an objective (Markham, Larmer & Ravitz 2008). Constructionism, on the other hand, promotes a verification of individual learning, detailing all the steps of a path; the human being learns better as he builds and shares knowledge with others (Grant, 2002).

Modern engineering education programmes aim to enrich student's with the necessary knowledge, skills and attitudes for becoming successful young engineers. Assessment is defined as the process of gathering evidence on student's outcomes that can be used to draw reasonable inferences about the impact of the education process and what student's know (Wengrowicz, Dori & Dori, 2017).

Developing a process for this, that meets learner, employer and higher education framework requirements is challenging and, therefore, is a demanding and valuable focus for further research in this innovative area of higher education (Brodie & Irving, 2007).

Pereira, Flores & Niklasson (2016) in their work about to Assessment and Evaluation in Higher Education address the following:

- Assessment methods used in higher education, their effectiveness, fairness and their influence on learning and impact on teaching. Aspects related to the assessment methods and their impact on the student's performance are also tackled;
- Forms of assessment in higher education, related to self- and peer assessment practices and the monitoring of learning, taking into consideration formative, continuous and summative assessment;
- Learning and teaching practices and their impact on assessment. The influence of certain learning environments and contexts on student learning and assessment is also discussed.

This study aims to compare the assessment outcomes from an Industrial Engineering and a Business Administration courses by the application of three different innovative learning methodologies:

- Team-Based Assessment (TBA);
- Peer Assessment (PA);
- Project-Based Assessment (PBA).

The comparison was conducted by the statistical tools analysis and the results were presented and discussed.

2. Theoretical Background

2.1 Assessment in higher education

The assessment of learning is a key component in higher education. One of the models, probably one of the most used, was proposed by Donald L. Kirkpatrick in 1959 (Kirkpatrick, 1959; Kirkpatrick & Kirkpatrick, 2007). It aims at measuring the quality and effectiveness of a formative intervention by analysing the four levels of assessment:

- Level 1 - Evaluation – Reaction:

  Evaluation of the Participants Reaction to the System aims to evaluate the degree of satisfaction of the different players in the training process.

- Level 2 - Evaluation – Learning:

  Participant Knowledge Assessment aims to assess the extent to which trainees are acquiring intended and predefined knowledge in the diagnosis of training needs (knowing how to know).

- Level 3 - Evaluation – Transfer:

  Participant Behaviour Assessment aims to assess the extent to which new knowledge acquired is modifying behaviours and being put into practice by improving their performance.

- Level 4 - Evaluation – Results:

  Evaluation of Training Results and their Impact aims to assess the variation detected from tangible indicators.

J. Phillips proposed a fifth level of evaluation, return on investment (ROI approach) (Phillips, 1996). In addition to assessing potential impacts on an organization's performance, cost-benefit analysis should be conducted to determine whether the impacts justify the investment made. For Kirkpatrick & Kirkpatrick (2006, 2007) the ROI approach is implicit at level 4 (outcome evaluation).

Miller & Lesks (2005), proposed 5 levels of assessment, ranging from from student to institution level assessment:

- Level 1 – Assessing individual student learning within courses;
- Level 2 – Assessing individual student learning across courses;
- Level 3 – Assessing Courses;
- Level 4 – Assessing Programs;
- Level 5 – Assessing the Institution.

Regarding assessment implementation, i.e. a process perspective (Tyler & Hlebowitsh, 2013) sees it as the constant comparison between student's results and their previously defined performance or goals. Assessment is thus the process of determining the extent to which educational objectives are realized. In a massified engineering education scenario, a disciplinary approach to learning is much easier to implement, resulting in assessment being often restricted to knowledge assessment.

This perspective has to evolve due to technological, human and societal factors. Engineering systems are becoming increasingly complex, quite often multidisciplinary, thus requiring the student's to practice actual project/product development and to learn how to work in teams. There is the need to teach student's real-world engineering practice. The worldwide CDIO (conceive-design-implement-operate) initiative (www.cdio.org) aims to bridge this gap between scientific/disciplinary and practical engineering learning. CDIO proposes a syllabus and 14 standards to guide program design and deployment. Most important, it covers all 5 Leskes assessment levels in an integrated and coherent approach.

CDIO it is not prescriptive regarding assessment methods, but they must be aligned with the active and project-based learning methodologies used in CDIO programs. Based on CDIO principles, we will focus on 3 student assessment methods: PA; TBA; PBA. All these innovative/non-traditional assessments methods are especially useful to assess team and process skills (section 3 and 4 of the CDIO syllabus).

2.2 Peer Assessment

Eric Mazur is known for introducing in 1991 a teaching strategy called peer instruction (Mazur, 2013). In 1997 he published the book "Peer Instruction: A User's Manual", which provides details of this strategy, including the application of peer assessment.

In peer assessment, a collaborative learning technique is used in which student's evaluate or partially evaluate the work of their peers. In the context of peer-learning, this type of assessment is regarded as integral part of the learning process, as it contributes to learning by all participating student's, both the evaluated and evaluators.

The evaluation techniques and criteria should be clear and well defined in advance. Student's should also train/rehearsal the process before applying it for grading. This PA can have the following advantages:

- As student's are evaluating their peers, they are also learning. As a result, they are helping teachers in the classroom and they also feel to contribute to the learning process (Searby & Ewers, 1997);
- In pedagogical terms (Liu & Carless, 2006) (Bryant & Carless, 2010) and metacognitions (Sadler & Good, 2006) state that student's can learn from their peers;
- As student's are evaluating their peers, it leads them to reflect on their work. This reflection process can stimulate actions for improvement. (Kristanto, 2018) .

Monterola, Roxas & Carreon-Monterola (2009) showed that teachers who used peer instruction (PI) in class plan to continue to use it in the future.

2.3 Team Assessment

Group work is a method of instruction where student's are working together. There are great advantages to group work as a teaching and learning tool. Despite the advantages of group work, there is a major challenge in assigning grades and feedback to group student's.

Student's have different characteristics and therefore have different learning rates and therefore cannot be treated equally. The current school must be multi and intercultural, with the concern to integrate all student's (Gundara & Sharma, 2013). Teaching differentiation mechanisms will need to be accompanied by assessment differentiation mechanisms for clearly differentiated student groups.

It is also necessary to take into account other types of knowledge, skills, attitudes and values, to which Rotgans et al. (2019) adds that we need to value cross-cutting organization, communication and problem-solving skills. The teaching process, by focusing on the student, implies that different teaching methods and techniques begin to be used more, such as group work, research and project.

However, evaluating a group work is not easy it is difficult to accurately monitor the learning only by the final work presented. The development process and teamwork are also paramount. Individual effort assessment, if required, is also difficult to assess in some contexts. There are always some student's who work less than their peers, others who, on the contrary, try to the whole work themselves. Both are undesirable outcomes.

While there may be differences between group members, some may have more difficulties than others, the effort of each member of the group should be balanced. This heterogeneity can help student's with more learning difficulties.

Truong et al. (2014) conducted IT student's teamwork capabilities measurement by assessment and analysis using a rubric which used five different dimensions:

- Shared leadership;
- Team orientation;
- Effort redundancy;
- Learning results;
- Team's autonomy.

2.4 Project-Based Assessment

In terms of Pedagogy, Project-Based Learning (PBA), can be allocated within the constructivist approach of learning education, since in opposite side of traditional approaches the teachers assume the facilitator profile.

Mitchell & Rogers (2019) demonstrated using the PBA application that the student's involved in the project-based learning activities, were more engaged than any other student. The staff verified that they had improved the student's abilities in terms of skills in problem-solving, teamwork and design. Conclusions drawn support the use of authentic project-based engineering activities as a vehicle to best develop these skills.

Experiments were conducted in Finland University of Applied Science, in order to discover the impact of using PBA to improve first-year retention and its effect on the practical pedagogical methods already in place (Vesikivi et al. 2019). The student's had their competence development assessed by a questionnaire and scored high in the scores targeted by the course through project-based teamwork (collaboration around shared objects, integration of efforts, feedback practices, persistent development and exploiting technology for collaboration).

Chu et al. (2017) work on the effectiveness of PBA, compared the perceptions and actions among the student's in three undergraduate courses of different disciplines, English Language, Information Management, and Mechanical Engineering, after the application of the PBA.

Project-based learning can be an effective approach in implementing principles and practices of sustainability within a leaning environment and can be useful in helping student's become more effective problem solvers and professionals (Leal Filho, Shiel & Paço, 2016).

3. Research questions

This research study incites the following questions:

(1) What is the relationship between the assessment methodology and the average of the student's outcomes?;

(2) What is the assessment methodology in which the outcomes are the highest?;

(3) What is the difference of the results when comparing outcomes from the Industrial Engineering Course with the Business Administration Course, applying the three methodologies?. This research question is important to address because we were treated two different courses from two different Faculties located in different cities each one with its own culture;

(4) What is the influence of the gender on the outcomes results considering the Industrial Engineering and Business Administration groups?. This question makes sense since we had conducted this research work in Brazil, where the culture directs male student's for Engineering courses and female student's to Business Administration. Hence some differences between the outcomes could be expected, especially when innovative assessment methodologies are used. Maybe It might be as important other countries without a discriminatory culture.

4. Method

4.1 Participants

The subjects were student's of the Industrial Engineering (IE) and Business Administration (BA) courses. The group of 81 student's of IE were submitted to three different types of assessment (TBA, PA, PBA), coursing the discipline Product Engineering. Furthermore, 67 student's of BA, enrolled in the subject "Production Management", were submitted to the same methodologies.

The student's belonged to two different colleges located in São Paulo state, Brazil.

4.2 Research procedures and sources data

A quantitative-experimental study was conducted on the validation of the null hypothesis, which deemed the average among the outcomes from the three methodologies were the same.

Both groups were submitted to the three assessments during one semester of classes. The data of the assessment outcomes was collected, first from the TBA, followed by the PA and, finally, the PBA.

In the TBA, the outcomes were the grades of individual student. The examination was conducted in teams with a maximum of six student's. Each team member student had a different exam. The examination was on the theoretical concepts covered until the examination. The team members had to help each other, but the grade was individual. The instructor corrected the individual examination and gave a grade for each student.

The PA was conducted two weeks after the TBA. It was an individual examination and in this phase, each student received an examination solved by an unknown colleague and had to correct it. However, the student's were not given the solutions, they had to correct their colleagues' exams by themselves. Afterwards, the instructor assessed the correction and graded the student.

Finally, in the PBA, which was conducted in a group, the student's were evaluated in accordance with three events:

a) The scientific paper to be submitted to a journal: each group must present a finished paper about the project they developed and implemented (weight 40% of the PBA grade);

b) Oral presentation: all the members of the group must present the project to be evaluated (weight 30% of the PBA grade);

c) Individual examination: the instructor selects one member to undergo an examination regarding the contents of his group project (weight 30% of the PBA grade).

At the end of PBA, the instructor graded each student individually.

The data analysis was conducted in accordance with the following sequence using ANOVA in the first approach, Tukey and t-test:

- Comparison of the assessment methodologies grade outcomes for IE;
- Comparison of the assessment methodologies grade outcomes for BA;
- Comparison between IE and BA:
  - General average grade IE x BA.
- Average grade considering each kind of assessment methodology:
  - BA PA x IE PA;
  - BA TBA x IE TBA;
  - BA PBA x IE PBA.

4.3 Data processing methods

For the purposes of the article, it was important to identify if there were significant differences between the average strata of the surveyed population, such as: a) Is there a significant difference between the grades obtained by the student's depending on the three methods used? b) Is there a significant difference between the performance of the two educational institutions? c) Is there a difference between performances depending on the gender of the student's? In these cases, we were used to a tool will Variance lysis to a criterion (Montgomery, 2009) as to whether and there is significant difference between the measurements obtained by the student's in each case.

The null hypothesis tested, of equality between the n averages and the alternative hypothesis can be presented as follows:

H_0: $\mu_1 = \mu_2 = \ldots = \mu_n$

H_1: There is at least one different average (1)

The test value of the Analysis of Variance is the quotient between the variance estimate between the samples and the residual variance estimate (Triola, 2017). This quotient between variance estimates follows a Snedecor F curve, with degrees of freedom depending on how many populations are being compared and the sample size of each pollution. The probability value of the Snedecor F be greater than the test value is value- p testing to be compared with the usual significance levels (10%, 5%, 2.5% and 1%). The lower the p-value, the greater the rejection force of the null hypothesis, indicating the existence of a significant difference between the averages.

But when the analysis of variance identifies the difference between the averages, the question remains: which average should be considered different from each other? To answer this question, there is need for continuity of analysis. Therefore, in this work, it was decided to use two methods: the multiple comparisons of Tukey's t-test and applied between the samples two by two (Walpole et al. 2009).

The Tuckey method uses critical values of the standardized amplitude denoted by q. The literature (include citation with table) provides critical values of q in the case of normal population. If we want to compare k samples, each with n elements (samples of equal size), the procedure recommends considering different averages μ_i and μ_j such that:

$$|\bar{x}_i - \bar{x}_j| > q_{k,\upsilon,\alpha}\sqrt{S_R^2/n}\,, \quad (2)$$

where α is the desired level of significance, $\upsilon = k(n-1)$ and S_R^2 is the residual variance.

When sample sizes are not the same, there are slight variations in the calculations, but the principle is the same. However, by this method, it is not possible to know from what level of significance can be suggested that there is a difference between these averages. For this, the *t*-test was applied, comparing the samples two by two (Montgomery & Runger, 2013).

An always present case in this study is the pairing of the samples. The identity of the student's presented himself as a criterion that makes it possible to match the data. Then the *t*-test is used to compare two averages with each other. When data from two samples are paired, it makes sense to calculate the differences d_i corresponding to each pair of values and test the hypothesis that the difference between the averages of the two paired populations is equal to a certain value Δ. This is equivalent to testing the hypothesis that the mean of all population differences is equal to **Δ.**

When pairing was not possible, the t-test values depend on whether the population variances are the same or different from each other. In these cases, there was a need for prior comparison of the variances to decide whether variances can be considered equal or not. Afterwards the null hypothesis was tested: $H_0: \sigma_1^2 = \sigma_2^2$, against the alternative hypothesis of difference between the variances.

Obviously, when the number of populations (or strata) compared is equal to two, the initial steps of analysis of variance and multiple Tuckey comparisons will not be performed. In these cases, the appropriate t-test was used directly, and the variance compared if necessary.

Variance Analysis, variance comparison (F-test), and t-tests were performed using Microsoft Excel® software, licensed by one of the authors, using the "*Data Analysis*" routines of the spreadsheet.

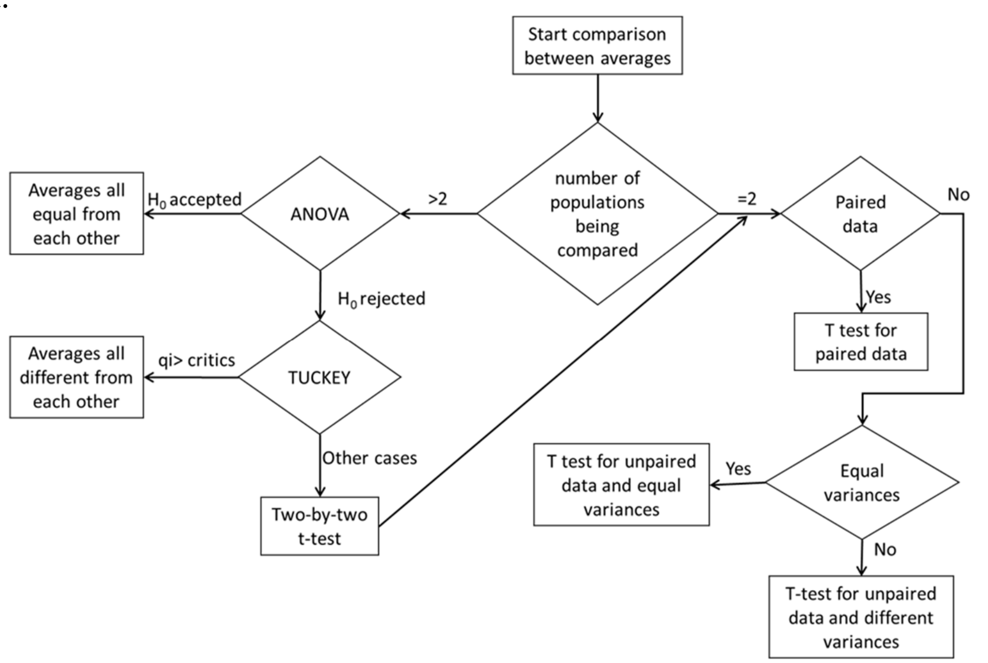

Figure 1: Data processing logic

5. Results and discussions

5.1 Comparison between assessment methods

5.1.1 Comparison between the methods in the IE course

The authors wish to find out whether there is a significant difference between the averages obtained by the student's, depending on the three methods used (TBA, PA and PBA).

For this, Variance Analysis was used, with a classification and same size samples.

The null hypothesis to be tested, of equality between the three averages and the alternative hypothesis can be presented as follows:

H_0: $\mu_{team} = \mu_{peer} = \mu_{pbl}$

H_1: There is at least one different average (3)

The results summary and the analysis of the variance table are presented in table 1 and 2, respectively:

Table 1: Summary of results for analysis of variance

Groups	*Score*	*Sum*	*Average*	*Variance*
TBA	81	607	7,49	3,36
PA	81	676,5	8,35	3,40
PBA	81	645	7,96	1,56

Table 2: Variance analysis chart

Variation Source	*Sum of Squares*	*Degrees of freedom*	*Medium square*	*F*	*P-value*
Among groups	29,90	2	14,95	5,38	0,0052
Within groups	666,35	240	2,78		
Total	696,26	242			

The *p*-value found is much lower than the usual significance levels. This indicates a significant difference between the 3 averages.

But the question remains: which average(s) should be considered different from which other? At first, the average of the PA method seems to be the largest and the TBA method the smallest. But it is necessary to continue the analysis as another conclusion maybe reached when using the difference between the three averages or partial differences, two by two. Thus, the authors decided to use two methods: Tuckey's multiple comparisons and the paired t-test applied between the samples, two by two.

For analysed cases, we have n = 81, k = 3 and S_R^2= 2.78. And by adopting the 5% significance level, we should use $q_{3,240,5\%}$ = 3,33. Therefore, the averages that differ from more than 0,62 $(=q_{3,240,5\%} \cdot \sqrt{\frac{S_R^2}{n}})$ should be considered different. The results are:

$$| \bar{x}_{TBA} - \bar{x}_{PA} | = 0{,}86$$
$$| \bar{x}_{PA} - \bar{x}_{PBA} | = 0{,}39 \quad (4)$$
$$| \bar{x}_{PBA} - \bar{x}_{TBA} | = 0{,}47$$

That is, the averages of the TBA method and the peer method are considered different from each other. In fact, the average of the PA method is greater than TBA's. Moreover, when analysed two by two there seems to be no significant difference: between PBA and TBA and between TBA and PBA.

Although, by using this method, it is not possible to know from what level of significance it can be suggested that there is a difference between these averages. Hence, the t-test was applied, with paired samples, two by two. The student's identity is the criteria that makes it possible to match the data.

In summary, thanks to the data matching, we have the desired and the performed tests, which are shown in Table 3.

Table 3: The t-tests performed

Test initially desired	***Test performed thanks to data pairing***
H_0: $\mu_{TBA} = \mu_{PA}$	H_0: $\mu_{d1} = 0$
H_1: $\mu_{TBA} \neq \mu_{PA}$	H_1: $\mu_{d1} \neq 0$
H_0: $\mu_{TBA} = \mu_{PBA}$	H_0: $\mu_{d2} = 0$
H_1: $\mu_{TBA} \neq \mu_{PBA}$	H_1: $\mu_{d2} \neq 0$
H_0: $\mu_{PBA} = \mu_{PA}$	H_0: $\mu_{d3} = 0$
H_1: $\mu_{PBA} \neq \mu_{PA}$	H_1: $\mu_{d3} \neq 0$

The *t*-tests were performed with the aid of Microsoft Excel ® software, using the "*T-test: two paired samples for averages*", available in "*Data Analysis*". The summary of results are in Table 4.

Table 4: Results of t-tests

	H_0: $\mu_{TBA} = \mu_{PA}$		***H_0: $\mu_{TBA} = \mu_{PBA}$***		***H_0: $\mu_{PBA} = \mu_{PA}$***	
	TBA	***PA***	***TBA***	***PBA***	***PBA***	***PA***
mean	7,49	8,35	8,35	7,96	7,96	7,49
variance	3,37	3,40	3,40	1,56	1,56	3,36
notes	81	81	81	81	81	81
difference hypothesis	0		0		0	
degrees of freedom	80		80		80	
***t* Student experimental**	-3,20		1,90		1,88	
one- tailed *p*-value	0,00		0,03		0,03	
one-tailed *t* critic	1,66		1,66		1,66	
two- tailed *p*-value	0,00		0,06		0,06	
two- tailed *t* critic	1,99		1,99		1,99	

From these results, especially by analysing the p-values, one may conclude that the data is compatible with the mean difference, with strong or moderate evidence, as summarized in Table 5.

Table 5: Summary of t-test results

	PA	***PBA***
TBA	One- tailed *p*-value = 0.00 Two-tailed *p*-value = 0. 00 It is concluded that the results are compatible with the difference between the averages (strong evidence).	One- tailed *p*-value = 0.03 Two-tailed *p*-value = 0.06 It is concluded that the results are compatible with the difference between the averages (moderate evidence).
PA		One- tailed *p*-value = 0.03 Two-tailed *p*-value = 0.06 It is concluded that the results are compatible with the difference between the averages (moderate evidence).

These results are always compatible with the difference between the three averages, although to different degrees of intensity in the certainty of these conclusions.

The results obtained by both methods (Tuckey method and two-by-two t-test) are compatible and can be summarized as follows:

$$\mu_{PA} > \mu_{PBA} > \mu_{TBA} \quad (5)$$

This maybe because the group of student's had never been assessed by TBA before and, since it was the first type of assessment the grade results were inferior to PA and PBA. Also, PA presents the highest score as the student's were more familiar with the content assessed, as added to the fact that they assessed their peer on the same content that was applied in the TBA. The performance of the PBA was expected to be the highest but the final grade of the this methodology was composed of three parts: a scientific article (40% of the total PBA grade), oral presentation by the entire group (30% of the total PBA grade) and individual examination by one group member (30% of the total PBA grade).

5.1.2 Comparison between the evaluation methods in the BA course

As in the first course, we sought to know if there was a significant difference between the averages obtained by the student's in the three methods used (TBA, PA and PBA). The null hypothesis, equality between the three averages, and the alternative hypothesis, can be presented as follows:

H_0: $\mu_{TBA} = \mu_{PA} = \mu_{PBA}$

H_1: There is at least one different average (6)

The summary of results and the analysis of variance table are presented in Tables 6 and 7.

Table 6: Summary of results for analysis of variance

Groups	*Count*	*Sum*	*Average*	*Variance*
TBA	67	454	6,78	3,68
PA	67	584,5	8,72	2,68
PBA	67	532,5	7,95	1,80

Table 7: Variance analysis chart

Variation Source	*Sum of Squares*	*Degrees of freedom*	*Medium square*	*F*	*p-value*
Between groups	128,84	2	64,42	23,67	6,06E-10
Within groups	538,85	198	2,72		
Total	667,69	200			

The p-value found is much lower than the usual significance levels. This suggests that one could reject the null hypothesis. To further determine which average (s) should be considered different, the Tuckey method is applied again.

For the case under analysis, we have: n = 67; k = 3; S_R^2 = 2,72. And by adopting the 5% significance level, we should use $q_{3,198,5\%}$ =3,33. Therefore, the averages that differ from more than 0,67(=$q_{3,198,5\%} \cdot \sqrt{\frac{S_R^2}{n}}$) should be considered different. . The results are:

$$|\bar{x}_{TBA} - \bar{x}_{PA}| = 1{,}95$$
$$|\bar{x}_{PA} - \bar{x}_{PBA}| = 0{,}78 \quad (7)$$
$$|\bar{x}_{PBA} - \bar{x}_{TBA}| = 1{,}17$$

The conclusion is that the three averages are distinct from each other, with very clear certainty.

$$\mu_{PA} > \mu_{PBA} > \mu_{TBA} \text{ , } very\ sure \quad (8)$$

For this school, it was not necessary to perform the *t*-tests, two by two. Tukey's method was enough to reach the conclusion that the averages of the three methods were different.

This result was similar to the previous one and reinforces that PA outcomes are higher than the others obtained from the PBA and TBA.

5.1.3 Comparison between assessment methods GROWING the two courses

The ANOVA results for the comparison of averages, according to methods, when the two schools were grouped, are presented in Tables 8 and 9.

Table 8: Results summary for analysis of variance

Groups	*Count*	*Sum*	*Average*	*Variance*
TBA, 2 schools together	148	1061	7,17	3,61
PA, 2 schools together	148	1261	8,52	3,09
PBA, 2 schools together	148	1177,5	7,96	1,66

Table 9: Variance analysis chart

Variation Source	*SQ*	*gl*	*MQ*	*F*	*P-value*
Between groups	136,36	2	68,18	24,46	8,42345E-11
Within groups	1229,18	441	2,79		
Total	1365,54	443			

One must reject the null hypothesis. Using Tuckey again, we have n = 148; k = 3; S_R^2 = 2,79.

And by adopting the 5% significance level, we should use $q_{3,198,5\%}$ = 3,31. The averages averaging more than 0,46 (=$q_{3,441,5\%} \cdot \sqrt{\frac{S_R^2}{n}}$) should be considered different. The results are:

$$| \bar{x}_{TBA} - \bar{x}_{PA} | = 1{,}35$$
$$| \bar{x}_{PA} - \bar{x}_{PBA} | = 0{,}56 \quad (9)$$
$$| \bar{x}_{PBA} - \bar{x}_{TBA} | = 0{,}79$$

It follows that

$$\mu_{PA} > \mu_{PBA} > \mu_{TBA}\text{, very sure} \quad (10)$$

It is noticed that there is consistency within the results, either in the comparison between each course, as well as in the grouping of the two courses.

5.2 Comparison between courses

5.2.1 Between IE and BA, as for the overall average

Since we want to compare the averages of two populations (between two colleges), the t test was used directly. Firstly, the average of the total grades of the three methods was calculated for each student. Data was unpaired, but it was not known whether the variances could be considered equal or not. Consequently, it was initially tested:

$$H_0: \sigma_{IE}^2 = \sigma_{BA}^2$$
$$H_1: \sigma_{IE}^2 \neq \sigma_{BA}^2 \quad (11)$$

The p-value of this F test (with 66 and 80 degrees of freedom) was 71%, which indicates that H_0 should be accepted. Thus, we opted for the t-test with variances considered equal for both populations, shown in Table 10.

Table 10: *t* test for course average

	H_0: μ_{IE} = μ_{BA}	
	BA	**IE**
average	7,82	7,94
variance	1,30	1,19
notes	67	81
grouped variance	1,24	
hypothesis of difference	0	
degrees of freedom	146	
statistic t	-0,69	
uni-flow p-value	0,25	
t single- tailed critic	1,66	
two-tailed p-value	0,49	
t two-tailed critic	1,98	

By analysing the *p*-values, ene may conclude that the null hypothesis should be accepted. There was no significant difference in student's averages, according to the course attended.

5.2.2 Comparison between schools in each method

Initially, the equality of variances was tested. In conclusion, they may be considered statistically equal, as summarized in Table 11.

Table 11: Summary of Variance Comparison Tests

	$H_0;\ \sigma^2_{IE,Team} = \sigma^2_{BA,Team}$	$H_0;\ \sigma^2_{IE,Peer} = \sigma^2_{BA,Peer}$	$H_0;\ \sigma^2_{IE,Pbl} = \sigma^2_{BA,Pbl}$
Bi-tailed *p*-value	0.70	0.32	0.55
Gl`s	80 and 66	80 and 66	80 and 66
Conclusion	Accept H_o	Accept H_o	Accept H_o

Next, t-tests were performed for each pair of populations, the results of which appear in Table 12.

Table 12: Summary of comparison tests of averages according to teaching methods

	$H_0;\ \mu_{BA,Team} = \mu_{IE,Team}$		$H_0;\ \mu_{BA,Peer} = \mu_{IE,Peer}$		$H_0;\ \mu_{BA,Pbl} = \mu_{IE,Pbl}$	
	BA e TBA	***IE e TBA***	***BA e PA***	***IE e PA***	***BA e PBA***	***IE e PBA***
Average	6,78	7,49	8,72	8,35	7,95	7,96
Variance	3,68	3,37	2,68	3,40	1,80	1,56
Notes	67	81	67	81	67	81
Grouped variance	3,510		3,078		1,668	
Mean difference hypothesis	0		0		0	
Degrees of freedom l	146		146		146	
statistic t	-2,32		1,28		-0,07	
uni-flow p-value	0,01		0,10		0,47	
t single- tailed critic	1,66		1,66		1,66	
two-tailed p-value	0,02		0,20		0,94	
t two-tailed critic	1,98		1,98		1,98	
Conclusion	Reject H_0		Accept H_0		Accept H_0	

The only significant difference when comparing two courses was in the TBA methodology. The IE student's performed significantly better in this method than the BA student's. This could be because the PBA was applied for the first time in the BA course and the BA student's had never undergone this kind of assessment method.

It is possible to observe that all p-values are much high than the usual significance levels. This allows us to conclude that there is no difference in performance outcomes between genders. Considering both the general approach and the individual method of assessment.

6. Conclusion

This research mainly contributed to the present knowledge by providing a comparison of assessment outcomes when different innovative learning methodologies were applied.

Answering the research questions:

What was the relationship between the method of assessment and the average of the student's outcomes?

The results are compatible and can be summarized as follows:

$$\mu_{PA} > \mu_{PBA} > \mu_{TBA}, very\ sure \quad (12)$$

The group of student's had never been assessed by TBA before and, since it was their first type of assessment, the grades were lower than PA and PBA. PA presents the highest scores because the student's were more familiar with the contents evaluated, furthermore, they assessed their peers using the same knowledge as in the TBA assessment.

- What was the assessment methodology in which the outcomes are the highest?.

The outcomes obtained using PA are higher than the ones obtained from PBA and TBA.

- By applying these three methodologies, what was the difference of the results when comparing outcomes from the Industrial Engineering Course with the Business Administration Course?.

There was no significant difference in student averages no matter what course they were attending.

This research can be used as a guide for lecturers and researchers, assisting them when choosing the most suitable assessment methodologies, and can be applied independently of the high education course they are involved in.

A possible a limitation of this study, is that the comparison was conducted between two high education courses in the same Brazilian region of the State of Sao Paulo and in the same university. Another limitation that may have influenced the results, is that only one significance level was used in the analysis.

For future researches, one suggests that the same approach could be used in other different high education courses and regions, and different levels of significance could be tested in order to give more accurate results.

References

Brodie, P. B., & Irving, K. (2007). Assessment in work□based learning: investigating a pedagogical approach to enhance student learning. *Assessment & Evaluation in Higher Education*, 32(1), 11-19. https://doi.org/10.1080/02602930600848218

Bryant, D.A., & Carless, D.R. (2010). Peer assessment in a test-dominated setting: empowering, boring or facilitating examination preparation?. *Educational Research for Policy and Practice*, 9, 3–15. https://doi.org/10.1007/s10671-009-9077-2

Chu, S. K.W, Zhang, Y., Chen, K., Chan, C.K., Yi Lee, C.W., Zou, E., & Lau, W. (2017). The effectiveness of wikis for project-based learning in different disciplines in higher education. *Internet and Higher Education*, 33, 49-60. https://doi.org/10.1016/j.iheduc.2017.01.005

Dewey, J., (1897). My pedagogic creed. *The School Journal*, 54, 77-80.

Gundara, J.S., & Sharma, N. (2013). Some issues for cooperative learning and intercultural education. *Intercultural Education*, 24(3), 237-250. https://doi.org/10.1080/14675986.2013.797202

Grant, M. M. (2002). Getting a grip on project-based learning Theory, cases. *Meridian: A Middle School Computer Technologies Journal*, 5(1). http://www.ncsu.edu/meridian/win2002/514/2.html

Kirkpatrick, D. L. (1959). Techniques for Evaluation Training Programs. *Journal of the American Society of Training Directors*, 13, 21-26.

Kirkpatrick, D.L., & Kirkpatrick, J.D. (2006). *Evaluating Training Programs: The Four Levels. (3ª ed.)*. Berrett-Koehler Publishers, ISBN-13: 978-1576753484, ISBN-10: 1576753484.

Kirkpatrick, D.L., & Kirkpatrick, J.D. (2007). *Implementing the Four Levels. (1ª ed.)*. Berrett-Koehler Publisher, ISBN-13: 978-1576754542, ISBN-10: 1576754545.

Kristanto, Y.D. (2018). Technology-enhanced pre-instructional peer assessment: Exploring students' perceptions in a Statistical Methods course. *REiD (Research and Evaluation in Education)*, 4(2), 105–116.

Leal Filho, W., Shiel, C., & Paço, A. (2016). Implementing and operationalizing integrative approaches to sustainability in higher education. *Journal of Cleaner Production*, 133, 126-135. https://doi.org/10.1016/j.jclepro.2016.05.079

Liu, Ngar-Fun, & Carless, D. (2006). Peer Feedback: The Learning Element of Peer Assessment. *Teaching in Higher Education*, 11(3), 279–290. https://doi.org/10.1080/13562510600680582

Markham, T. Larmer, J., & Ravitz, J. (2008). *A Aprendizagem Baseada em Projetos*. Artmed Editora S/A, Porto Alegre.

Mazur, E. (2013). Peer Instruction: *A User's Manual*. Kindle Edition. ISBN-13: 978-0135654415, ISBN-10: 0135654416.

Miller, R., & Leskes, A. (2005). *Levels of Assessment: From the Student to the Institution*. Washington, D.C.: Association of Amer project-based system versities. Association of American Colleges and Universities. ISBN 0-9763576-6-6.

Mitchell, J.E., & Rogers, L. (2019). Staff perceptions of implementing project-based learning in engineering education. *European Journal of Engineering Education*, 45(3), 349-362. https://doi.org/10.1080/03043797.2019.1641471

Monterola, C.; Roxas, R.M., & Carreon-Monterola, S. (2009). Characterizing the Effect of Seating Arrangement on Classroom Learning Using Neural Networks. *Complexity*, 14 (4), 26–33. ISSN 1076-2787. https://doi:10.1002/cplx.20237

Montgomery, D. C. (2009). *Introduction to Statistical Quality Control*. *(6ª ed.)*. New York: John Wiley and Son.

Montgomery, D. C; & Runger, G. C. (2013). *Applied Statistics and Probability for Engineers. (6ª ed.)*. Wiley. ISBN 10: 1118539710 ISBN 13: 9781118539712.

Pereira, D., Flores, M.A., & Niklasson, L. (2016). Assessment revisited: a review of research in Assessment and Evaluation in Higher Education. *Assessment & Evaluation in Higher Education*, 41(7), 1008-1032. https://doi.org/10.1080/02602938.2015.1055233

Phillips, J.J. (1996). How much is the training worth?. *Training and Development*, 50(4), 20-24.

Rotgans, J.I., Rajalingam, P., Ferenczi, M.A., & Low-Beer, N. (2019). A Students' Model of Team-based Learning. *Health Professions Education*, 5(4), 294-302. https://doi.org/10.1016/j.hpe.2018.10.003

Sadler, P. M., & Good, E. (2006). The Impact of Self- and Peer-Grading on Student Learning. *Educational Assessment*, 11(1), 1–31.

Searby, M., & Ewers, T. (1997). An Evaluation of the Use of Peer Assessment in Higher Education: A Case Study in the School of Music. *Assessment & Evaluation in Higher Education*, 22(4), 371-383. https://doi.org/10.1080/0260293970220402

Triola. (2017). *MF Introduction to Statistics (Translation and technical revision of Farias, AML & Flores, VRLF. (12ª ed.)*. Rio de Janeiro: LTC Gen.

Truong, Vu T., Bao, N. Le, Nguyen, Man D., & Nguyen T.M. (2014, June 16 – 19). *Assessing the Maturity of Teamwork Capabilities Through CDIO Projects [Conference presentation]*. Proceedings of the 10th Annual International CDIO Conference, Universitat Politècnica de Catalunya, Barcelona, Spain. http://www.cdio.org/files/document/cdio2014/123/123_Paper.pdf

Tyler, R.W., & Hlebowitsh, P.S. (2013). *Basic principles of curriculum and instruction (1ª ed.)*. University of Chicago Press, Revised edition, ISBN-13: 978-0226086507, ISBN-10: 022608650X.

Vesikivi, P., Lakkala, M., Holvikivi, J. & Muukkonen, H. (2019). The impact of project-based learning curriculum on first-year retention, study, experience, and knowledge work competence. *Research Papers in Education*. https://doi.org/10.1080/02671522.2019.1677755

Walpole, R. E., Myers, R.H., Myers, S. L, & Keying, Ye. (2009). *Probability and Statistics for Engineering and Science. (Translation by Vianna, LFP)*. São Paulo: Pearson Prentice Hall.

Wengrowicz, N., Dori, Y. J., & Dori, D. (2017). Meta-assessment in a project-based systems engineering course. *Assessment & Evaluation in Higher Education*, 42(4), 607-624. https://doi.org/10.1080/02602938.2016.1173648