Evaluating the effectiveness of a cloud-based laboratory for teaching Linux operating systems to Computer Science students

Vasyl P. Oleksiuk1,2, Oleg M. Spirin2,3, Olena S. Holovnia4 and Olena G. Glazunova5

1Ternopil Volodymyr Hnatiuk National Pedagogical University, 2 M. Kryvonosa Str., Ternopil, Ukraine
2Institute for Digitalisation of Education of the NAES of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine
3University of Educational Management, 52A Sichovykh Striltsiv Str., Kyiv, 04053, Ukraine
4Zhytomyr Polytechnic State University, 103 Chudnivska Str., Zhytomyr, 10005, Ukraine
5National University of Life and Environmental Sciences of Ukraine, 15 Heroyiv Oborony Str., Kyiv, 03041, Ukraine

Abstract
The paper is a study of using cloud technologies to teach the Linux operating system to future computer science teachers. The authors integrated the NDG Linux Essentials MOOC course, Apache Cloudstack and Proxmox VE private cloud platforms into the CL-OS Cloud Lab. A model of such a lab was developed. The paper describes the author’s teaching materials such as tests, essays and assignments. They complement the resources of the cloud lab. The specifics of students’ activities in the conditions of the real educational process according to the methodology of blended learning were analyzed. To confirm the effectiveness of the Cloud Laboratory methodology, a pedagogical experiment was conducted in the study. Its results were tested using statistical methods such as ANOVA for repeated measures and Spearman’s rank correlation coefficient. These methods made it possible to investigate some of the factors that influence the educational performance of future computer science teachers in learning Linux.

Keywords
Cloud Lab, Cisco Network Academy, future Computer Science teacher, Operating Systems course, experiment

1. Introduction

The Operating Systems course has always been a fundamental discipline in the education of computer science teachers. Currently, competence in working with different operating systems is included as an essential component in many international and national training standards for IT professionals [1]. Although Operating Systems courses have been fundamental for many years, difficulties persist in organizing their study. One of the most significant challenges is providing access to various types of operating systems and arranging remote work with them for the students [2, 3]. This was affirmed by the imposition of numerous lockdowns throughout the COVID-19 pandemic and Russia’s ongoing invasion of Ukraine [4, 5, 6].

Academics suggest incorporating virtualization technologies in the delivery of computer science education [7, 8, 9, 10, 11]. It allows for the creation of a realistic environment while still retaining the ease of debugging that comes from running the operating system and user programs as normal processes on a host machine [12]. When teaching operating systems, lecturers face various issues [13]:

- OS composition requires explaining many abstract concepts which can be difficult for students to understand and manage.
• Though students can understand the principles of kernel primitives explained in the class, they have no idea how to take advantage of these primitives and improve the stability of their applications. They do not know how to locate the required primitives and use them correctly.

• Even though students can make good use of the kernel primitives taught in the class, they lack the enthusiasm to generate innovative applications in their final projects.

• Lacking time for further discussion, students may be forced to limit the scale of their final projects, and may also miss opportunities to interact and cooperate with other students on larger projects.

Cloud labs offer the sharing of study objects published on one computer by several students. When it comes to computer science, such labs are not reliable and scalable enough. Therefore, their educational content is usually taken from MOOCs. For example, in the study [14], the remote electronics laboratory in the MOOC environment was expanded.

MOOCs provide various benefits such as the ability for students to study at their own convenient time, compare various course materials and teaching styles, engage in discussions, and receive self and mutual evaluations [15, 16]. Moreover, MOOCs facilitate the enhancement of listening, reading, and writing skills in English, create an opportunity for reflecting on pedagogical activities, encourage digital creativity, and foster collaboration with other participants [17, 18]. Studies suggest that MOOCs, on account of their scalability and reusability, can offer a cost-effective approach to delivering higher education [19, 20]. MOOCs could also be a means to tackle the widening income gaps affecting higher education access [21]. All of them employ cloud-based tools to provide unrestricted access to knowledge. However, this alone is no longer sufficient. That is because such systems must now guarantee adaptability in learning. According to Ukrainian researchers, adaptable cloud-based learning systems will be the backbone to establish new pedagogy, personalized education strategies, and increased openings for active learning [22]. At present, several platforms have been developed and are working effectively, implementing the MOOC concept.

The first MOOCs developed by non-profit or private organizations featured prominent educators and scientists as instructors. Their emphasis was not on content but on audience engagement. At this time, the phrase cMOOC appears in academic literature. The primary objective of these courses was to facilitate communication among the participants. The course participation model aimed to create a personal learning network. It provides access to content and communication that carries significance for a listening audience. These MOOCs were developed based on the theory of connectivism and raised questions about academic and institutional responsibility [23]. Connectivism theory emphasizes that networks and connections are the principal means of learning. According to the concepts of connectivism, learning is not only an individual activity but a social process that involves communication and interaction with other participants in the educational process. Today, computer networks, the Internet, cloud technologies, and similar technologies form the technical basis for such interaction. Connectivism posits that with the availability of additional educational resources, learning is continually developing and evolving. In this theory, knowledge is viewed as a non-static entity that cannot be fixed or mastered completely. Thus, education is regarded as an ever-changing dynamic process. Importance is given to recognizing patterns and connections between diverse sources of information, rather than mere memorization of facts in Connectivism. This approach can assist students in achieving a comprehensive understanding of the subject matter and evade assimilating superficial knowledge. Since learning takes place in a networked environment, it cannot be limited to one source of information or the personality of the teacher [24]. These networks include social media, MOOC online communities, and the applicant’s connections and relationships [25].

At present, most MOOCs encourage enhanced communication, particularly in a blended format where learners of the same academic discipline undertake the same module [26, 27]. Incorporating these courses into the training process of future IT specialists and computer science teachers necessitates the conceptual development and experimental validation of pertinent methodological models [28, 29]. The primary objective of these programmes is to enable students to join the social and technological educational system where teaching professionals are not central but rather nodes in the broader network [30, 31].
Cisco is developing a strategy to provide widespread access to computer and network technology knowledge through the Cisco Network Academy. The Academy offers both professional and commercial courses. In addition, the academy provides basic courses that are accessible to a broad audience, including educational institutions.

This article aims to examine how certain factors affect the learning outcomes of the Linux operating system when using a cloud-based laboratory.

2. Cloud lab as a learning tool

Today, cloud technology has emerged as a highly efficient framework for providing access to computing resources. Many experts recommend its use as a valuable tool for training future IT professionals [32, 33, 34, 35, 36, 37].

The studies by Nosenko et al. [22], Popel and Shyshkina [36], Papadakis et al. [38], use the term “cloud-based learning and research environment” highlighting its affiliation with a higher educational institution and define it as “the environment in which the virtualized computer-technological (corporate or hybrid-based) infrastructure is purposefully built for the realization of computer-procedural functions (such as content-technological and information-communication functions)”.

To clarify the “cloud laboratory” term we analyzed some studies. Cloud labs use a special user interface (first of all based on the web), which allows working with the user interface of OS in any case. It should not depend on whether the network is running on the virtual machine (VM) or not. Such objects in cloud labs are an integral part of logical network infrastructure with a flexible architecture, which, according to its structure and time, corresponds to some personal needs of a user [39]. Markova et al. [34] describe a ready-made cloud solution VEL (Virtual Education Laboratory, iNetwork, Inc.) as a product that provides “a remote computer service for educational experiments in the information technology industry”.

Cloud labs have been identified as advantageous over traditional ones by Zhang et al. [40]. These advantages include high efficiency in the utilization, and allocation of computing resources, and convenient, fast, and efficient management and maintenance [40]. However, compared to traditional laboratories, the cloud also has its drawbacks and shortcomings, such as difficulties in management and maintenance due to all the resources being virtualized on servers. The non-transparent nature of cloud computing is another disadvantage of cloud labs. As a result, users feel like they lose control over their virtual machines, data, networks, etc. Therefore, they are unsure whether cloud providers can be trusted [41].

An automated control system design environment was developed by Bucher [42]. This environment is Python programming language-based. All control design tasks such as simulation, identification, controller design, and simulation can be performed by students in this environment. Once completed, they can generate RT code for PCs, Raspberry Pis running Linux RT OS automatically [42].

Svatos et al. [43] presented the “Home Laboratory” that was developed to remotely conduct practical lessons on the microelectronics course. The home laboratory consists of two functional groups: a laboratory experimental device and a system of measuring devices, microcontrollers. The laboratory model was successfully tested through distance learning during the COVID-19 pandemic. There is evidence that the laboratory can be deployed successfully for training according to the project methodology. A VPN server was set up by the researchers to enable students to collaborate on projects [43].

Nevertheless, depending solely on digital technologies is inadequate to harness the complete potential of cloud laboratories in education. Integrating high-quality educational content with each cloud lab is essential, in parallel with the incorporation of tools that encourage personalized and adaptable learning experiences [44]. Gillet and Li [45] have delved into cloud laboratories’ concept as collaborative spaces that integrate various applications. Many studies have focused extensively on the incorporation of MOOCs in the learning environment. According to researchers, cloud labs have the potential to facilitate MOOCs’ inclusion, allowing teachers or students to access and monitor the usage of openly accessible
learning resources.

In this study, the integration was performed in a laboratory to learn about the Linux operating system. Generally, our original course “Operating Systems” is a two-semester program. During the first semester, students study Windows, and in the second semester, they study Linux. The research was carried out in the joint laboratory of the Institute for Digitalisation of Education at the National Academy of Educational Sciences of Ukraine and Ternopil Volodymyr Hnatiuk National Pedagogical University (Ukraine).

Numerous studies have found that MOOCs can effectively enhance the efficacy of distance learning, blended learning, and traditional face-to-face educational approaches [23, 46]. Wu [47] discovered that online reviews (as reviewing autonomy) of MOOC (massive open online course) learners are influenced by their progress in the course (as learning abilities), as well as the number and length of posts (as indicators of learning and social engagement) within assignments, quizzes, and discussion. Therefore, to implement the CL-OS cloud lab, we adapted the NDG Linux Essentials course, which was originally developed by the Cisco Networking Academy [48]. This course is structured in a MOOC-style format, which enables it to be taught without requiring prior teacher training or accreditation. However, we completed the entire course ourselves before making it available to students. Following the conclusions drawn from the study by Bralić and Divjak [46] certain actions have been taken. For instance,

- The Content of the course has been carefully examined to ensure a reasonable workload and expectations from students.
- Learning outcomes have been taken into consideration to properly connect online and offline learning and to deploy a cloud lab that ensures achieving those outcomes.
- Fully completion of a MOOC with a certificate was not strictly required (the main reasons for this will be discussed later).

The cloud laboratory CL-OS implemented the following technologies and educational resources.

1. Graphical User Interface. The NDG Linux Essentials course does not include learning the graphical user interface (GUI). Perhaps this is due to the many different implementations of the GUI. Nevertheless, we added a suitable module and requested students to conduct a comparative analysis of various GUIs for Linux. We created several VM templates and deployed them in our cloud laboratory.

2. Mini-lectures from the original course. We did not hold traditional lectures. Instead, the instructors delivered mini-lectures to the students. It made no difference whether this happened face-to-face or online. During these mini-lectures, the basic concepts of the module were explained and links to concepts that students would later learn were systematically repeated and reinforced.

3. Tasks for students’ independent work. We made these tasks different in complexity, grading them as the first, second and third levels. Tasks are also selected randomly from a bank of quizzes. If such tasks were created in sufficient numbers (in our case about 200), it ensured that the student completed his/her own block of tasks. We analyzed the capabilities of LMS MOODLE. Cisco Networking Academy currently runs courses based on this system. A good option is to use the “random task” plugin. However, it is not available at the Cisco Networking Academy. Therefore, we used the “test” module. The tasks were classified according to complexity levels and presented in an essay format. This course’s particular redesign seeks to augment comprehension and encourage students to regard the coursework as more germane and valuable [12].

4. Public cloud services were used. To organise the training, the Google Meet cloud service from the Google Workspace for Education package was used. Recording and sharing of all laboratory work was performed. The assessment of learning outcomes was conducted through personal communication between the teacher and the student in a separate room. Chapter tests were evaluated as well.

Our cloud laboratory offers students and teachers ubiquitous on-demand access to a shared pool of configured computing resources. Based on studies by Selviandro and Hasibuan [49], Togawa and Kanenishi [50] we propose the architecture of such a laboratory (see figure 1).
The CL-OS laboratory model is presented as a 4-layer structure. Infrastructure, application, access, and user layers implement technical and educational requirements for the cloud laboratory. This laboratory corresponds to the cloud concept: it is fully automated, inexpensive and scalable in design; Apache Cloudstack or Proxmox VE cloud platforms architecture provides simultaneous work of students with Linux VM from 50 to 100 students [51]. The maximum number of VMs will be in the case of the command-line mode of VMs running. If students need to work with a graphical interface, the number of VM will be approximately 2 times less.

We can conclude that more qualified IT professionals are needed to manage and maintain cloud labs based on Apache Cloudstack or Proxmox VE platforms. For example, in our cloud lab implementation, some issues (forcibly shutting down, deleting, or transferring machines to another user) are sometimes solved by directly editing the database. The laboratory is built according to the IaaS model. Apache CloudStack or Proxmox VE infrastructure allows teachers, students, and administrators to aggregate, route and filter traffic from different networks such as the local network of the university, virtual local area networks of cloud infrastructure, and virtual private networks through which students access the university’s cloud resources.

To ensure efficient completion of laboratory tasks, the cloud laboratory incorporates virtual machines from either the “sandbox” or the Apache CloudStack cloud platform. Now, let us scrutinise some of these situations.

- Multi-user mode allows students to understand how Linux performs tasks such as putting processes in the background, finding and changing the priority of a process, and forcefully terminating a process.
- Comparison of Linux working in Sandbox and Apache CloudStack. An example is the Investigation of the first parent process name in various Linux distributions and releases (module 13 “Where Data is Stored”). Students find the name of the first parent process using at least two different commands (ps, top etc) and by analyzing the content of the /proc/1/ directory. They complete this task in Sandbox and Linux VM from the CL-OS and compare the obtained results.
- Investigate virtualization tools of CL-OS lab. (module 12 – Understanding Computer Hardware). Students get the list names of the virtualization platform under their Linux OS. They use the...
systemd-detect-virt command. Alternative ways to do this are commands dmidecode or lshw. Students compare the results from different Linux environments and complete a corresponding table.

3. An experiment to test the effectiveness of using a cloud laboratory

A previous study examined the disparity in student academic achievement before and after using the lab CL-OS [52]. Most of the participants in the educational experiment observed an increase in their scores, as recorded.

3.1. Research of the influence of some factors on the students’ academic achievements

An additional study was conducted to verify the earlier assumption. We investigated whether the students had prior experience in using Linux before enrolling in the course. Furthermore, the gender of the respondents was included in the updated survey data. This study aims to evaluate the impact of these factors on the grades achieved in the authors’ course. We opted for the Three-way Mixed ANOVA as the method to assess the impact of these factors [53]. Our study has two between-subjects factors – gender and experience – along with one within-subjects factor. The purpose of this factor is to investigate variations between groups over time. It refers to measurements that are taken repeatedly. The study dataset is accessible through the provided link. The data contains the following variables:

- Assessment of academic achievement (dependent variable) measured during the course at two-time points, t2 and t3.
- Two between-subjects factors: gender (levels: male and female) and previous experience in Linux using (levels: yes, no)
- One within-subjects factor, time, which has two time points: t2 and t3
- Course Feedback Assessments. The names of these variables begin with the letters "S_" and "A_". They will be discussed later.

The dataset underwent statistical processing through R packages, such as tidyverse (for visualization and data manipulation), ggpubr (for producing publication-ready plots), and rstatix (for conducting statistical analyses). The data was first read from the csv-file and then converted into a variable as a factor.

```r
achievement <- read.csv2(file = "OS_dataset.csv", sep = "\,");
achievement <- achievement %>%
gather(key = "time", value = "score", t2, t3) %>%
convert_as_factor(id, time)
```

Then we performed data analysis by groups of respondents. The obtained descriptive statistics are shown in table 1.

The study groups comprised an equal number of males and females. 40% of students had prior experience with Linux, with 44% of males and 37% of females having experience. An analysis of averages or medians indicates that students previously acquainted with Linux recorded a significant improvement in their academic performance during the second stage of the course (medium-term exam). During the third and final stage of the exam, there was almost no discernible impact of experience on the academic performance of female and male students. However, additional verification is required to confirm this assertion.

To use the mixed ANOVA method, we need to make some assumptions about our data.

There are no significant outliers in each student’s answers. This can be checked by using the function identify_outliers(). The output of this function shows that there were no extreme outliers (table 2).
Table 1
The summary statistics of the score variable.

<table>
<thead>
<tr>
<th>GroupID</th>
<th>Experience</th>
<th>Gender</th>
<th>Time</th>
<th>N_students</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>no</td>
<td>female</td>
<td>t2</td>
<td>17</td>
<td>77.90</td>
<td>12.90</td>
</tr>
<tr>
<td>2</td>
<td>no</td>
<td>female</td>
<td>t3</td>
<td>17</td>
<td>84.00</td>
<td>10.30</td>
</tr>
<tr>
<td>3</td>
<td>yes</td>
<td>female</td>
<td>t2</td>
<td>10</td>
<td>76.40</td>
<td>11.00</td>
</tr>
<tr>
<td>4</td>
<td>yes</td>
<td>female</td>
<td>t3</td>
<td>10</td>
<td>83.70</td>
<td>7.04</td>
</tr>
<tr>
<td>5</td>
<td>no</td>
<td>male</td>
<td>t2</td>
<td>15</td>
<td>71.90</td>
<td>13.80</td>
</tr>
<tr>
<td>6</td>
<td>no</td>
<td>male</td>
<td>t3</td>
<td>15</td>
<td>83.50</td>
<td>10.70</td>
</tr>
<tr>
<td>7</td>
<td>yes</td>
<td>male</td>
<td>t2</td>
<td>12</td>
<td>81.80</td>
<td>8.01</td>
</tr>
<tr>
<td>8</td>
<td>yes</td>
<td>male</td>
<td>t3</td>
<td>12</td>
<td>83.60</td>
<td>7.06</td>
</tr>
</tbody>
</table>

Table 2
The summary statistics of the score variable.

<table>
<thead>
<tr>
<th>GroupID</th>
<th>Experience</th>
<th>Gender</th>
<th>Time</th>
<th>N_students</th>
<th>Score</th>
<th>Is.Outlier</th>
<th>Is.Extreme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>no</td>
<td>female</td>
<td>t3</td>
<td>32</td>
<td>61</td>
<td>TRUE</td>
<td>FALSE</td>
</tr>
<tr>
<td>2</td>
<td>no</td>
<td>female</td>
<td>t3</td>
<td>33</td>
<td>62</td>
<td>TRUE</td>
<td>FALSE</td>
</tr>
<tr>
<td>3</td>
<td>no</td>
<td>male</td>
<td>t2</td>
<td>34</td>
<td>38</td>
<td>TRUE</td>
<td>FALSE</td>
</tr>
<tr>
<td>4</td>
<td>yes</td>
<td>male</td>
<td>t3</td>
<td>7</td>
<td>60</td>
<td>TRUE</td>
<td>FALSE</td>
</tr>
</tbody>
</table>

Normality. The dependent variable (students score in t2 and t3 period) should be approximately normally distributed in each row of achievement. We have done this using the shapiro_test() function. The output of this function (table 3) shows that almost all cases have a normal distribution (except row with 2 and 3 IDs).

Table 3
Checking the normality of the distribution of the dependent variable.

<table>
<thead>
<tr>
<th>GroupID</th>
<th>Experience</th>
<th>Gender</th>
<th>Time</th>
<th>Variable</th>
<th>Statistic</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>no</td>
<td>female</td>
<td>t2</td>
<td>score</td>
<td>0.927</td>
<td>0.197</td>
</tr>
<tr>
<td>2</td>
<td>no</td>
<td>female</td>
<td>t3</td>
<td>score</td>
<td>0.871</td>
<td>0.023</td>
</tr>
<tr>
<td>3</td>
<td>yes</td>
<td>female</td>
<td>t2</td>
<td>score</td>
<td>0.843</td>
<td>0.048</td>
</tr>
<tr>
<td>4</td>
<td>yes</td>
<td>female</td>
<td>t3</td>
<td>score</td>
<td>0.911</td>
<td>0.289</td>
</tr>
<tr>
<td>5</td>
<td>no</td>
<td>male</td>
<td>t2</td>
<td>score</td>
<td>0.947</td>
<td>0.475</td>
</tr>
<tr>
<td>6</td>
<td>no</td>
<td>male</td>
<td>t3</td>
<td>score</td>
<td>0.933</td>
<td>0.304</td>
</tr>
</tbody>
</table>

Since our sample size is greater than 50 students, the Shapiro-Wilk test becomes very sensitive even to a minor deviation from normality. Therefore we built the normal plot using ggqqplot() function (figure 2).

Another prerequisite for using mixed Anova is the homogeneity of variances. The scores’ variance needs to be even among the groups of between-subject factors. We utilised the levene_test() function. The results are presented in table 4. Therefore, the requirement of homogeneity of variances is also met.

Table 4
Checking the homogeneity of variances.

<table>
<thead>
<tr>
<th>ID</th>
<th>Time</th>
<th>df1</th>
<th>df2</th>
<th>Statistic</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>t2</td>
<td>3</td>
<td>50</td>
<td>0.560</td>
<td>0.644</td>
</tr>
<tr>
<td>2</td>
<td>t3</td>
<td>3</td>
<td>50</td>
<td>0.544</td>
<td>0.654</td>
</tr>
</tbody>
</table>

The last requirement is an assumption of sphericity. It means that the variance of the differences between gender and experience should be equal. The corresponding Mauchly’s test of sphericity is
automatically applied when using the anova_test() R function. Using the following fragment of the R code, we calculated three-way interaction between- and within-subject factors.

```r
res.aov <- anova_test(
    data = achievement, dv = score, wid = id,
    within = time, between = c(gender, experience))
```

Table 5 contains the results of the call of the function get_anova_table(res.aov). As can be seen from table 5, there was a statistically significant three-way interaction between gender, experience and scores obtained by students during repeated measures.

**Table 5**
Assessing the impact of gender, experience and time factors.

<table>
<thead>
<tr>
<th></th>
<th>Effect</th>
<th>DFd</th>
<th>F</th>
<th>P</th>
<th>P&lt;0.5</th>
<th>Ges</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>gender</td>
<td>50</td>
<td>0.011</td>
<td>9.17e-01</td>
<td>1.93e-04</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>experience</td>
<td>50</td>
<td>0.545</td>
<td>4.64e-01</td>
<td>1.00e-02</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>time</td>
<td>50</td>
<td>43.009</td>
<td>2.95e-08*</td>
<td>9.20e-02</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>gender:experience</td>
<td>50</td>
<td>1.120</td>
<td>2.95e-01</td>
<td>1.90e-02</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>gender:time</td>
<td>50</td>
<td>0.001</td>
<td>9.74e-01</td>
<td>2.56e-06</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>experience:time</td>
<td>50</td>
<td>4.463</td>
<td>4.00e-02*</td>
<td>1.00e-02</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>experience:gender:time</td>
<td>50</td>
<td>7.254</td>
<td><strong>1.00e-02</strong></td>
<td>1.70e-02</td>
<td></td>
</tr>
</tbody>
</table>

To determinate which levels of the dependent variable are statistically significant for this interaction, we have performed post-hoc tests such as

1. Two-way interaction test at each time levels. This test did not provide statistically significant significance between gender and experience factors. In addition, this test did not have a correct interpretation for our study.
2. Two-way interaction test to inspect the effect of experience on the students’ scores at every level of gender. The results of the function `anova_test(dv = score, wid = id, between = experience)` proved that there is a statistically significant relationship between the experience and learning achievements of men in the second stage of the study (table 6).

### Table 6
Relationship between the experience and learning achievements.

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender</th>
<th>Time</th>
<th>Effect</th>
<th>DFd</th>
<th>F</th>
<th>P</th>
<th>P&lt;0.5</th>
<th>Ges</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>female</td>
<td>t2</td>
<td>experience</td>
<td>25</td>
<td>0.092</td>
<td>0.764</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>male</td>
<td>t2</td>
<td>experience</td>
<td>25</td>
<td>4.82</td>
<td><strong>0.038</strong></td>
<td>*</td>
<td>0.162</td>
</tr>
<tr>
<td>3</td>
<td>female</td>
<td>t3</td>
<td>experience</td>
<td>25</td>
<td>0.0073</td>
<td>0.936</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>male</td>
<td>t3</td>
<td>experience</td>
<td>25</td>
<td>0.001</td>
<td>0.974</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

The last outcome can be proved by multiple pairwise comparisons. We grouped the data by time and gender and performed a pairwise comparison between experience levels (yes, no) using the Bonferroni adjustment. We performed the function `pairwise_t_test(score ~ experience, p.adjust.method = "bonferroni")` and got the value `p = 0.0376`. It is statistically significant. This confirms that there was an impact from previous experience of using Linux in male students in the second stage of the authors’ course. There was no significant difference between the groups of women with and without experience at both stages of the study (`p2 = 0.764` and `p3 = 0.974`). This outcome can be visualized using plots of academic achievement score by gender coloured by experience faceted by time (figure 3).

\[ Anova, F(1,50) = 7.25, \rho = 0.01, \eta^2 = 0.02 \]

![Figure 3: The box plots of summary statistics.](image)

Having analyzed figure 3, we can state some additional outcomes such as

- Medians indicate an increase in students’ scores from the 2nd to the 3rd stage of the study. This was observed for all groups of students (men, women, with or without experience).
- The women’s experience in the 2nd stage of the study did not have a positive effect on their success. This may be due to their inadequate self-assessment.
The value of all upper quadrants of assessments from the 2nd to the 3rd stage has increased. This indicates that the number of students with a high level of academic achievement has increased.

- Non-extreme emissions were observed in the study. They usually indicated students who did not complete the course (received less than 60 points in total).
- According to the results of the 3rd stage of the research, the success of students of all groups became almost the same.

3.2. Study of correlations between factors of satisfaction and students’ abilities

Extending the research, we tried to process some data from the questionnaire “End of Course Feedback”. This questionnaire is a standard component of the Cisco NDG Linux Essentials course. We have analyzed two items of the questionnaire. Here they are.

1. Please rate your level of satisfaction with the following aspects of this course:
   - On-line curriculum materials.
   - Lab exercises.
   - Access to virtual machines.
   - Assessments (including quizzes, chapter exams, and the final exam).

2. Please rate how confident you feel in your ability to do each of the following:
   - Explain Open Source.
   - Explain the value of Linux.
   - Provide examples of Linux in industry.
   - Use the Linux command line.

These questions have a Likert-scale style. We transformed students’ answers into a 5-position scale (from 0 points – Very Dissatisfied (Not at All Confident) to 4 points – Very Satisfied (Completely Confident)). The scores corresponding to this scale are contained in the dataset (reference was given above). The table contains columns such as S_OnlineMatherials, S_Labs, S_AccessVM, S_Assessments. They answer the first question (about the level of satisfaction). The columns A_ExplainOpenSource, A_LinuxValue, A_UseCLI, and A_LinuxIndustry answer the second question (about students’ abilities).

We selected these questions because they are crucial to comprehend students’ perceptions towards the course’s contents, techniques, and tools. They enable us to establish the association between self-esteem and academic performance. This allows learning about students’ self-assessment impartiality. We examine if the points distribution of each column is normal to decide the correct correlation method. The results were obtained after invoking the shapiro.test function are presented below.

- \( p\text{-value}(S_{\text{OnlineMaterials}}) = 2.004\text{e-08} \)
- \( p\text{-value}(S_{\text{Labs}}) = 7.922\text{e-09} \)
- \( p\text{-value}(S_{\text{AccessVMs}}) = 4.38\text{e-10} \)
- \( p\text{-value}(S_{\text{Assessments}}) = 3.621\text{e-08} \)
- \( p\text{-value}(A_{\text{ExplainOpenSource}}) = 4.38\text{e-10} \)
- \( p\text{-value}(A_{\text{LinuxValue}}) = 6.867\text{e-09} \)
- \( p\text{-value}(A_{\text{UseCLI}}) = 7.61\text{e-09} \)
- \( p\text{-value}(A_{\text{LinuxIndustry}}) = 1.085\text{e-06} \)

As all \( p\)-values are below 0.05, it is evident that the distributions are not normal. Consequently, the Spearman rank coefficients should be calculated. Spearman rank coefficients are a statistical measure of the strength of a monotonic relationship between paired data. The correlation is a measure of the effect size. This coefficient assesses whether the quantitative factor (scores of students from the third stage) influences the quantitative response (S_ and A_ variables). The absolute value of the coefficient is interpreted based on the modified Cheddock scale [54].
• 0 – 0.19 – relationship is very weak;
• 0.2 – 0.39 – relationship is weak;
• 0.40 – 0.59 – relationship is moderate;
• 0.60 – 0.79 – relationship is strong;
• 0.80 – 1.0 – relationship is very strong.

A positive coefficient indicates the presence of a direct relationship between the factor and response variables. On the other hand, a negative coefficient suggests an inverse relationship between the two variables. The rank correlation coefficients were computed and displayed using the “corrplot” library in R. Table 7 and table 8 display the correlation coefficients that were computed.

Table 7
The Spearman rank coefficients for the level of satisfaction.

<table>
<thead>
<tr>
<th></th>
<th>S_OnlineMatherials</th>
<th>S_Labs</th>
<th>S_AccessVMs</th>
<th>S_Assessments</th>
</tr>
</thead>
<tbody>
<tr>
<td>t3_scores</td>
<td>0.54</td>
<td>0.50</td>
<td>0.61</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Table 8
The Spearman rank coefficients for the level of ability.

<table>
<thead>
<tr>
<th></th>
<th>A_ExplainOpenSource</th>
<th>A_LinuxValue</th>
<th>A_useCLI</th>
<th>A_LinuxIndusy</th>
</tr>
</thead>
<tbody>
<tr>
<td>t3_scores</td>
<td>0.37</td>
<td>0.61</td>
<td>0.46</td>
<td>0.23</td>
</tr>
</tbody>
</table>

According to table 8, there is a strong correlation between students’ understanding of Linux principles and scores. Unfortunately, the correlation between the same scores and students’ understanding of the Open-Source ideology and the possibilities of using Linux in the industry was weak.

All questions in the course feedback questionnaire were analysed for their correlations. From this figure we can draw a few more conclusions:

• There are no negative values of correlation coefficients in this questionnaire. This means that for all questions there is a positive relationship between student grades in the course feedback.
• The greatest positive relationship exists between questions of satisfaction with laboratory works and the ability to use CLI.
• Satisfaction with access to virtual machines has positive correlations in self-assessment of all student abilities. It is hoped that students understand the importance of practical experience using Linux and the convenience of working with this OS using modern cloud technologies.
• There is no correlation between job satisfaction and online materials. Students may have been dissatisfied with a significant number of additional assignments, including essays. These tasks are individual and increase the amount of time required to complete the course. In general, this correlation needs further study.

4. Conclusions

Teachers can utilise the technological and methodical benefits of a cloud-oriented environment through a combination of face-to-face and online training. Implementing combined learning approaches using a cloud lab method leads to a more efficient use of computing resources and acquiring time. Additionally, this enhances the learning process by teaching students how to manage their own time, which better prepares them to complete the course successfully.

The analysis of experimental data suggests that practical laboratory exercises are preferred by students over theoretical ones. Students understand the concept better when a short theoretical lecture is immediately followed by appropriate practical tasks. These lectures provide students with an
The authors aim to develop students' competencies in using Linux by assigning them laboratory work, essays, or tests that require them to complete their answers. A correlation analysis conducted in the Operating Systems course revealed that students understand the significance of the assigned tasks. The pedagogical experiment in the mentioned course did not provide evidence of the impact of previous experience or gender of students on their academic performance. Consequently, it can be reasonably concluded that the author’s methodology of using cloud laboratories as a primary factor has a positive effect on enhancing the success rate of studying the Operational system Linux by future computer science teachers.

Our study discovered a minimal influence of students’ previous experience or gender on their academic performance. Thus, the primary means of enhancing our students’ performance lies in the authors’ approach of employing cloud laboratories. It can be asserted that it is advantageous to introduce such laboratories in universities and colleges.

References


