Teaching and Learning of Introduction to Software Engineering Experimentation to Distance-Learning Students: a Quasi-Experiment

1st Carlos D. Luz
Informatics Department
State University of Maringá
Maringá, Brazil
carlos.danilo.luz@gmail.com

2nd Elaine I. Moreira
Informatics Department
State University of Maringá
Maringá, Brazil
elaine.ignaciomoreira@gmail.com

3rd Nelson Tenório
Informatics Department
State University of Maringá
Maringá, Brazil
nelson.tenoriojr@gmail.com

4th Edson OliveiraJr
Informatics Department
State University of Maringá
Maringá, Brazil
edson@din.uem.br

5th Ellen F. Barbosa
Computing Department
University of São Paulo
São Carlos, Brazil
francine@icmc.usp.br

***Abstract*—This research paper describes a controlled quasi-experiment conducted with 24 distance-learning undergraduate students on teaching introduction concepts of Experimentation in Software Engineering (ESE) with different materials and procedures. The use of experimentation is a *de facto* way of obtaining data to statistically answer assumptions or concerns, especially in recent decades in the Software Engineering area. The ESE teaching-learning process in undergraduate and graduate programs still lacks evidence of its effective contribution. For instance, no study explores different ways of giving classes, such as traditional in-person or distance education (online). Our focus in this research is answering the following research question: "Is there any difference in applying three different sets of materials in teaching and learning introduction concepts of experimentation in Software Engineering for distance-learning undergraduate students?". We used three different combinations of materials during learning (i.e., only slides, only recorded class, and both). We conducted activities with such materials in a higher education institution. Despite being incipient, the results demonstrate that in each group/profile of enrolled students, the teaching-learning process mediated by hybrid content (i.e., both slides and recorded classes) had no significant difference concerning students who studied only with recorded classes or slides. The results were obtained using descriptive statistics and inferential (hypothesis) tests. Our contributions here offer evidence-based statistical analysis for ESE education through distance learning, thereby enriching current academic research regarding distance education and facilitating knowledge dissemination in the field.**

***Index Terms*—experimental software engineering, evaluation, teaching-learning process, distance education, students**

I. Introduction

Experimentation[1] is commonly used to assess Software Engineering (SE) practices [1]. In both the software industry and scientific research in SE, experimentation is relevant for understanding the software life-cycle and highlighting different perspectives of the software development process and decision-making regarding the SE education process [2].

Instructors, students, and distance education managers categorically state that there are no differences in the students' learning process based on the content of the course and the way it is presented in the distance education (EaD) modality [3], [4]. This hypothesis is defended from the viewpoint of those closely linked to this type of teaching. The statement is entirely based on the theory and opinions of the parties involved as they work directly with the distance learning modality. One method to gain a deeper understanding of this hypothesis is to conduct experiments to collect data on the approach to introducing new concepts to distance learning students.

In computer science, experimental software engineering (ESE) plays a fundamental role in scientific communities and the professional SE field. In this sense, ESE aims to provide evidence regarding a certain theory or technology related to the software development life cycle [2].

According to Wohlin et al. [2], an experiment is a research method used to investigate a set of testable hypotheses suitable for observing phenomena, formulating theories, confirming conventional knowledge, exploring relationships, validating measures, and evaluating the prediction of models.

In this paper, we aim to demonstrate how the knowledge gained from an experiment can be important in the broader development context. Controlled experiments are important from the perspective of research or technology transfer. Regardless, we need to improve how we teach and conduct studies in software engineering [2].

Therefore, we propose an experiment to assess the degree of students' experience by measuring the level of knowledge the student obtained after the topic "introduction to experimental

[1]In this paper, we focus exclusively on controlled experiments and quasi-experiments.

software engineering", among other perspectives from the data obtained.

This paper is organized as follows: Section II presents background; Section III discusses the planning and conducting of the experiment, presents its results and discussions; Section IV discusses the experiment results; and VI presents the contributions, limitations, and suggestions for future work on this research.

II. Background and Related Work

This section presents relevant concepts regarding ESE, its practice and teaching in higher education, the evaluation of the teaching-learning process, and related work.

A. Software Engineering Experimentation

In the Software Engineering area, experimentation pertains to conducting controlled experiments to evaluate research, tools, techniques, and any other artifacts from the SE process [2]. In the context of experimentation, empirical studies serve to demonstrate the practical application of an idea and can also be referred to as field research. These studies are fundamental for consolidating a proposed solution [5].

An experiment is conducted mainly based on different phases as depicted in Figure 1.

The **Experiment definition** phase is concerned with identifying the context where the experiment will take place; the **Experiment planning** is focused on defining the main elements as hypotheses, variables, design, and validity concerns; **Experiment operation** regards data collection; **Analysis and interpretation** is focused on applying proper techniques to data collected, thus providing an interpretation of the evidence; and **Presentation and package** prepares all the experiment artifacts to be shared to a prospective experiment.

Experimentation in the context of software production combines facts, assumptions, speculations, and beliefs [2]. Hypotheses are formulated along with an evaluation strategy for them. The outcome provides evidence of whether the assumption, speculation, or belief holds validity and under what circumstances [5].

Carrying out an experiment considers the relationship between cause and effect through a theoretical model between two or more related phenomena to determine whether the proposed model can be considered correct. Thus, the model creates a hypothesis regarding the particular changes in the phenomena (the cause) that will lead to changes in the other (the effect). Therefore, the role of the experiment is to test the hypothesis to decide whether it is true or false [6].

Historically, the validation of a hypothesis was determined by whether or not people utilized the idea. Although this is not a confirmed phenomenon, it can serve as an example of what might have occurred due to the lack of validation through experiments of tools and techniques in the field. For instance, consider the difficulties software developers and users faced in the 1960s, a period known as the software crisis. The demands for new applications were high, the problems were increasingly complex to solve, and the absence of well-established development techniques complicated the process [5].

The absence of experimental rigor was highlighted several years ago by Tichy [7] and Zelkowitz and Wallace [8]. Upon studying works published in various software journals, it was identified that 30% of the works were not experimentally validated, and only 10% presented some form of experimental validation following a formal approach [8]. Another study pointed out that only 8% of the works included considerable quantitative evaluation, and only two of the evaluations were formally conducted [7].

In 1998, in the U.S., a workshop by the National Science Foundation [9] presented a report emphasizing the importance of software to the nation. Among the different strategies presented is the relevance of empirical investigation in Software Engineering. It is believed that an increase in the evaluation of software techniques and tools would enhance quality, boost productivity, and reduce errors.

The 610.12 standard [10] for Software Engineering stipulates that scientific knowledge about software development, operation, and maintenance should be applied in disciplines, with experimentation falling under scientific knowledge. The intention was to replace perceptions and prejudices with analysis and information. Solid ideas may not be absolute.

The use of experimentation in software engineering has exponentially increased due to the efforts of the SE community [11]. However, there is still little discussion and knowledge about the teaching and learning of experimentation in software engineering [12]. Industry professionals may find the practice challenging and inappropriate, possibly due to the difficulty of finding industry professionals to participate, making students a valid alternative for conducting tests and gathering results [13].

B. Teaching-Learning ESE

The teaching of ESE is presented as a new subject within the academic environment, gradually gaining recognition. There is a discussion [2] about different ways to teach ESE within a given curriculum. These methods include

i) integration with SE courses, where ESE is part of other disciplines and is evaluated based on learning assessments and practical work, such as in a requirements engineering course where students can compare different requirement specification techniques;
ii) as a standalone discipline, where the main advantage lies in the specific focus on experiments as practical work; and
iii) as part of a research methodology course, which is independent of whether the students are enrolled in undergraduate, master's, or doctoral programs.

Meireles and Bonifácio [14] state that software technologies' increasingly globalized production process has changed how solutions are developed to keep up with current technological advances. This scenario added new challenges to the teaching process of the ES subject, where

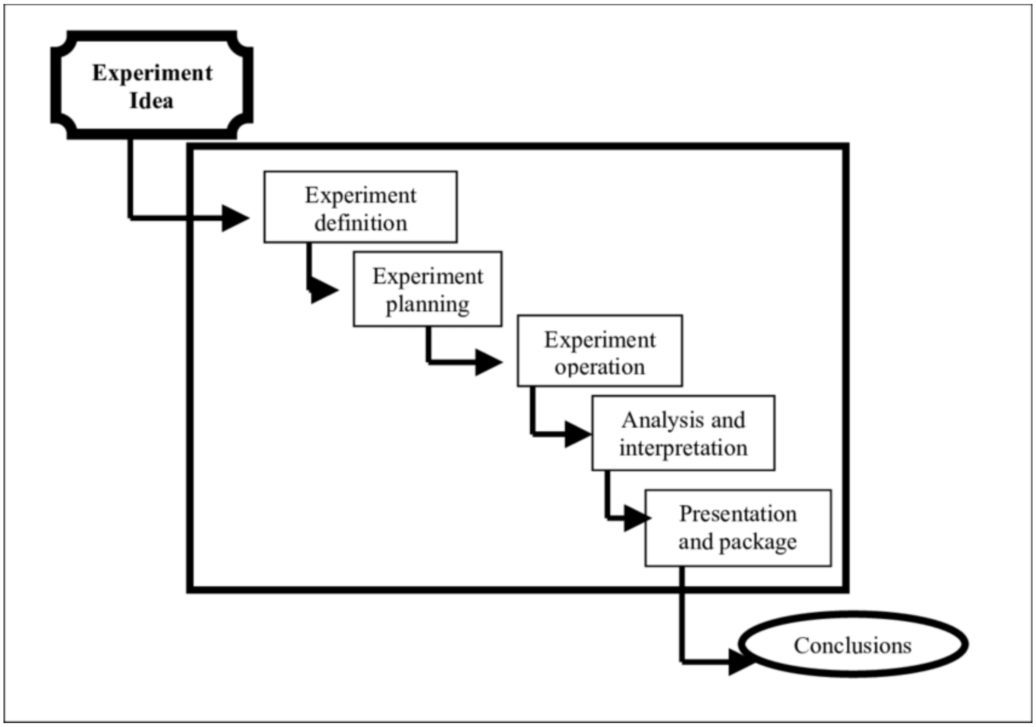

Fig. 1. The Experimentation Process by Wohlin et al. [2]

this new way of producing software needs to become a fundamental element.

Assumptions about the software construction process can be effectively tested based on the concepts of ESE [5], thereby providing reliable evidence about software production instead of waiting for its consolidation over years, even decades [13], [2]. This accelerates the transfer of technology [15].

C. Characterization of the ESE Course

The teaching process aimed to present those involved with the basic concepts of experimentation in Software Engineering, considering that few or none of those involved in the study had prior knowledge of the introductory ESE subjects to be covered.

The total time of the course was 60 minutes. The contents covered are as follows:

i) What is an experiment;
ii) Motivation on why to carry out experimentation in SE;
iii) Introduction to empirical study: Quantitative and qualitative investigation;
iv) Types of Empirical Studies;
v) Experimentation and Learning;
vi) The Experimentation Process; and
vii) Formulation of Hypotheses.

The course was performed as an extra-class activity. This means that only previously invited SE students who were interested attended it. The students received an experiment certificate of attendance for their CVs.

As presented in the section III-C5, to conduct the study and teaching process, participants would be separated into 3 groups. The content to be covered in each group would be the same, only the way it was made available would change (video, base text, video and base text). This way, we could determine that regardless of the group to which the participant would be allocated, the contents to be covered would be the same.

The teaching process was conducted by providing participants with recorded videos and basic study material (articles and texts). The Learning Management System (LMS) of the institution where the study would be carried out was used. Iteration with participants (invitation to the study, sending information about the conduct of the study, sending study materials to participants, link to the response form on knowledge of the subject covered) was carried out by sending messages through the LMS. During the teaching period, participants could send messages with their doubts and observations to those involved, thus creating a fluid and effective means of communication.

D. Related Work

With the rapid and growing development of Distance Education (DE), several researchers have dedicated themselves to studying and evaluating how students learn best, whether through reading, listening, or a hybrid approach. Even before the swift development of distance education, this topic was already being studied, and the aim was to apply content to students in the most effective way possible for optimal absorption. Two examples are the Cognitive Load Theory by Sweller [16] and the Multimedia Learning Theory by Mayer [17].

Sweller's theory [16] advocates the idea that the load of information provided to humans should be compatible with their processing capacity. It is believed that by developing educational materials based on this theory, students will learn much more, as the material will only contain what their minds can process. Mayer's theory [17] defends the idea that the use of multimedia can further deepen students' knowledge, going beyond texts and images. This author grounded his theory in experiments and constructed principles to assist in developing multimedia educational material, asserting that combining multimedia with texts and images enhances learning.

Considering research focused on experimental software engineering education, Kuhrmann [1] developed a course project to teach empirical software engineering, where students carried out mini-projects of theoretical, practical, or cross-cutting nature, addressing a variety of different empirical methods.

Other research reports experienced regarding software engineering teaching, but none focus on experimentation in software engineering [18], [19], [20], [21]. As observed, although these studies are related to learning, there are also no experiments that solely evaluate the three modalities being studied here: hybrid, text-only, and multimedia.

III. The Experiment

In this section, we present the controlled quasi-experiment conducted to evaluate the teaching-learning process of experimentation in software engineering for distance learning students.

A. Methodology

We followed the methodology proposed by Wohlin et al. [2] and Juristo and Moreno [5] with the evidence-based software engineering experiment guidelines by Jedlitschka et al. [22] and Furtado et al. [23]. The experimentation process is illustrated in Figure 1.

B. Goal and Research Question

The objective of this study is to **analyze** the teaching of introduction of experimentation concepts in software engineering with the aim of characterizing it with regard to the effectiveness of students' study process **from the perspective of** different ESE content materials **in the context of** distance undergraduate students of the Software Engineering at a Brazilian university.

Therefore, our research question is: the level does the absorption of introductory knowledge about ESE in higher-level distance learning courses vary based on how the content is presented to students? Do students learn the same way regardless of the format of the study material offered to them?

C. Planning

This section presents the plan for the experiment.

1) Hypotheses Formulation: Based on the research question and the block separation procedures (Section III-C5), we formulated the following hypotheses:

- **Null Hypothesis (H0):** there is no difference in the teaching-learning process of introductory ESE using recorded lectures and/or slides. $H0 : \mu(F1) = \mu(F2) = \mu(F3)$
- **Alternative Hypothesis (H1):** there is a difference in the teaching-learning process of introductory ESE using recorded lectures and/or slides. $H1 : \mu(F1) \neq \mu(F2) \neq \mu(F3)$

2) Definition of Variables: As an independent variable for this experiment, we have **the teaching material**, which is a factor with three treatments (recorded lectures, slides, and both).

As a dependent variable, we defined the **effectiveness** of the students learning, which is the number of correct answers to the questionnaire (Section III-C3) after classes.

3) Instrumentation: For the formulation of the questions, we used Bloom's Taxonomy [24], known for the classification and organization of questions. Thus, we understand that the results are meaningful and contribute to ESE's teaching and learning process. The terminology is common among instructors and guided by various coordinators, allowing each taxonomic category to present what the individual has learned. This theory was developed by Bloom (1956) and conceived by the North American Association of Psychology in 1948 [25].

Conklin [26] states that Bloom's taxonomy became so important and considerable in the 1950s one of the major problems in educational literature was the lack of consensus between words and their instructional objectives.

Easy-level questions were used to assess the student's knowledge and understanding. They analyzed their ability to remember previously addressed information and content, understand and give meaning to it, and address it in different contexts.

Medium-level questions were used to assess the student's application and analysis, thus their ability to use the content in different situations and subdivide it to understand the final stage can be analyzed.

Difficult-level questions are aimed at synthesis; thus, the ability to join and create something new can be assessed.

The instrument was evaluated by three senior researchers on software engineering education and experimentation. One with 25 years of education experience and one with 15 years of experimentation education experience. Another researcher has 13 years of software engineering education and 20 years of experimentation.

Instruments availability information is in Section V.

4) Selection of Participants: The experiment was conducted with students enrolled in the 2nd and 3rd years of a distance education undergraduate Software Engineering course. These student's program periods were chosen because these students are more accustomed to the distance learning modality. The decision to exclude the 1st year was made to minimize potential issues arising from students who are inexperienced with the teaching modality in question.

As experimental software engineering is a completely new topic in the undergraduate course, no prior knowledge of the subject was required from the participants. Therefore, all students (60) who completed the form were accepted as participants.

5) Block Separation: We separated participants into three blocks as follows:

- **Group 1:** learn ESE only using recorded lectures (**F1**);
- **Group 2:** learn ESE only using slides (**F2**); and
- **Group 3:** learn ESE using recorded lectures and slides (**F3**).

6) Threats to Validity: During this research, some threats that may have directly interfered with data collection for final analysis were detected. The first is that the execution of stage 1 occurred in the middle of the student's curricular discipline. For this reason, we had many participants, but when applying the stage 3 form, we had a very low number of responses based on the registrants.

The low response rate may be linked to the end of the discipline and the delivery of academic activities, as the study period for the experiment took place in parallel with the studies for the delivery of activities. The final deadline for academic responses also coincided with the deadline for work to close the disciplines they were studying when the experiment occurred.

Even with low response numbers, data analysis can be performed, and some relevant assumptions can be made based on the hypotheses raised for the research development.

Another threat might be the fact that the ESE classes were given as an extra class activity. Thus, the sample size might be impacted as students were invited to attend the experiment outside of their typical scheduled hours of study.

D. Operation

The experiment was divided into four steps.

Step 1: students were invited to participate in the experiment through a message sent via the university Learning Management System (LMS) Moodle. Along with the message, a link was sent that took interested students to a characterization form with personal questions and others about the topic of ESE. After collecting the data, the candidates were separated into three groups as indicated in Section III-C5 and demonstrated in Table I.

The separation of participants was carried out randomly, without considering their knowledge or year of graduation, only regarding the unevenness in the number of participants in each group.

We invited around 250 students from 2nd to 4th years to participate in the experiment. The students had a period of 15 days to fill out a form indicating their interest in attending the experiment.

Step 2: with the participants divided into groups, study materials on the topic of ESE were sent to each group by email. In addition, we sent individual messages to the participants through the LMS.

TABLE I
PARTICIPANTS GROUPING AND RESPECTIVE COURSE YEAR

Groups	2nd year	3rd year	Total
Group 1 (F1)	12	8	20
Group 2 (F2)	12	8	20
Group 3 (F3)	12	8	20
Total	**36**	**24**	**60**

Step 3: after we sent the material and the time for the study, we made available a form to collect their knowledge on ESE with several questions, separated into four sections:

- **Section 1:** four easy-level questions;
- **Section 2:** four medium-level questions;
- **Section 3:** four difficult-level questions; and
- **Section 4:** four open questions on self-evaluation and how the ESE content was approached.

The link for filling out the form was sent via email and individual message through the LMS as follows.

Step 4: we analyzed all data collected from the students' forms based on statistical techniques and tools such as the R environment.

E. Results

This section presents the analyses based on the participants' responses obtained in Step 3,consideringt normality and hypothesis tests.

We plot the three samples in boxplots and presented their descriptive statistics in Figure 2.

1) Demographics: We had the participation of 24 students from the SE undergraduate program. Table II presents the summarized participants' profiles.

The experiment carried out included only Brazilian participants, covering the South, Southeast, and Northeast regions, as shown in table III.

Amongst the 24 responding students, the predominant age range is 21 to 34 years old, followed by the age range 35 to 49 years old, as can be seen in Table IV. Therefore, we have 12.5% in the age group from 18 to 20 years old, 45.83% in those from 21 to 34 years, 37.5% in those from 35 to 49 years, and 4.17% in those from 50 to 64 years.

Regarding the gender of the 24 participants, we have a male predominance, with 83.3% (20) respondents and 16.7% female (4) respondents.

Regarding the academic level of the students, 50% were taking their first higher-degree program, and the other 50% had already completed at least one higher-degree program. Of the 12 students who have already completed a higher degree, two studied in the area of Information Technology, among the other listed: Environmental Management Administration, Social Communication, Physiotherapy, Medicine, Industrial Automation, and Pedagogy.

Amongst the 12 students who had already completed some higher degree, four belong to Group 1, two to Group 2, and six to Group 3, according to Table V.

Data Summary								
Groups	N	Min	Q_1	Median	Q_3	Max	Mean	SD
Group 1	6	5	8	8.5	9	10	8.1667	1.7224
Group 2	6	6	7.25	8	8.75	11	8.1667	1.7224
Group 3	12	2	5.75	7	9	10	6.8333	2.3677

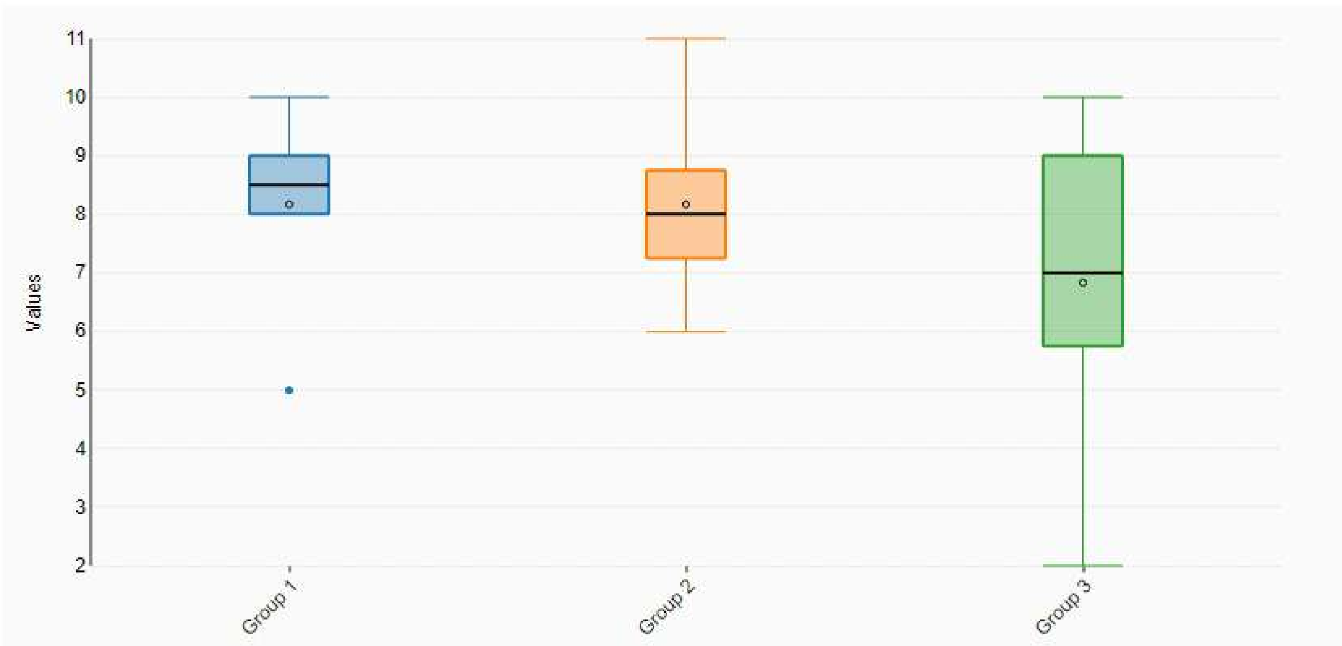

Fig. 2. Boxplots and Descriptive Statistics for Samples F1 (*Group 1*), F2 (*Group 2*), and F3 (*Group 3*).

Among the 24 respondents, the only student who reported having already carried out an experiment and having knowledge of Experimental Software Engineering is among the 12 students who have already completed some degree. He/she reported having a degree in Administration and Social Communication. Of the 12 graduated students, in addition to the aforementioned student, we have one more who has already carried out an experiment but was unfamiliar with Experimental Software Engineering. This one has a degree in Pedagogy. Another student reported knowing the term ESE but never having carried out an experiment. He/she already has a degree in the area of Information Technology. Among the 12 students taking their first degree, 50% were in the 2nd year students, and 50% were in the 3rd year students. None of them knew the term ESE, nor had they carried out an experiment before.

The two students who have already carried out some type of experiment reported having more than 10 years of experience in a software development company. Among the participants, we have 29.2% of students with experience in software development companies, one student with 1 to 2 years of experience, three with 2 to 5 years of experience, and three with more than 10 years of experience.

2) Descriptive Statistics: Figure 2 presents the descriptive statistics for the three samples. The individual values of correct answers are presented in Table VI. Note that the maximum correct answer (effectiveness) value is 12.

The F1 sample has a median value of 8.5, a mean value of 8.17, and a standard deviation value of 1.72. It means that at least 50% of the sample had the number of correct answers greater than 8.5, which represents 71% of effectiveness. In addition, we can observe a low standard deviation value, which characterizes a similar level of ESE comprehension based on recorded lectures.

The F2 sample has a median value of 8.0, a mean value of 8.17, and a standard deviation value of 1.72. It means that at least 50% of the sample had the number of correct answers greater than 8.0, which represents 66.7% of effectiveness. In addition, we can observe a low standard deviation value, which characterizes a similar level of ESE comprehension based on slides as teaching materials.

The F3 sample has a median value of 7.0, a mean value of 6.83, and a standard deviation value of 2.37. It means that at least 50% of the sample had the number of correct answers greater than 7.0, which represents 50% of effectiveness. In addition, we can observe a not-so-low value of standard

TABLE II
SUMMARIZED PARTICIPANTS' PROFILES

Group	St. ID	Pr. Year	Gender	Has Prior Higher Education?	Replied Open-Ended Question?
F1	S1	3	M	No	Yes
F1	S2	3	M	Yes	Yes
F1	S3	2	M	Yes	Yes
F1	S4	2	M	Yes	Yes
F1	S5	2	M	No	Yes
F1	S6	2	M	Yes	Yes
F2	S7	3	M	No	Yes
F2	S8	3	F	No	Yes
F2	S9	3	M	No	Yes
F2	S10	3	M	No	Yes
F2	S11	2	M	Yes	Yes
F2	S12	2	M	Yes	Yes
F3	S13	3	F	Yes	Yes
F3	S14	3	M	Yes	Yes
F3	S15	3	M	No	Yes
F3	S16	3	M	Yes	Yes
F3	S17	2	M	No	Yes
F3	S18	2	M	No	Yes
F3	S19	2	M	No	Yes
F3	S20	2	M	No	Yes
F3	S21	2	M	Yes	Yes
F3	S22	2	F	Yes	Yes
F3	S23	2	M	No	Yes
F3	S24	3	F	Yes	Yes

TABLE III
BRAZILIAN REGIONS PER SAMPLE

Groups	South	Southeast	Northeast
Group 1 (F1)	4	2	0
Group 2 (F2)	3	3	0
Group 3 (F3)	7	3	2
Total	**14**	**08**	**02**

TABLE IV
AGE RANGE IN YEARS PER SAMPLE

Groups	18-20	21-34	35-49	50-64
Group 1 (F1)	0	3	2	1
Group 2 (F2)	2	3	1	0
Group 3 (F3)	1	5	6	0
Total	**03**	**11**	**09**	**01**

TABLE V
PARTICIPANTS' PREVIOUS HIGHER-EDUCATION PER SAMPLE

Groups	No Higher-Education	With Higher-Education
Group 1 (F1)	2	4
Group 2 (F2)	4	2
Group 3 (F3)	6	6
Total	**12**	**12**

deviation, which leads to an assumption for a similar level of ESE comprehension based on recorded lectures and slides as teaching materials.

By comparing F1 and F2, we can say such samples had similar results, thus affecting recorded lectures and slides as potential teaching materials. On the other hand, by combining such material, the F3 sample had median and mean values lower than F1 and F2. In addition, its standard deviation and median and mean values provide insights for future assumptions and further investigations.

TABLE VI
CORRECT ANSWERS PER STUDENT ACCORDING TO THEIR GROUPS (F1, F2, AND F3)

Student ID	F1	F2	F3
S1	8	–	–
S2	10	–	–
S3	8	–	–
S4	9	–	–
S5	5	–	–
S6	9	–	–
S7	–	9	
S8	–	6	–
S9	–	8	–
S10	–	11	–
S11	–	8	–
S12	–	7	–
S13	–	–	7
S14	–	–	4
S15	–	–	10
S16	–	–	2
S17	–	–	8
S18	–	–	6
S19	–	–	9
S20	–	–	7
S21	–	–	5
S22	–	–	9
S23	–	–	9
S24	–	–	6

3) Normality Tests: Based on the results of Table VI, we performed normality tests using Shapiro-Wilk with 95% (0.05) of significance, as follows:

- **F1 sample:** p-value = 0.2102 $>$ 0.05 leading to a normal distribution;
- **F2 sample:** p-value = 0.8302 $>$ 0.05 leading to a normal distribution; and
- **F1 sample:** p-value = 0.5938 $>$ 0.05 leading to a normal distribution.

Therefore, we chose a parametric hypothesis test to support our hypothesis.

4) Hypotheses Tests: For hypothesis testing, we chose a non-paired T-test. To test our hypotheses, we grouped each of the three samples in 2x2 tests with a significant level of 95% (0.05), as follows:

- **F1 vs. F2:** obtained p-value = 1.0 $>$ 0.05;
- **F1 vs. F3:** obtained p-value = 0.1963 $>$ 0.05; and
- **F2 vs. F3:** obtained p-value = 0.1963 $>$ 0.05.

Based on the results of each test, we do not have statistical significance to reject our null hypothesis (H0), which means **there is no difference in the teaching-learning process of ESE using recorded lectures and/or slides** (H0: μ(F1) = μ(F2) = μ(F3)).

5) Qualitative Results based on the Open Questions: We performed the Grounded Theory procedure for Coding to gather results from the open questions, as suggested and revisited by Hoda [27] and Stol et al. [28].

At the end of the evaluation instrument (Section III-C3), we provided open-ended questions in which the participants could give their opinion on their expected performance when taking the experiment, how the materials were approached/made available to participants, and express themselves freely about the conduct of the experiment.

When analyzing the question about their **own expected performance** in the experiment, we observed that 37.5% (9) of the participants a **Regular** one, followed by 33.3% a **Good** performance, 16.7% expected a **Poor** performance, 8.3% a **Great** one, and 4.16% expected an **Excellent** performance.

Regarding the participants' understanding of **how the ESE contents were introduced**, the majority (16) indicated that the approach was **Great** and **Excellent**. Seven responded **Relevant**, and one indicated **Indifferent** for the contents introduction.

Concerning the **way the ESE contents were made available** to participants, 12 responded as **Excellent** and **Great**, six indicated that the process was **Relevant** and six that the process addressed was **Indifferent**.

When analyzing the open-ended answers, we noted that the majority of participants stated that the experiment process and conduction were positive, for instance, highlighting as follows (our free translation from Portuguese):

- **Student S5**: "*...very relevant content, especially for those who have no knowledge of experimentation, but perhaps more bases or examples of the application of Experimental Software Engineering are needed...*"
- **Student S8**: "*...I think the scientific basis is important, as we still don't know the consequences of software development since we first develop and empirically analyze its consequences...*"
- **Student S21**: "*...high expectations to develop Software Engineering skills with this project...*"

IV. Discussions of Results

In this section, we discuss our experimental results.

A. Participants Knowledge

Certain students who attended the experiment already had experience in the software area. Out of 12, five had more than 10 years of experience in the software industry, two participants had 5-8 years of experience, and three had 2-5 years of experience. The remaining had 1-2 years.

We understand that the previous knowledge did not affect the experimental results. Although ESE is encouraged to be applied in the industry, the ESE content is rarely reached in undergraduate programs. Therefore, little influence was observed in the results. Then, a further investigation must be conducted to establish new hypotheses to be tested for this matter.

B. Effectiveness of Participants

The statistical analysis of the results demonstrated no difference regarding the type of content material (F1, F2, or F3). We should otherwise note the low number of participants per group.

Even as the tests have shown no difference in teaching materials, we observed that higher education students' level of knowledge absorption in the distance learning modality provided proportional adherence among the groups. The groups have similar median and mean values, except for F3, with a higher internal difference and a greater standard deviation than F1 and F2.

Therefore, we need to analyze such results further, refine our hypotheses, and run another experiment with a considerably large sample for each material. Thus, we will have more evidence for performing a meta-analysis to combine results.

C. Participants Self-Assessment

Based on the results of the open questions about the participants' attendance at the experiment, we understand that the experimentation process was positive in general.

Certain participants stated they found it important to attend the ESE classes and responding our instrument to test their gathered knowledge and participation perspectives, as highlighted in Section III-E5.

Even with a low sample size and the ESE contents given as extra classes, we realized that, in the context of this experiment, there was no difference in which materials were made available to students. Therefore, we highlight the need to obtain a larger sample and refine our hypotheses for prospective actions.

V. Data Availability

As our research group advocates for Open Science practices [29], all data from this study is available in a public, permanent, and DOI-based repository at https://doi.org/10.5281/zenodo.11223397.

VI. Conclusion

This paper presented a controlled quasi-experiment with 24 distance-learning undergraduate students on teaching ESE with different materials and procedures. We assessed the participants' knowledge by distributing different ESE materials, such as video and text, and asking questions regarding them. By conducting this research, we noticed that the way in which the teaching content of disciplines in the distance modality is presented is of great importance.

Even though the experiment was subject to some threats, we performed some analyses to obtain important data for future research on the subject of ESE education, such as the strategy of offering content for distance education students.

Based on data analysis, we observed that higher education students' level of knowledge absorption in the distance learning modality did not present significant variations depending on how the content was presented. In addition, based on the hypothesis tests, we could not reject H0, which means there is no difference among the content materials used for teaching ESE.

We shared our thoughts on the growing use of experiments in the software engineering discipline. This practice is becoming increasingly common and valuable because ESE learning contributes significantly to research and the generation of new hypotheses for further investigation. However, this relationship between ESE learning and research should be considered in the context of prospective ESE teaching-learning experiments and practices.

Acknowledgments

The authors thank the Brazilian university, the coordinator of the distance-learning undergraduate program in Software Engineering, and all the students who participated in this experiment. We also thank CAPES for supporting this study (code 001). Edson OliveiraJr thanks CNPq/Brazil grant #311503/2022-5.

References

[1] M. Kuhrmann, "Teaching empirical software engineering using expert teams," in *CEUR Workshop Proceedings*, 02 2017, pp. 1–12, university of Southern Denmark. [Online]. Available: https://findresearcher.sdu.dk/ws/files/124532226/2016_12_16_final.pdf

[2] C. Wohlin, P. Runeson, M. Höst, C. O. Magnus, R. Björn, and Anders, *Experimentation in Software Engineering*. Springer Berlin, Heidelberg, 2012.

[3] C. Chisadza, M. Clance, T. Mthembu, N. Nicholls, and E. Yitbarek, "Online and face-to-face learning: Evidence from students' performance during the covid-19 pandemic," *African Development Review*, vol. 33, no. S1, pp. S114–S125, 2021.

[4] E. Duman, "The challenges of distance education and evidence-based solution suggestions," *International Journal of Academic Studies in Technology and Education*, vol. 1, no. 1, May 2023.

[5] N. Juristo and A. M. Moreno, *Basics of Software Engineering Experimentation*. Springer New York, NY, 2001.

[6] B. A. Kitchenham, D. Budgen, and P. Brereton, *Evidence-based software engineering and systematic reviews*. CRC press, 2015, vol. 4.

[7] W. F. Tichy, "On experimental computer science," 1993. [Online]. Available: https://link.springer.com/content/pdf/10.1007/3-540-57092-6_97.pdf

[8] M. Zelkowitz and D. Wallace, "Experimental models for validating technology," *Computer*, vol. 31, no. 5, pp. 23–31, 1998.

[9] NSF, "Final report nsf workshop on a software research program for the 21st century greenbelt maryland october 1998," 1999.

[10] IEEE, "Ieee standard glossary of software engineering terminology," *IEEE Std 610.12-1990*, pp. 1–84, 1990.

[11] D. Sjoeberg, J. Hannay, O. Hansen, V. Kampenes, A. Karahasanovic, N.-K. Liborg, and A. Rekdal, "A survey of controlled experiments in software engineering," *IEEE Transactions on Software Engineering*, vol. 31, no. 9, pp. 733–753, 2005.

[12] J. M. Gonzalez-Barahona and G. Robles, "Revisiting the reproducibility of empirical software engineering studies based on data retrieved from development repositories," *Information and Software Technology*, vol. 164, p. 107318, 2023.

[13] W. Tichy, "Should computer scientists experiment more?" *Computer*, vol. 31, no. 5, pp. 32–40, 1998.

[14] M. C. Meireles and B. Bonifácio, "Use of agile methods and problem-based learning in software engineering teaching: An experience report," in *Brazilian Symposium on Computers in Education*, vol. 26, no. 1, 2015, p. 180.

[15] D. I. K. Sjøberg, B. Anda, E. Arisholm, T. Dybå, M. Jørgensen, A. Karahasanović, and M. Vokáč, *Challenges and Recommendations When Increasing the Realism of Controlled Software Engineering Experiments*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 24–38.

[16] J. Sweller, P. Ayres, and S. Kalyuga, "Cognitive load theory." Springer, 2003.

[17] R. E. Mayer, "Multimedia learning," New York: Cambridge University Press, 2009, 2009.

[18] M. Kuhrmann and J. Münch, "Enhancing education engineering software through experimentation: An experience report," in *IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 2018, pp. 1–9.

[19] N. E. Cagiltay, "Teaching software engineering by means of computer-game development: Challenges and opportunities," *British Journal of Educational Technology*, vol. 38, no. 3, pp. 405–415, 2007.

[20] M. Gnatz, L. Kof, F. Prilmeier, and T. Seifert, "A practical approach of teaching software engineering," in *International Conference on Software Engineering Education and Training*, 2003, pp. 120–128.

[21] D. Dahiya, "Teaching software engineering: a practical approach." ACM SIGSOFT Software Engineering Notes, 2010.

[22] A. Jedlitschka, M. Ciolkowski, and D. Pfahl, *Reporting Experiments in Software Engineering*. London: Springer London, 2008, pp. 201–228.

[23] V. Furtado, E. OliveiraJr, and M. Kalinowski, "Guidelines for promoting software product line experiments," in *Brazilian Symposium on Software Components, Architectures, and Reuse*. New York, NY, USA: ACM, 2021, p. 31–40.

[24] L. A. Sosniak, *Reflections on the development and use of the taxonomy*. National Society for the Study of Education (USA), 1994.

[25] A. P. C. M. Ferraz and R. V. Belho, "Bloom's taxonomy: theoretical review and presentation of the instrument's adjustments for defining instructional objectives." São Carlos, SP, 2010.

[26] J. Conklin, "Taxonomy for learning, teaching and assessing: a revision of blooms's taxonomy of educational objectives." Educational Horizons, v. 83, n. 3, 2005, pp. 153–159.

[27] R. Hoda, "Decoding grounded theory for software engineering," in *International Conference on Software Engineering: Companion Proceedings*, ser. ICSE '21. IEEE Press, 2021, p. 326–327.

[28] K.-J. Stol, P. Ralph, and B. Fitzgerald, "Grounded theory in software engineering research: a critical review and guidelines," in *International Conference on Software Engineering*, ser. ICSE '16. Association for Computing Machinery, 2016, p. 120–131.

[29] S. United Nations Educational and C. O. (UNESCO), "UNESCO Recommendation on Open Science," Nov. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.5834767