



Effective Metaheuristic Assignment to Improve Equivalence in Small Experimental Pedagogical Groups

Diego Roldán¹, Víctor Aguilar², Francisco Roldán³, Manuel Villavicencio⁴

¹ University of Cuenca, Faculty of Economics, Av. 12 de Abril y Av. Loja, Cuenca – Ecuador, 010203, diego.roldan@ucuenca.edu.ec

² University of Cuenca, Faculty of Economics, Av. 12 de Abril y Av. Loja, Cuenca – Ecuador, 010203, victor.aguilar@ucuenca.edu.ec

³ University of Cuenca, Faculty of Economics, Av. 12 de Abril y Av. Loja, Cuenca – Ecuador, 010203, roldan.arauz@yahoo.com

⁴ University of Cuenca, Faculty of Philosophy, Av. 12 de Abril y Av. Loja, Cuenca – Ecuador, 010203, manuel.villavicencio@ucuenca.edu.ec

Abstract. We propose a new metaheuristic algorithm to find “good” solutions for the assignment of small treatment-control groups, minimising the random resource. Using simulated cases, we achieved 100% groups with equivalence levels equal to or higher than those generated with the simple random assignment, complete random assignment and block random assignment designs. In addition, as a secondary objective to test the new algorithm, we found that short out-of-class essays implied that treatment group marks were 14% higher than in the control group.

Keywords: *Effective Meta-Heuristic Assignment, learning measurement, small treatment and control groups, group equivalence.*

Introduction

In the design of pedagogical experiments, it is common to resort to the use of treatment-control groups. The proper formation of these groups requires attention to some important aspects. One of them is to ensure that these groups are as similar as possible in order to prevent certain factors from hindering the adequate evaluation of a new pedagogical resource.

From a theoretical perspective, the balance between the treatment and control groups is achieved with random methods, which requires large sample sizes.

According to Gertler et al. (2017), a *completely random assignment* of students to the treatment-control groups produces groups with a high probability of being statistically equal (equivalent in averages and distributions in all their characteristics), provided that the set of eligible students is sufficiently large. For Farrington and Ttofi (2009), the size of the groups may not be less than 100 each, if it is desired to achieve realistic pre-test balance.

Unfortunately, in practice, it is not always possible to have a design of this nature, frequent having only a limited population of teachers and students willing to collaborate with research; thus, in this scenario, random resources are not efficient. In this regard, Brown and Harrigan (1983) argue that when you have smaller sample sizes, you are more likely to get large differences between groups simply by chance at the site level than at the model level. Even Suresh (2011) argues, that a completely random assignment with equal probability for all students often leads to severe imbalances.

In this study, we propose an *effective meta-heuristic assignment* (EMA) algorithm that improves equivalence between treatment and control groups when these are small. Additionally, to measure its effectiveness, we use a case study through which we compare it with other assignment methods present in the literature.

Not every new pedagogical resource is more efficient. Education must be adapted to a society that is evolving more and more rapidly, which requires a constant innovation of pedagogical resources to guarantee better results (Gavrilyuk, Tareva, & Lakhno, 2019; Waters, 2009). Each new resource seeks to replace a traditional one in use, which is justified as long as it is more effective, but this does not necessarily happen. Thus, Mendezabal and Tindowen (2018) in a quasi-experimental research design, in which two different groups were compared, found that the use of Microsoft Mathematics in learning calculus is no more effective than the traditional approach. While Jelatu, Sariyasa and Ardana (2018) concluded that the geometry concepts are better learned using the Geo Gebra-aided REACT strategy than with conventional (expository) teaching.

In the two examples cited, as in many others, the good quality of the conclusions has depended on an adequate construction of equivalent experimental groups (Kover & Atwood, 2013) to ensure that the effects of the experiment are basically the results of the didactic resources to be evaluated.

Small groups, in the opposite direction to the law of large numbers

The law of large numbers (probability theory) asserts that the average of the results when performing an experiment, a large number of times tends to be the *expected value* and is closer as more trials are performed. Based on this, in sampling methods, a random resource is more efficient while sample and population grow in the direction of the law of large numbers (i.e. if these decrease, random resources are less efficient). In other words, with small samples, everything indicates that we should go in the opposite direction,

minimizing the random resource by substituting it with heuristics, like blocking or pairing, that tries to strengthen the equivalence of the groups of the experiment. Reynolds and West (1987) state that in practice we find large obstacles to the application of random methods, this being a common scenario in quasi-experiments. Sekhon (2009) states that, in real scenarios, the treatment and control groups are almost never balanced, especially if they are small populations and samples, “*unless something special is done*” and that in our case corresponds to a heuristic matching strategy that we will describe later.

In the literature we find some heuristic proposals that allow us to compensate for the limitations of random resources in small samples. For example, Box et al. (1978) and Imai et al. (2008) state that we must *block* the confounding factors as far as possible and randomize them only when not. This strategy allows the compensation of the limitations of small samples to achieve group equivalence: blocking controls the effects of confounding factors by placing students into groups of similar characteristics. Imai et al. (2008, 486) assert categorically that “Blocking guarantees that the treated and control groups are identical with respect to these covariates so that they cannot affect our inferences.” In any case, blocking methods (also called stratification) work very well if the number of strata is small (Zagoraiou, 2017), as is our case study¹. As Weisburd and Gill (2014) indicate, the blocking procedure improves both the equivalence and statistical power of the experiment groups. These authors demonstrated by simulations that with groups of 28 cases based on blocking procedure, the same or better results in equivalence are achieved (in terms of baseline characteristics) than with groups of 50 formed only with random resources

Another heuristic alternative to a completely random process is *matched pair design*. It consists of matching the students of each course section (students attending the same classroom) with the closest ones based on their characteristics. Then, for each matched pair, one student is randomly assigned to the control group and the other to the treatment group. Researchers such as Li et al. (2014) justify the use of a matched-pair design to optimize randomization, applying it to a case study with small treatment and control groups of size equal to 20. They argue that this increased the potential of achieving balanced intervention groups.

According to the authors, this resource allows the improvement of the unbalances; however, the individual random assignment in each pair still allows unbalances when it comes to small samples, although less significant compared to the completely random assignment. When matched pairs are obtained, the frequent use is a random assignment to the treatment and control groups, as in the case of Weisburd and Gill (2014). A description of other alternatives may be reviewed in Suresh (2011), Grischott (2018) and King, Lengerich and Bai (2018).

¹ Few teachers and course sections decided to collaborate. This problem is a common obstacle in pedagogical research.

Greevy et al. (2004) in their study argue that “blocking or pairing before randomization” is a basic principle of experimental design and propose to improve covariate balance (equivalence) using an algorithm applied to a case-study. Similarly, in our study we propose a meta-heuristic algorithm that allows treatment and control groups to be formed with an improved level of equivalence, which we apply to a case study.

In general terms, this algorithm EMA has four stages: in the first one, traditional blocking resources are used by course section. In a second stage, each block (course section) proceeds to perform student matching according to what we call learning potential; they are sorted ascendingly, then, every two consecutive students are sequentially matched. In a third stage, the formation of two groups not yet assigned to any treatment begins. At this stage, the students of each matched pair are assigned alternately (not randomly) to each of these two groups. In a fourth iterative stage, using a non-random method, the matched pair that best contributes to improve the equivalence of the two groups under construction is located and exchanged. This procedure is repeated until there is no matched pair that allows such improvement.

Once the two groups are constituted, one is randomly assigned to control and the other to treatment². Like Greevy et al. (2004), we will measure the effectiveness of this meta-heuristic assignment through a case study in which short out-of-class essays are applied to improve learning.

In summary, in this study we use a quasi-experimental research design, using a meta-heuristic assignment algorithm that allows to improve the level of equivalence between the treatment and control groups, especially when it comes to small samples. Specifically, this study states that:

Accordingly, *the research object* was an assignment method between the treatment and control groups. *The research question* of this paper asks: if you only have small treatment-control groups, is it possible to improve their equivalence by resorting to metaheuristic assignment processes? *The aim of the paper* is to provide a resource for pedagogical experiments that have only small samples which limits the adequate conformation of treatment-control groups using pure random processes.

The objectives of the paper: 1) The main objective is to verify whether the proposed new meta-heuristic assignment method EMA allows small treatment-control groups to improve their level of equivalence; 2) As a secondary objective, and a case study of the previous objective, we seek to measure whether the use of short out-of-class essays improves student learning.

The conceptual positions are based on the view that if you only have a small sample, insufficient to apply random methods that allow you to properly balance the confounding factors by resorting to the law of large numbers, it is possible to compensate for

² This meta-heuristic algorithm will be explained in detail in the section on Methods.

this equivalence by resorting to metaheuristic methods, this based on a review of the literature and its results.

The research method is quantitative and uses two-group comparison quasi-experimental study conducted in five course sections in which, as a case study, the pedagogical efficiency of short out-of-class essays is evaluated. The study was conducted at a public university in Ecuador.

On the case study: Short essays written out-of-class as a didactic reinforcement across the curriculum

Each didactic resource must be well justified before its time-consuming effectiveness evaluation. In this sense, using a *short out-of-class essay* (SOCE) as a didactic reinforcement may be convenient for the following reasons:

First, writing can support the learning process of curricular content. In other words, students' discursive competence represents their learning in-class, and not their writing ability. Secondly, a writing programme is defined as transversal to the curriculum, so it cannot be based on resources which are not generally applicable to the different topics of all disciplines.

Some studies have evaluated didactic sequences that require significant time from the students and teachers involved. Nurnberg (2017) used a 16-week teaching sequence and evaluated using the Science Writing Heuristic (SWH) approach; Bargate (2015) applied an 18-week writing programme based on the Writing to Learn (WTL) approach; Caukin (2010) performed a 5-week intervention evaluating the SWH strategy. Cohen and Spencer (1993) evaluated the impact of essays between 15 and 20 pages on the class of Economics; Andueza (2016) carried out a didactic sequence of eight sessions with the writing of an article of scientific dissemination as a final product.

In all cases, they found that their proposed writing treatments contributed to meaningful learning. However, long didactic sequences, essays or scientific articles, are types of writing that can serve as a didactic resource to achieve greater depth only in a few topics. In practice, they cannot be applied comprehensively to all topics in all courses because of the excessive (if not impossible) workload for the student and for the teacher who must carry out the respective assessments.

On the contrary, if the SOCE's effectiveness is confirmed, it can be widely applied as a resource to reinforce across-the-curriculum learning. In this respect, according to a survey conducted by the American Economic Association, *short one-minute essays* at the end of the class summarizing what has been learned are often used for this purpose (Bazerman et al., 2016). However, an objective evaluation of the pedagogical resource was not included. In addition, the students' writing for a short period at the end of class implies a rush that would limit the understanding of concepts and the level of critical analysis required. To correct this limitation, we suggest that the short essay should be out-of-class writing.

Regarding the case study, the hypothesis that we propose is that SOCE, independently of its formal part, potentiates the student to achieve greater significant learning.

Methodology

Participant Selection and Sample. In the present study, a quasi-experiment was carried out, selecting for convenience five course sections from the Faculty of Economics with a total of 134 students. We used a convenience sample that included students from five different course sections ranging in size from 24 to 30 students. The data needed to evaluate the treatment's effectiveness were collected at the regular scheduled examination during the term.

In this experiment we obtained the voluntary cooperation of three instructors (P1, P2 and P3), each with different years of teaching experience (17, 18 and 3 years respectively). This voluntary incorporation into a pilot project implies more tolerance to several measures and adjustments that may be required, especially at an early stage.

These instructors cover four courses from the Faculty of Economics with different levels of mathematical requirement: Financial Mathematics (FM), Corporate Finance (CF), Professional Ethics (PE), and Social Programme Evaluation (SPE). The first of these, FM, includes two sections (see Table 1).

Table 1
Characteristics of the course sections

Course	Instructor	Students	Female	Mathematical requirement
FM1	P1	24	15	High
FM2	P1	30	20	High
CF	P2	26	15	medium
PE	P3	28	13	None
SPE	P3	26	17	medium
Total		134	80	

Note: female 59% (sd=0.089)

The treatment (G_p) and control (G_c) groups that were finally obtained for the experiment (67 students each) were made up of half of the students in each of these five course sections. The division of each course section into two sub-groups (SG_{k1} and SG_{k2}) was done through the new meta-heurist method EMA of matched-pair assignment.

Measurement instruments

In order to measure the equivalence of the experimental groups, the student's *grade point averages* (GPA) obtained from the university's administrative databases and using the Student's t-test. For the analysis of the efficiency of the evaluated methods (SOCE), written tests (post-test) applied to the students were required. The comparison of the results was also carried out using the same test. In order to measure the equivalence of the two experimental groups, the GPA of each student was used. This average is justified as a natural synthetic indicator of the confounding factors related to the learning process, which we justify below.

Learning Potential (LP) as a fundamental synthetic Confounding Factor. Weisburd and Gill (2014) explain that the knowledge we have about the individuals involved "can help us create [heuristic] equivalence over the factors that are related to the observed outcomes".

There are both observable and unobservable confounding factors. It is understandable to think that not all of them have the same importance when it comes to influencing the results of the treatment, which in our case is the level of learning. According to Powers and Witmer (1951) and Greevy et al. (2004), of the observable factors, very few proved to be relevant. Thus, for example, Caukin (2010) uses only three factors: gender, socioeconomic status, and GPA. Powers and Witmer (1951), in another context, determine only age and family background as key factors. In our case study, we must analyse which factors are relevant to give them focus in our analysis.

To achieve this, we must first identify the confounding factors that can positively or negatively affect the learning process of a specific curricular topic, in addition to the didactic treatment evaluated. In order to identify these confounding factors, we interviewed and surveyed experienced teachers and students in order to obtain an adequate listing of these factors and a measure of their importance.

The survey inquired about factors that may influence the level of learning of a specific topic in the syllabus of a given subject. The students and teachers expressed more than seventy confounding factors with different levels of incidence, positive in some cases and negative in another, even for the same factor (for example, having their girlfriend or boyfriend as a classmate).

Of those interviewed, 99% stated that these factors are 73% interrelated. The same percentage of respondents said that between these positive and negative factors there are trade-offs in the learning process of 67%. This situation reveals that the teaching-learning process is extremely complex.

Based on the above, it is feasible to think that each student has a certain *learning potential* (LP) depending on the internal and external conditions that surround him/her. This potential could be negatively affected by strenuous salaried work but could be partially compensated by the help of a study group. It could also grow with a high IQ but could be decimated by problems of indiscipline in study habits. A rejection of subjects

with mathematical content could be compensated to some degree with good motivation from the corresponding instructor.

On the other hand, from a psychological point of view, each student has a certain level of resilience, the capacity that living beings have to cope with adverse situations. This is difficult to model since it mitigates the effects of some confounding factors; this can be mitigated up to 67% based on those surveyed.

In this sense, the learning potential of each student LP_i is conditioned by the set X_i of observable and unobservable confounding factors that we have explained above.

$$LP_i = f(X_i).$$

This functional relationship can be very complex due to the links that may exist between many of these factors. However, given the fact that this variable LP_i already synthesizes both observable and unobservable confounding factors, if we had direct measurements of it, we would avoid relying on disaggregated confounding factors. We would also avoid establishing their functional relationship, and then we could more adequately and easily balance the treatment and control groups.

How to obtain measurements of each student's LP_i ? One option would be to teach different curricular topics to a sample of students involved in an experiment under random circumstances of their complex confounding factors; however, this procedure would be very complicated and remarkably costly, thus prohibiting. Nevertheless, the fact that each student throughout his or her student life has already been assessed in several subjects under very diverse internal and external circumstances constitutes a *natural experiment*, like that needed to measure his or her LP_i . We can then assume that students' GPA is a good *proxy* measure of their individual LP_i . In any case, we must note that each LP_i is a cause, and the GPA is an effect.

In summary, from this perspective, in our experiment we must measure whether the methodological reinforcement of SOCE has a positive impact on the level of student learning. To achieve this, we must fundamentally control the students' LP_i as a synthetic confounding factor to isolate the effect of treatment and facilitate its measurement.

The counterfactual problem

It is frequent in pedagogical experiments to test whether a new teaching method is better than one traditionally used, under the assumption that with both methods there is always some level of learning. In our case study, the new method ($T=1$) differs from the traditional method ($T=0$) only in that it includes a didactic reinforcement constituted by the addition of SOCEs.

$T_i \in \{0, 1\}$: Binary treatment, 0= without SOCE, 1= with SOCE.

$Y_i(0)$: LP_i level (outcome) for student i without SOCE,

$Y_i(1)$: LP_i level for student i with SOCE (treatment).

The impact of the treatment (T), or treatment effect, is the result of the difference Δ between the learning level $Y_i(1)$ through the method ($T=1$) which includes SOCE, minus the learning level $Y_i(0)$ through the traditional method ($T=0$) (Iacus, King, & Porro, 2018).

$$\Delta_i = (Y_i(1) | T = 1) - (Y_i(0) | T = 0).$$

But this requires that, for each student i , the learning level $Y_i(0)$ should be measured with the traditional method, and simultaneously the learning level $Y_i(1)$ with the treatment method, although addressing the same topic is obviously not possible.

If we have the measurement of the learning level with treatment, we cannot know, for the same student and at the same time, what his learning level would have been, if he had learned under the traditional method, i.e., a term known as counterfactual. This impossibility is what is called a *counterfactual problem*. “Since it is not possible to observe the counterfactual directly, it is necessary to *estimate it*” (Gertler et al., 2017).

According to Gertler et al. (2017), one option for solving the counterfactual problem is to replace its individual character with a counterfactual at the group level, i.e., although each student in the treatment group does not have another student in the control group whose characteristics are equal, one can choose to establish a control group that as a whole is statistically *equivalent* to the characteristics of the treatment group. However, this option requires a sufficiently large number of students, which in practice is not always feasible. In fact, this limitation results in mostly assessments of pedagogical programmes with quasi-experimental designs in the literature, such as Caukin (2010) with an average of 10 students per course section (3), Nurnberg (2017) with an average of 32 students per course section (4) and Schmidt (2004) with an average of 22 students per course section (4). In our case study we have an average of 27 students per course section (5).

Experimental Design

This is a quasi-experimental study, the most frequent type in pedagogical research, conditioned by the circumstances described above. Despite being quasi-experimental, it is a prospective impact assessment (Gertler et al., 2017). This study focuses on the conformation of treatment-control groups, particularly on the improvement of the assignment process by measuring their efficiency through t-test and resorting to the *learning potential* as a synthetic indicator of the confounding factors of this pedagogical experiment. Subsequently, we adopt post-test design variant to evaluate the effect of SOCE. To improve the balance between the experimental groups we will initially use a process of blocking by course sections and subsequently matching pairs within course sections.

Blocking some confounding factors by course section

It is not always possible to find two sections that are experimentally equivalent between them. Frequently, these differ in the type of subject and teacher, in number and type of students, in timetables and other confounding factors that hinder a balance in

the groups. To address these constraints, blocking can be applied by splitting the total sample into blocks corresponding to each of the course sections, as does Caukin (2010), Nurnberg (2017) and Schmidt (2004), and then split each of those course sections into two experimental subgroups. With this initial step, it is possible to control the influence of some confounding factors shared by each course section (block) such as the differences between teachers, subjects, timetables, the topics discussed, etc.

In an academic experiment, one of the most important confounding factors represents the courses used. Students differ in their preferred subjects influenced by different reasons, such as sympathy or resistance to mathematical content. This fact can distort the assessment of a didactic treatment if this covariate is not controlled. Another relevant confounding factor represents the different teaching methods of each instructor as well as his/her personal characteristics including level of knowledge, didactic experience, temperament, and even *beliefs* as held by Randler and Bogner (2008).

In order to control the effects of the confounding factors described above, as a blocking mechanism, we divided each course section into two equal subgroups, which were subsequently assigned one to treatment and the other to control. This blocked assignment additionally allowed the two subgroups to be evaluated simultaneously with the same rubric within each course section.

With this assignment, the treatment and control groups are constituted equally by students of all course sections in different proportions, which is what usually happens, according to King, Lengerich and Bai (2018): 11 students from FM1, 15 from FM2, 13 from CF, 24 from PE and 13 from SPE, totalling 67 students in each group, which does not affect the equivalence pursued.

On the matched pair assignment within course sections

Once some confounding factors external to the student have been blocked through stratification by course sections, it is necessary to review the assignment process within each course section, considering the small size of these strata.

At this stage, the most common alternative is randomization (Suresh, 2011). If this is done properly, the causes of the treatment effect will probably not be at the beginning of the experiment (Shadish, Cook, & Campbell, 2002), especially if the sample size grows and reaches sufficiently large sizes (Graham et al., 2014; Sekhon, 2009). However, the common small number of students within each course section weakens the equivalence between covariates related to student characteristics which can be strengthened by matching pairs.

Matched pair design

The objective of this second stage is to divide each block (course section) into two equivalent subgroups (SG_{k1} y SG_{k2}) on a non-random way.

Equivalence through optimization

Agreeing with Greevy et al. (2004) and Weisburd and Gill (2014), the first measure of covariate imbalance (Preliminary Tests of Equivalence) was simply the absolute value of the difference in means.

From the perspective of operational research, the optimal equivalence can be obtained by considering two objective functions: on the one hand, minimizing the difference (*Diff*) between the averages of learning potentials (\overline{LP}_s) of each $s \in \{1,2\}$ subgroup:

$$\min Diff = |\overline{LP}_1 - \overline{LP}_2| \quad \text{objective function (1)}$$

However, this is not enough as this objective function can lead us to subgroups with similar means, but with different dispersions (standard deviation). This implies that one group could include only *LP* students very close to the average (low dispersion), while the second would include at the same time students with very high *LP* and others with very low *LP* (high dispersion). Therefore, it is also required to minimize the difference between their *LP* variances.

$$\min Diff = |VAR(LP_1) - VAR(LP_2)| \quad \text{objective function (2)}$$

On the other hand, it is convenient to balance the size n_s of the two subgroups by establishing their equality as a restriction:

$$n_1 = n_2 \quad \text{constraint (1)}$$

Evidently, each student i can only belong to one of the two groups:

$$g_{i1} + g_{i2} = 1 \quad g_{i1}, g_{i2} \in \{0,1\} \quad \text{constraint (2)}$$

is the belonging of student i to group 1 or 2; therefore, it can only be an integer 0 or 1.

Then the objective functions (1) and (2) can be expressed by (3) and (4) respectively:

$$\min Diff = \left| \frac{1}{n} \sum_{i=1}^n LP_i g_{i1} - \frac{1}{n} \sum_{i=1}^n LP_i g_{i2} \right| \quad \text{objective function (3)}$$

$$\min Diff = \left| \frac{1}{n-1} \left(\sum_{i=1}^n (LP_i g_{i1} - \overline{LP}_1)^2 - \sum_{i=1}^n (LP_i g_{i2} - \overline{LP}_2)^2 \right) \right| \quad \text{objective function (4)}$$

To solve this model requires *nonlinear programming, integer and multi-objective*, all the less desirable scenarios for classical optimization techniques, which forces the development of metaheuristic techniques, which do not necessarily lead to the optimal solution, but get “good solutions.”

Metaheuristic algorithm to balance treatment and control groups

In the case of small samples, the proposed algorithm attempts to include all the students in each course section, except in the case of odd sized sections in which only one student is randomly excluded, so all course sections are of size $2n$ which allows n pairs of students to be established using the following algorithm:

Step 1: Sort the LP_i

The list of students is sorted in ascending order according to their LP , establishing the ordered set:

$$\{LP_1, LP_2, LP_3, \dots, LP_{2n}\} \quad \text{where}$$

$$LP_1 \leq LP_2 \leq LP_3 \leq \dots \leq LP_{2n}$$

Step 2: Unconditional sequential matching

Pairs of students are established by sequentially taking unconditionally every two sequential students from the ordered list:

$$\{LP_1, LP_2, LP_3, \dots, LP_{2n}\} \Rightarrow (LP_1, LP_2), (LP_3, LP_4), (LP_5, LP_6) \dots (LP_{n-1}, LP_n)$$

It is possible that LP_2 is a value closer to LP_3 (even equal) than its pair LP_1 ; when that is the case, there is the temptation to pair LP_2 with LP_3 (which decreases the variance of the pair). However, this would force LP_1 to pair with LP_4 , or a more distant LP_i , increasing the variance more. In this sense, this would affect the homogeneity between counterfactuals at the individual level (pairs).

The proposed unconditional sequential matching provides more homogeneous individual counterfactuals. This strategy tends to a better balance between variances because extreme values (outliers) are very strange in GPA records. To verify this behaviour, we performed 1000 simulations of which 100% confirmed this behaviour.

Step 3: Assignment algorithm from pairs to groups G_T y G_C

Of each pair of students, one is then assigned to the treatment group and the other to the control group using two alternatives (see *Table 2*):

Table 2
Student Assignment Patterns

Simple alternating		Double alternating	
S ₁₁		S ₁₁	
	S ₁₂		S ₁₂
S ₂₁			S ₂₁
	S ₂₂	S ₂₂	
S ₃₁		S ₃₁	
	S ₃₂		S ₃₂

- a. In a simple alternating pattern: from all pairs the first student to group 1 and the second to group 2.

- b. In a double alternating pattern: from the first pair, the first student to group 1 and the second student to group 2; in the next pair, the first student to group 2 and the second student to group 1, so on and so on.

From the simulations carried out, it could be seen that the single alternating assignment better balances the variance of the groups, while the double alternating assignment better balances the means. In this sense, the most relevant parameter (mean or variance) should be established in each case. However, when the number of pairs grows, these differences diminish, but not enough because the course sections in the sample are small.

Step 4: Adjusting the balance of LP averages using pair permutations

The balance achieved with the unconditional assignment explained above can be improved with an additional procedure:

- a. Check the difference $Diff_g = \overline{LP}_1 - \overline{LP}_2$ between the subgroups. To decrease this difference, students of a given pair j (matched in step 2.) can be exchanged among the assigned subgroups, provided that the condition in the next point (b.) is met.
- b. Considering that the permutation of the students of a pair j will alter the values of the means and standard deviations of the subgroups, whenever their LP_i are different, it is possible to analyse which pair should be permuted to diminish the difference $Diff_g$.
- c. When the pair j is permuted, \overline{LP}_1 decreases LP_{j1}/n and increases LP_{j2}/n , while \overline{LP}_2 decreases LP_{j2}/n and increases LP_{j1}/n , thus the new mean difference between groups is:

$$Diff_g^{+1} = \left(\overline{LP}_1 - \frac{LP_{j1}}{n} + \frac{LP_{j2}}{n} \right) - \left(\overline{LP}_2 - \frac{LP_{j2}}{n} + \frac{LP_{j1}}{n} \right) = (\overline{LP}_1 - \overline{LP}_2) - 2 \left(\frac{LP_{j1} - LP_{j2}}{n} \right)$$

where is the difference between the LPs of permuted pair j .

Determine the pair j corresponding to the minimum difference. Step (b.) must be repeated as long as there is some pair j whose difference is minimal and also meets the condition:

- a. In each iteration, the means and must be recalculated, whose difference $Diff_g$ will be reduced in absolute value or module. Iterations cease when there is no longer a pair that meets the previous condition in (c.).

With the above algorithm, the formation of two equivalent subgroups in each course section is achieved through a metaheuristic strategy. The final step is limited to randomly determining which subgroup will receive the treatment; therefore, the remaining subgroup will be the control one.

Effectiveness test using simulations

To test the effectiveness of this proposed method in group balancing, we performed 1000 simulations based on the data from our case study and compared with the assignments by EMA and three other methods that are very common in the literature and have commands in R: *Simple random assignment (SRA)*, *Complete random assignment*

(CRA) and *Block random assignment* (BRA). The commands used were `simple_ra()`, `complete_ra()` and `block_ra()` respectively, available in the package “randomizr” in R. The results showed that in 100% of the cases, the new assignment method produces smaller or equal differences than the other methods.

Additional Resources for Assessing Treatment Outcomes

Didactic sequence

Once the two subgroups were established in each course section, the instructors proceeded with the didactic sequence that consisted of the following steps:

- 1) A topic of each subject was selected and included in the experiment.
- 2) Each selected topic was taught in each course section jointly to the two subgroups. The expected learning was summarized in an adequate conceptualization and in a critical analysis of them.
- 3) At the end of each class, the students of the treatment subgroup were instructed to write a short out-of-class essay on the topic covered. The essay would first include an explanation of the concepts addressed and then a personal (critical) analysis of them. They were given a deadline and the length was two pages. In order to avoid copying as much as possible, they were asked to write by hand.
- 4) At the end of the class, each instructor published the list of students who had to write the short essay.
- 5) Once the deadline was met, the out-of-class essay was received for assessment and feedback.

Rubric

This experiment aimed to be the least invasive in instructors’ current methods. To this effect, the evaluation of the level of learning of the two groups of each selected topic was carried out within the examinations programmed in each syllabus. Since the essays were short critical writings, the rubric was meant to assess 50% of the student’s understanding of the concepts learned and 50% the student’s critical analysis of these concepts.

Handling missing data

Of the students selected, only three of them did not take the scheduled examinations. In this case, in order not to unbalance the groups, we proceeded to fill the missing marks with the GPA of their academic record, under the assumption that, if they had taken the examination, it was very likely that they would obtain a grade similar to their grade average. One could also opt to follow the suggestion of Ho et al. (2011) who propose to complete the data with the average of the students’ GPA of the subgroup to which the student belongs; however, the average imputation is highly questioned (Acock, 2005). In

any case, we consider that the first option is the most accurate to fill missing data from this study under the assumption that the student with the missing grade tends more to his/her own GPA than to the grade average of a group of students.

Results

An objective of this investigation is to verify if for the case study, the method we propose is a good alternative to the designs Simple random assignment (SRA), Complete random assignment (CRA) and Block random assignment (BRA) for which we have made simulations of assignments using the corresponding commands in R.

As can be seen in Table 3, considering the last simulation, only the SRA has no balance between the group sizes (fourth column). The three methods show differences in the mean of the G_c and G_t groups, except for the proposed assignment method (fourth column).

The last column presents the absolute values of the mean differences in the simulations performed. The three “traditional” assignment methods present differences with average values between 0.46 and 0.48, while using EMA is 0.007, a significantly lower value. The standard deviations are in parentheses, between 0.35 and 0.36 for “traditional” assignments, whereas in the case of the proposed new method it is zero, because it does not depend on random factors, and therefore the differences are always the same.

Table 3
Equivalence by type of assignment (simulations)

Design	Last simulation				All simulations
	Treat_<fct>	n_students<int>	mean_avg<dbl>	std_error<dbl>	Average of abs(diff) *
Simple random assignment	0	62	73.4	0.427	> mean(simple_ests)
	1	72	72.5	0.409	0.4814147 (0.37266)**
Complete random assignment	0	67	73.7	0.452	> mean(complete_ests)
	1	67	72.1	0.365	0.4958254 (0.36453)
Block random assignment	0	67	73.4	0.468	> mean(block_ests)
	1	67	72.4	0.362	0.4601236 (0.34606)
EMA	0	67	72.9	0.405	> mean(new_ests)
	1	67	72.9	0.439	0.0073134 (0.00000)

* Mean of the absolute value of the differences in the simulations, between means of “avg” in the treatment and control groups. ** Standard deviations in parentheses.

Balancing results

With the proposed matching method, a balance was achieved in the size of the groups G_T and G_C ($n_students$). On the other hand, there is no difference (or it is very small) between the averages of their learning potentials ($mean_avg$) as well as low dispersions (std_error) (see Table 4 y Figure 1).

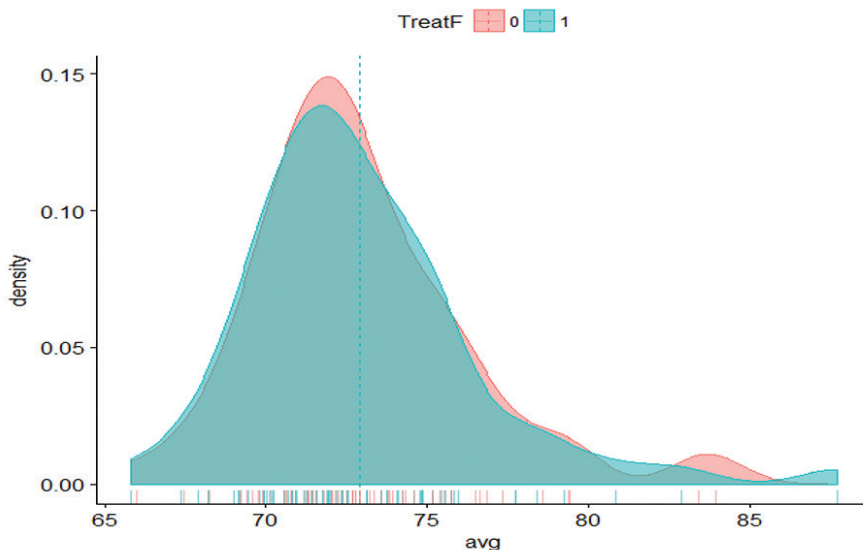


Figure 1. Balance of learning potential (LP) between treatment and control group

Table 4

Overall balance

TreatF	n_students	mean_avg	std_error
0	67	72.9	0.405
1	67	72.9	0.439

At the level of each course section (stratum) the learning potential is also highly balanced (see Table 5).

Table 5

Balance by course section with the proposed method EMA

Course	TreatF	n_students	mean_avg	std_error
1	0	12	73.00	1.660
1	1	12	73.00	1.860
2	0	15	72.10	0.464
2	1	15	72.10	0.460
3	0	13	73.20	0.810
3	1	13	73.20	0.875
4	0	14	71.90	0.583
4	1	14	71.90	0.545
5	0	13	74.70	0.741
5	1	13	74.70	0.836

Statistical assessment of the difference

According to Ho et al. (2011), the nearest and optimal methods seek to obtain the minimum difference between the averages of the two groups (G_T and G_C), which coincides with our main objective. However, these authors, as well as Amusa (2018), warn that “the widely used procedure of doing t-tests of the difference in means is highly misleading and should never be used to assess balance.” In any case, applying paired t-test, we obtained p-value 0.93 (>0.05); thus, we can conclude that the means of the groups G_T and G_C are significantly similar, which implies having reached a very high balance between the confounding factors synthesized in the learning potential (see Table 6):

Table 6

Balance: Paired t-test

data: avg by TreatF
t = -0.087996, df = 66, p-value = 0.9301
95 percent confidence interval: -0.1732494 0.1586226
sample estimates: mean of the differences -0.007313433

Treatment Outcomes

The examinations in each course section were administered jointly to all the students, and they are evaluated with the same rubric. From the treatment groups, it can be observed that, in general, the students who wrote the short essays (treatment) scored an average of 8.4 points more than those who learned with the “traditional” method. This implies having a positive impact of 14% (see Table 7).

Table 7
Means of Treatment Evaluations (test)

	Treat F	n_students	mean_test	std_error
1	0	67	60.2	4.16
2	1	67	68.6	3.22

At the course section level, the percentages of increase were not homogeneous. In the first course section of the list the results increased by 16%; in the second course section they only increased by 2.4%; the third course section presented the greatest impact since the results increased by 160%; in the fourth course section the increase was only 8%; and in the fifth course section there was a negative impact of 4.7% (see Table 8) .

Table 8
Averages of treatment (test) evaluations per course section

	Course	TreatF	n_students	mean_test	std_error
1	1	0	12	59.8	9.29
2	1	1	12	69.4	6.75
3	2	0	15	78.1	7.62
4	2	1	15	80	5.18
5	3	0	13	18.3	6.43
6	3	1	13	47.5	9.47
7	4	0	14	72.4	8.96
8	4	1	14	78.2	6.50
9	5	0	13	68.6	2.70
10	5	1	13	65.4	4.70

According to the t-test, with paired data, the mean difference was 8.41 points which, statistically presents a p_value of 0.057, a value slightly higher than 0.05, so we can assert that these mean scores are not statistically equal under the 6% CI (see Table 9).

Table 9

Treatment assessment: Paired t-test

data: testscore by Treat

t = -1.9373, df = 66, p-value = 0.05699

94 percent confidence interval: -16.71985 -0.10242
--

sample estimates: mean of the differences -8.411137

Graphically (see Figure 2) we can observe that the two modes are on very close values of the scores (testscore), the frequency of the students of the treatment group being greater. This behaviour is a consequence of the better performance of the treatment group.

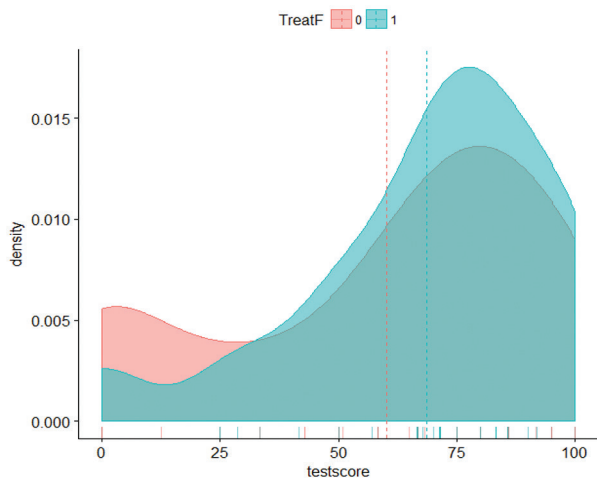


Figure 2. Comparison of treatment results

Finally, in order to test whether the treatment has a better effect on students with a higher or lower GPA, we performed a linear regression of the difference between the pairs ($Diff = \beta_0 + \beta_1 * avg$, being $Diff = Y(1) - Y(0)$) in relation to the academic average (avg). The results showed that $\beta_1 = 0.29$, implying that the treatment is more efficient the higher the GPA the student has, although the coefficient of this variable (avg) was not statistically significant ($p\text{-value} = 0.83$), i.e. there is no relationship between $Diff$ and (avg), which can be interpreted as indicating that the treatment improves in the same way all students regardless of their academic record.

Discussion

The common presence of quasi-experimental studies in pedagogical issues is favored to a large extent by the limited willingness of institutions, students, and even families to

collaborate in studies of this nature. In these scenarios, the basically random methods of experimental group formation encounter obstacles and, as suggested by Sekhon (2009), we must do something more in order to compensate for these limitations.

Weisburd and Gill (2014) suggest strengthening weakened quasi-experimental random resources through heuristic resources, such as block randomized design. Li et al. (2014) justify the use of a matched-pair design for the same purpose. However, there is concern that some quasi-experimental studies, such as those of Mendezabal and Tindowen (2018) and Rosário et al. (2017), use only random assignment without carrying out a pre-test evaluation of the degree of equivalence between their experimental groups, which in this type of design should be necessary.

The main objective of our study was to verify if the proposed Effective Metaheuristic Assignment (EMA) allows the improvement of the degree of equivalence between treatment and control groups in small samples with greater effectiveness than the heuristics of blocking and pairing.

Within the pedagogical field, in order to be able to measure and compare the EMA's effectiveness, we justify the use of what we call *learning potential* (LP) as a synthetic factor that summarizes observable and unobservable confounding factors, simplifying the modelling of the current didactic experiment. This synthetization is important because, according to the students, these factors have complex relationships between them in addition to their trade-offs, and particularly because of the levels of psychological resilience and homeostasis that make it more difficult to model and measure them.

The pre-test balance evaluation showed that the proposed EMA is highly efficient in getting balanced treatment and control groups (i.e. to start with equal averages of the confounding synthetic factor *LP*). This equivalence was achieved even at the level of course sections (strata). In our case study, the results showed that 100% of the simulations conducted achieved the same or better levels of equivalence than the Simple Random Assignment, Complete Random Assignment and Block Random Assignment procedures, very frequent in the literature.

In this case, we agree with the conclusions of Suresh (2011), Weisburd and Gill (2014), Li et al. (2014) who, from different approaches, argue that when small samples are available it is convenient to improve random allocations through various heuristic resources proposed. We have not found any studies that support the opposite in these scenarios.

On the other hand, the results obtained confirm the internal validity of the EMA; however, the convenience sample used weakens its external validity, which, agreeing with Gertler et al. (2017), can be improved later by resorting to a random and greater representative sample of course sections.

As a secondary objective, we seek to measure the pedagogical efficiency of short out-of-class essays in the process of learning. We consider that, for the case study, this hypothesis is confirmed according to the empirical evidence, since the evaluations carried out show that the group of students who received this didactic reinforcement obtained

an evaluation 14% above the control group. However, due to the dispersion of the scores of each group, this difference is barely statistically significant according to the t-test, although we must consider that for these cases this test can be questioned by Amusa's arguments (2018).

This result agrees with the study conducted by the American Economic Association (Bazerman et al., 2016), although the short essays used in that experiment were in-class rather than out-of-class. This suggests that writing short essays on a topic being studied always contributes to students' learning. In the literature review, no other pedagogical studies were found to evaluate the impact of SOCEs on other case studies with what we can compare, which restricts this discussion.

The information collected does not allow us to clearly explain the notable difference in results between course sections; however, the causes could be related to a combination of differences in the instructors' teaching experience and methods, as well as in the complexity of each topic and the level of difficulty of their examinations. In any case, although it does not appear in the objectives of this research, we recognize that it will be necessary to go deeper into this aspect.

On the other hand, although we proposed to use the *overall* grade point average as a proxy measure of each student's PL, to apply the experiment to any subject, we believe we can improve this factor to balance the experimental groups by using the grade point average only of the subjects related to the topic taught (prerequisites), coinciding in this aspect with what is proposed by Mendezabal and Tindowen (2018), although differing in the assignment method, but this requires further assessment.

Finally, EMA assessment can be extended to areas other than education where it is not sufficient or possible to use PL as a single confounding factor. In this case, a synthetic indicator can be established that aggregates the corresponding confounding factors, as simple as arithmetic mean or geometric mean, or complex as Mahalanobis distance or P2-distance, which requires addressing some theoretical drawbacks of aggregation related to the justification of weights, aggregation methods and more.

Conclusions

In conclusion, EMA was shown to be more effective than three other heuristic assignment methods seeking to improve randomisation in small experimental treatment and control groups. Therefore, for experiments in the pedagogical field in which the essential conditions for the application of traditionally random methods of assignment are not fulfilled, those meta-heuristic methods that better manage the experimental limitations are valid. In particular, we demonstrate that the one we propose is highly effective.

The limitations of these results correspond to those arising from a quasi-experiment, which could be improved, especially in its external validity, if we were provided with the

institutional collaboration of the case study to allow us to have an adequate sample size of teachers and course sections, promoting generalization.

In agreement with Mendezabal and Tindowen (2018), teachers should engage students in meaningful learning by diverse resources that enhance their learning, but previously evaluating the effectiveness of these pedagogical resources through techniques that guarantee the quality of the results, a task in which EMA can contribute. This method applied to our case study has allowed us to conclude in a more confident way that short essays out of class enhance learning.

Acknowledgement

This research is supported by the Research Department of the University of Cuenca (DIUC).

References

- Acock, C. A. (2005). Working With Missing Values. *Journal of Marriage and Family*, 67(4), 1012–1028. doi:<https://doi.org/10.1111/j.1741-3737.2005.00191.x>.
- Amusa, L. B. (2018). Reducing bias in Observational Studies: An Empirical Comparison of Propensity Score Matching Methods. *Turkiye Klinikleri Journal of Biostatistics*, 10(1), 13–26. doi:10.5336/biostatic.2017-58633.
- Andueza, A. (2016). La escritura como herramienta de aprendizaje significativo: un cuasiexperimento en la clase de ciencias. *Revista Complutense de Educación*, 27(2), 653–668. doi:http://dx.doi.org/10.5209/rev_RCED.2016.v27.n2.46918.
- Bargate, K. (2015). Effects of a Writing-to-Learn Tutorial Programme in a Higher Education Setting: An Interactive Qualitative Study. *Africa Education Review*, 12(3), 493–507. doi:<https://doi.org/10.1080/18146627.2015.1110914>.
- Bazerman, C., Little, J., Bethel, L., Chavkin, T., Fouque, D., & Garufis, J. (2016). *Escribir a través del Currículum. Una guía de referencia*. Córdoba: UNC.
- Box, G. E., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for Experimenters*. New York: Wiley-Interscience.
- Brown, R. S., & Harrigan, M. (1983). *The comparability of treatment and control groups at randomization*. Obtenido de U.S. Department of Health and Human Services: <https://aspe.hhs.gov/system/files/pdf/73976/compares.pdf>.
- Caukin, N. S. (2010). *Science Writing Heuristic: a Writing-to-Learn Strategy and Its effect on Students' Science Achievement, Science self-efficacy, and Scientific Epistemological View*. Tennessee State University. (Dissertation).

- Cohen, A. J., & Spencer, J. (1993). Integrating Writing Across the Curriculum into Economics: A Case Study, Benefits, and Costs. *Journal of Economic Education*, 24, 219–230.
- Farrington, D. P., & Ttofi, M. M. (2009). *School-Based Programs to Reduce Bullying and Victimization*. Retrieved from NCJRS: <https://www.ncjrs.gov/pdffiles1/nij/grants/229377.pdf>
- Gavrilyuk, O. A., Tareva, E. G., & Lakhno, A. V. (2019). Investigating the Association Between University Teachers' Professional Autonomy and Their Innovation Performance. *Pedagogika*, 133(1), 128–148. doi:<https://doi.org/10.15823/p.2019.133.7>.
- Gertler, P. J., Martínez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2017). *Impact Evaluation in Practice*. Washington DC: World Bank Group. doi:10.1596/978-1-4648-0779-4.
- Graham, J. W., Pettigrew, J., Miller-Day, M., Krieger, J. L., Zhou, J., & Hecht, M. L. (2014). Random Assignment of Schools to Groups in the Drug Resistance Strategies Rural Project: *Some New Methodological Twists*. *Prevention Science*, 15(4), 516–525. doi:10.1007/s11121-013-0403-9.
- Greevy, R., Lu, B., Silber, J. H., & Rosenbaum, P. (2004). Optimal Multivariate Matching before Randomization. *Biostatistics*, 5(2), 263–275.
- Grischott, T. (2018). The Shiny Balancer -Software and Imbalance Criteria for Optimally Balanced Treatment Allocation in Small RCTs and cRCTs. *BMC Medical Research Methodology*, 1–11. doi:<https://doi.org/10.1186/s12874-018-0551-5>.
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, 42(8), 1–28.
- Iacus, S. M., King, G., & Porro, G. (2018). A Theory of Statistical Inference for Matching Methods in Causal Research. *Political Analysis*, 27, 46–68. doi:10.1017/pan.2018.29.
- Imai, K., King, G., & Stuart, E. A. (2008). Misunderstandings between Experimentalists and Observationalists about Causal Inference. *J. R. Statist. Soc. A*, 171(2), 481–502.
- Jelatu, S., Sariyasa, & Ardana, I. M. (2018). Effect of GeoGebra-Aided REACT Strategy on. *International Journal of Instruction*, 11(4), 325-336. doi:<https://doi.org/10.12973/iji.2018.11421a>.
- King, T., Lengerich, R., & Bai, S. (2018). *Design and Analysis of Clinical Trials*. Retrieved from PennState Eberly College of Science: [https://newonlinecourses.science.psu.edu/stat509/node/67/\[2019.01.15\]](https://newonlinecourses.science.psu.edu/stat509/node/67/[2019.01.15]).
- Kover, S. T., & Atwood, A. K. (2013). Establishing Equivalence: Methodological Progress in Group-Matching Design and Analysis. *American Journal on Intellectual and Developmental Disabilities*, 118(1), 3–15. doi:<https://doi.org/10.1352/1944-7558-118.1.3>.
- Li, L., Liang, L.-J., Wu, Z., Lin, C., & Guan, J. (2014). Assessing Outcomes of a Stigma-reduction Intervention with Venue-based Analysis. *Soc Psychiatry Psychiatr Epidemiol*, 49, 991–999. doi:10.1007/s00127-013-0808-6.
- Mendezabal, M., & Tindowen, D. (2018). Improving Students' Attitude, Conceptual Understanding and Procedural Skills in Differential Calculus Through Microsoft Mathematics. *JTSE*, 8(4), 385–397. doi:<https://doi.org/10.3926/jotse.356>.
- Nurnberg, D. (2017). *Writing-to-Learn in High-School Chemistry: The Effects of Using the Science Writing Heuristic Increase Scientific Literacy*. (Doctoral dissertation): The University of San Francisco.

- Powers, E., & Witmer, H. (1951). *An Experiment in the Prevention of Delinquency*. New York: Columbia University Press.
- Randler, C., & Bogner, F. X. (2008). Planning Experiments in Science Education Research: Comparison of a Quasi-Experimental Approach with a Matched Pair Tandem Design. *International Journal of Environmental & Science Education*, 3(3), 95–103.
- Reynolds, K., & West, S. (1987). A Multiplist Strategy for Strengthening Nonequivalent Control Group Designs. *Evaluation Review*, 11(6), 691–714.
- Rosário, P., Högemann, J., Núñez, J. C., Vallejo, G., Cunha, J., Oliveira, V., Fuentes, S., Rodrigues, C. (2017). Writing Week-journals to Improve the Writing Quality of Fourth-graders' Compositions. *Read Writ*, 30, 1009–1032. doi:10.1007/s11145-016-9710-4.
- Schmidt, L. A. (2004). Evaluating the Writing-to-Learn Strategy with Undergraduate Nursing Students. *Journal of Nursing Education*, 4(10), 466–473.
- Sekhon, J. S. (2009). The Neyman–Rubin Model of Causal Inference and Estimation via Matching Methods. In J. M. Box-Steffensmeier, H. E. Brady, & D. Collier (Eds.), *The Oxford Handbook of Political Methodology*. Oxford Handbooks. doi:10.1093/oxfordhb/9780199286546.003.0011.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company.
- Suresh, K. (2011). An Overview of Randomization Techniques: An Unbiased Assessment of Outcome in Clinical Research. *Journal of Human Reproductive Sciences*, 4(1), 8–11. doi:10.4103/0974-1208.82352.
- Waters, A. (2009). Managing innovation in English language Education. *Language Teaching*, 4, 421–458. doi:doi:10.1017/S026144480999005X.
- Weisburd, D., & Gill, C. (2014). Block Randomized Trials at Places: Rethinking the Limitations of Small N Experiments. *J Quant Criminol*, 30, 97–112. doi:10.1007/s10940-013-9196-z.
- Zagoraïou, M. (2017). Choosing a Covariate-adaptive Randomization Procedure in Practice. *Journal of Biopharmaceutical Statistics*, 27(5), 845–857. doi:https://doi.org/10.1080/10543406.2017.1289944.

Efektyvus metaeuristinis priskyrimas siekiant pagerinti ekvivalentiškumą mažose eksperimentinėse pedagoginėse grupėse

Diego Roldán¹, Víctor Aguilar², Francisco Roldán³, Manuel Villavicencio⁴

¹ Kuenkos universitetas, Ekonomikos fakultetas, Av. 12 de Abril y Av. Loja, Kuenka – Ekvadoras, EC-010203, diego.roldan@ucuenca.edu.ec

² Kuenkos universitetas, Ekonomikos fakultetas, Av. 12 de Abril y Av. Loja, Kuenka – Ekvadoras, EC-010203, victor.aguilar@ucuenca.edu.ec

³ Kuenkos universitetas, Ekonomikos fakultetas, Av. 12 de Abril y Av. Loja, Kuenka – Ekvadoras, EC-010203, roldan.arauz@yahoo.com

⁴ Kuenkos universitetas, Filosofijos fakultetas, Av. 12 de Abril y Av. Loja, Kuenka – Ekvadoras, EC-010203, manuel.villavicencio@ucuenca.edu.ec

Santrauka

Norint, kad eksperimentinės tiriamosios ir kontrolinės grupės būtų tinkamai formuojamos, pageidautina, kad jos būtų pakankamai didelės, kad tinkamai atspindėtų populiaciją ir būtų pritaikytos tikimybinei klaidinančių veiksnių poveikio kontrolei naudojant atsitiktinius metodus. Vis dėlto praktiškai tai padaryti ne visada įmanoma. Iš tikrųjų literatūroje dažnai aptinkami kvaziekspertimentiniai tyrimai su mažomis imtimis. Tokiais atvejais, pasak skirtingų autorių, atsitiktinių metodų nepakanka atitinkamam ekvivalentiškumui tarp grupių pasiekti. Taigi, priskyrimas pagerinamas remiantis euristiniais metodais, tokiais kaip blokavimas ar poravimas.

Šio tyrimo pagrindinis tikslas – pateikti efektyvų metaeuristinį algoritmą, kuris padėtų rasti „gerus“ sprendimus mažose tiriamosiose ir kontrolinėse grupėse ir sumažintų atsitiktinius resursus. Naudojant imituotus atvejus, pasiekta 100 proc. grupių su ekvivalentiškumo lygiais, tokiais pat ar didesniais nei tie, kurie buvo sugeneruoti naudojant paprastą atsitiktinį priskyrimą, visišką atsitiktinį priskyrimą ir blokinį atsitiktinį priskyrimą. Papildomai, kaip antrinis tikslas ir atvejo tyrimas, pasiūlytam algoritmui įvertinti buvo apibrėžtas trumpos užklausinės esė vaidmuo ekonomikos mokymosi procese. Šie trumpi tekstai (esė), kaip metodologiniai ištekliai, gali būti dažnai naudojami įvairiems dalykams. Šiame pedagoginio eksperimento atvejo tyrime trumpos esė leidžia suprasti, kad tiriamojoje grupėje pažymiai buvo 14 proc. aukštesni nei kontrolinės grupės.

Esminiai žodžiai: *efektyvus metaeuristinis priskyrimas, mokymosi įvertinimas, mažos tiriamosios ir kontrolinės grupės, grupių ekvivalentiškumas.*

Gauta 2019 05 06 / Received 06 05 2019
Priimta 2019 10 15 / Accepted 15 10 2019