Zhang et al., Int J Comput Softw Eng 2022, 7: 181
https://doi.org/10.15344/2456-4451/2022/181

International Journal of
Computer & Software Engineering
Original Article Open Access

Summative Assessment of Undergraduate Learning Outcomes with Cross-Validation in C Programming Course

Yongbin Zhang[1], Hongxia Hu[2], Xiuli Fu[3*] and Yanying Zheng[4]
[1]Department of Mechanical and Electrical Engineering, Beijing Institute of Graphic Communication, Beijing, China
[2]School of Marxism, Beijing Institute of Graphic Communication, Beijing, China
[3]Department of Information Engineering, Beijing Institute of Petrochemical Technology, Beijing, China
[4]School of Biology and Environment, Beijing University of Agriculture, Beijing, China

Abstract

Learning outcomes have gained more and more attention in higher education. Course learning outcomes are essential for all stakeholders. There are many approaches to evaluate course learning outcomes. However, one single method cannot measure course learning outcomes preciously. Each method has its strength and weaknesses. For example, grades tend to measure knowledge but lack standardization. We presented a novel way to assess course learning outcomes with the cross-validation principle. Experiments show that this method can measure higher-level learning outcomes, as applying skills or knowledge in different contexts. The new approach provides a different perspective for assessing course learning outcomes.

Publication History:

Received: October 19, 2022
Accepted: November 02, 2022
Published: November 04, 2022

Keywords:

Summative assessment, Learning outcomes, Cross-validation, Higher education

Introduction

Learning outcomes in higher education have attracted more and more attention from various stakeholders. Universities collect learning outcomes as a source for self-improvement. Governments desire to obtain learning outcomes to compare the quality of universities across institutions, states, and countries. Employers want to know which universities could provide qualified students according to learning outcomes. Parents want to know if what they pay for is worth what they gain based on learning outcomes. And students want to know which university they should choose based on available information about learning outcomes. Therefore, all stakeholders regard learning outcomes as an indicator of the quality and effectiveness of universities [1].

Learning outcomes are also related to the sustainable development of universities. The COVID-19 pandemic has hurt numerous industries and slowed down the growth of the global economy. Due to the decline of economic development, governments have reduced the budget for higher education. At the same time, the reduction of population in developed countries also puts stress on universities to attract enough excellent students. Therefore, universities must persuade their governments to support them financially with efficient and sustainable learning outcomes. Meanwhile, universities attract potential students by the high-quality learning outcomes of their programs.

Learning outcomes can improve teaching and facilitate learning and are indispensable parts of meaningful education. Many researchers have demonstrated innovative approaches to improve learning outcomes because student-centered pedagogies and outcome-based education become more popular in higher education. For example, mobile devices improve learning outcomes in higher education[2], and Bloom's taxonomy achieves better learning outcomes in a PLC and robotics course[3].

However, learning is a complex process. It is almost impossible to measure students' learning outcomes preciously. Sometimes even an assessment shows outstanding learning outcomes, but it might not be the truth like the over-fitting phenomena in training a supervised model. Therefore, in this paper, we present a novel method for assessing course learning outcomes of undergraduate cognitive learning. One compulsory course, C programming language, was chosen in our research. Students were divided into two groups (classes) randomly. Two instructors taught the course. Each of them was in charge of one group. We designed and kept the two groups the same as possible to reduce external intervention. Two groups had the same syllabus, textbook, and class schedule except for the classroom and the instructor. At the end of the course, two groups participated in the same examination. At the end of the course, two groups participated in the same examination. One of two instructors designed the test while keeping the test from the other instructor. The results showed a statistically significant difference between the grades of the two groups. This summative assessment method can reveal student course learning outcomes from a fresh perspective.

Course Learning Outcomes

Learning outcomes refer to the learner changes as a result of learning [4]. The concepts of learning outcomes and learning objectives are two different terms often misused in the teaching and learning context. For example, the study defined learning outcomes as statements of the knowledge, skills, and abilities a student should possess at the end of the learning experience(s) [5]. It is the definition for learning objectives instead of learning outcomes. Learning outcomes describe what a student achieves while learning objectives are what instructors or institutions expect the student to gain by learning.

Therefore, it is possible to design and express learning objectives in specific statements. But learning outcomes have to be measured with

***Corresponding Author:** Xiuli Fu, Department of Information Engineering, Beijing Institute of Petrochemical Technology, Beijing, China, E-mail: fuxiuli@bipt.edu.cn

Citation: Zhang Y, Hu H, Fu X, Zheng Y (2022) Summative Assessment of Undergraduate Learning Outcomes with Cross-Validation in C Programming Course. Int J Comput Softw Eng 7: 181. doi: https://doi.org/10.15344/2456-4451/2022/181

different approaches. Keeping these differences in mind will help both instructors and students focus on outcome-based education, from which the term learning outcomes has derived.

Various learning outcomes categories exist according to different perspectives. Learning outcomes can be classified into generic and disciplinary learning outcomes [6]. The generic learning outcomes emphasize the skills and abilities regardless of the specific subject area and work as learning outcomes benchmarks across disciplines, institutions, or even countries. And the disciplinary learning outcomes are competencies related to a specific subject or profession.

Learning outcomes have also been categorized into program and course learning outcomes. The relationship between the two is like goals and objectives. The program learning outcomes are broad and abstract, while the course learning outcomes are specific and concrete. Both learning outcomes are strongly correlated. A mapping could be constructed between the two if related courses outcomes are available [7].

Learning outcomes can also be grouped into the cognitive domain, the affective domain, and the psychomotor one with the taxonomy of educational objectives, which has been designed by Bloom and his colleagues [8]. The definitions of objectives in the cognitive domain are classified into six levels, including remember, understand, apply, analyze, evaluate, and create from the lowest to the highest level. The objectives in the cognitive domain are the clearest and the most detailed among the three parts, and these objectives have been widely referenced.

Learning is a neurologic process that is internal and invisible. It is not always possible to directly measure learning outcomes. Primarily, it is difficult to state the objectives precisely in the affective and psychomotor domains [8]. Therefore, we focus on assessing the course learning outcomes in the cognitive field in this paper.

Assess Learning Outcomes Method

Various approaches have been employed in measuring learning outcomes. Universities and consulting companies usually ask students to evaluate their learning outcomes through surveys. Questions are organized under different dimensions such as knowledge, skills, and attitude. These self-reported outcomes are collected to reflect the quality of programs provided by universities. Some countries have developed testing programs to assess the learning outcomes of higher education institutions nationally. For example, the United States created the Collegiate Learning Assessment (CLA), and Australia developed the Australian Council for Educational Research (ACER). The Organisation for Economic Co-operation and Development (OECD) has initiated the study on the Assessment of Higher Education Learning Outcomes (AHELO) to measure learning outcomes internationally.

Different assessment approaches have their advantages and disadvantages. Self-report learning outcomes surveys are easy to implement and are resource effective. However, their reliabilities are doubtful because students tend to overestimate or underestimate their learning outcomes [9]. The CLA is designed to measure higher-order generic skills with open-ended questions. And the ACER combines multiple-choice questions with open-ended questions for evaluating specific skills. But both CLA and ACER are time-consuming for students, which is challenging to motivate students to take this kind of test [1].

A student's learning outcomes can be directly measured with grades and tests. The learning outcomes from students can be aggregated to assess the performance of a program or school. However, the basic problem with grades is the lack of standardization [6].

Although there are weaknesses with grades and tests, grades and tests are still used and accepted as a measure to assess students' acquired knowledge. Therefore, we present an innovative method to evaluate students' course learning outcomes with grades and tests.

Assessments can be categorized into formative and summative assessments based on their aims [7]. Formative assessments are assessments for learning, which means the purpose of formative assessments is to help students learn. Therefore, feedback on students' learning outcomes and change plans for improving learning outcomes often follow a formative assessment. Summative assessments are assessments of learning. It means summative assessments aim to evaluate students' learning outcomes and their achievements. Final examinations at the end of a course are examples of summative assessments. This research focuses on summative assessments of course learning outcomes to validate the new method.

Students' abilities to apply what they learn in different contexts are indicators of deep learning [10]. These learning outcomes are higher-level ones. However, instructors evaluate students' course learning outcomes within the same or familiar context under which students learn. Therefore, traditional assessments of course learning outcomes lead to suspicion of their reliability.

In data science, when predictive models are trained with samples, there is a potential risk of over-fitting, such as the one in the deep neural network[11]. Over-fitting means an artificial intelligence model fits precisely against its training sample but cannot perform accurately against unseen data [12]. Over-fitting leads to the lack of generalization of a model, which defeats the purpose of the model. There are different approaches to reduce or avoid the over-fitting issue for training a supervised model, including early stop or regularization [13].

The k-fold cross-validation method is commonly used to solve over-fitting issues. The main idea of the traditional k-fold cross-validation method consists of two steps [14]. The first one is dividing the dataset into k folds randomly. The second step is training the model with the cases of each fold withheld in turn from the training set.

These methods for evaluation the correct rate of predictive models give us hints to assess learning outcomes in higher education. Although there are many kinds of research on assessing learning outcomes, few consider the effects of learning over-fitting and applying knowledge in a new context. Effective learning requires students must learn when and how to apply their skills and knowledge [15].

Therefore, it is necessary to evaluate student course learning outcomes in a new context different from their familiar context. We presented a new way to assess student course learning outcomes based on the cross-validation idea in this paper.

Experiment Design

In our experiment, we evaluated students learning outcomes of the C programming language. It is a compulsory course for students majoring in mechanical engineering at the Beijing Institute of Graphic Communication. The C course comprises thirty-hour lectures and

Citation: Zhang Y, Hu H, Fu X, Zheng Y (2022) Summative Assessment of Undergraduate Learning Outcomes with Cross-Validation in C Programming Course. Int J Comput Softw Eng 7: 181. doi: https://doi.org/10.15344/2456-4451/2022/181

eighteen-hour labs. Students took two-hour lectures or labs twice a week for the C course during twelve successional weeks.

The C course is the first programming language for the students in the university. Therefore students' prior knowledge has little impact on their learning.

Participants were sophomore students from the same major, mechanical engineering. All students had just finished their first year of study and were in the third semester. These students were classified into two groups randomly for the C programming course in the 2020-2021 spring semester. One of the groups (Group A) consisted of fifty-three students (N1=53, male=36, female=17) and the other group (Group B) contained fifty-four students (N2=54, male=37, female=17).

Two instructors taught the two groups, respectively. One of the instructors (Instructor A) had a Ph.D. degree in computer science and taught the C programming course for more than six years. The other instructor (Instructor B) had a master's degree in mechanical engineering and led the C programming course for more than five years.

Two groups used the same textbook, course syllabus, and class time except for different classrooms and different instructors. The course lasted for twelve weeks. Then students spent two weeks preparing for the final examination. One of the instructors (Instructor B) designed the test for the final exam and kept the test confidential so that nobody else knew the content of the test until students finished their exam.

Both instructors (A, B) did not participate in the grading process. After the exam, we invited an experienced teacher (Instructor C) who had been teaching the C program course for more than ten years to grade all students' exam work. The whole process is shown in Figure 1.

In the experiment, Instructor B designed the test to assess the course learning outcomes of students taught by Instructor A. This approach is based on the idea of cross-validation methods. Students took the exam at the end of the course, which is a summative assessment.

Results

The distribution of grades for Group A is shown in Figure 2. And for Group B, it is shown in Figure 3. Both grades of groups are bimodal data. But the modes in Group B are higher than those in Group A.

The statistical descriptions for both groups are shown in Table 1. In group A the lowest grade is 21; the highest one is 96; the mean is 58.57; the standard deviation is 16.26. For group B, the lowest grade is 27; the highest grade is 90; the mean is 65.6; the standard deviation is 18.15. The mean of grades in Group B is higher than the mean in Group A.

Discussion

The experiment showed that students had higher grades in Group B. From the traditional perspective, students in Group B achieved more learning outcomes than students in Group A did. However, it is still early to draw this conclusion.

After the experiment, all three instructors (Instructor A, B, and C) discussed the test. All agreed that the designed test and the C course syllabus were aligned. Therefore, the test matches the learning objectives of the course.

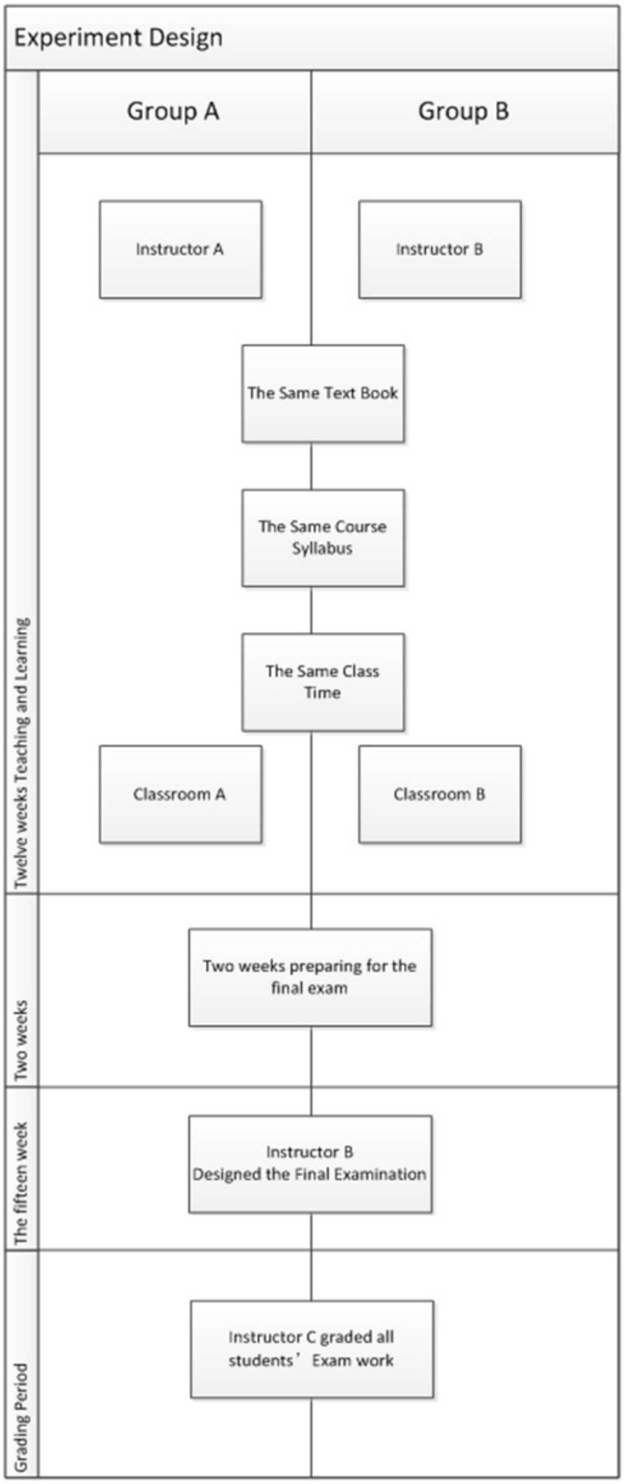

Figure 1: Process for the experiment.

Grade	Descriptive Statistics				
	N	Minimum	Maximum	Mean	Std. Deviation
Group A	53	21.00	96.00	58.57	16.26
Group B	54	27.00	90.00	65.56	18.15

Table 1: Descriptive statistics for both groups.

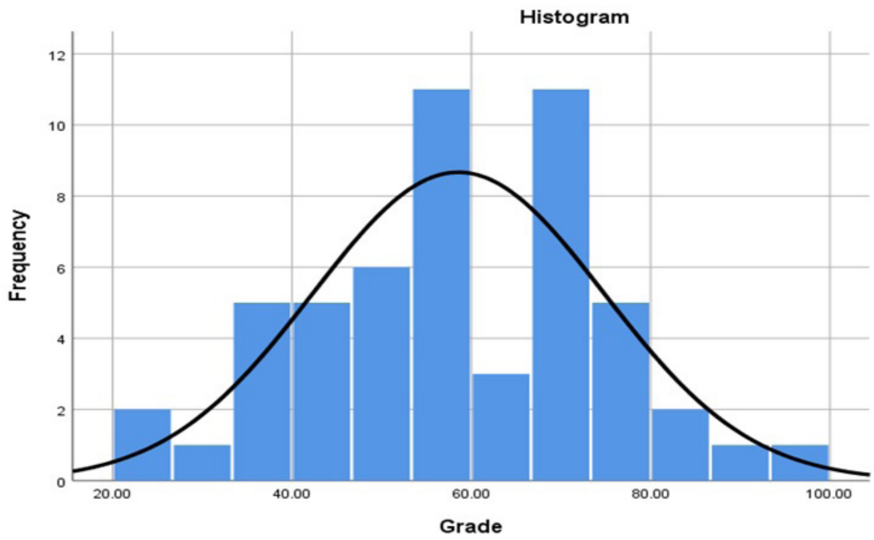

Figure 2: Grades distribution for Group A.

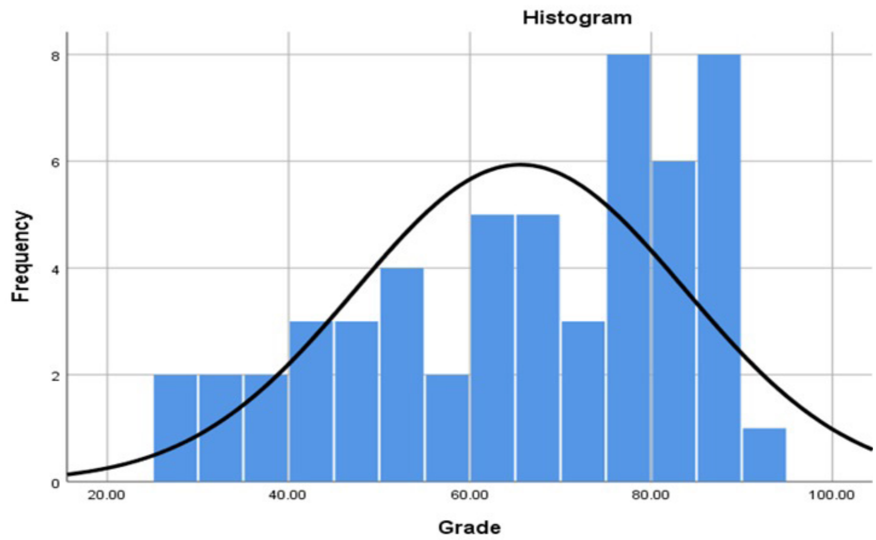

Figure 3: Grades distribution for Group B.

We also reviewed the grades of the C course in the 2019-2020 spring semester for the same major but different students. In the 2019-2020 spring semester, there were two instructors and two classes for the C course. Each taught one class, respectively. Each instructor designed a test to evaluate students learning outcomes. There was no significant difference in grades between the two classes ($n^1 = 38$, mean = 82, sd = 8.83; n^2=47, *mean* = 84, sd = 8.39, $p = 0.385 > 0.05$).

The results mean no statistical difference in grades between students from two groups when an instructor was responsible for teaching and designing the test to evaluate the same students. However, students got better marks when their instructor created the test than students whose instructor did not participate in the assessment test.

Therefore, we can conclude that the cross-validation method provides a novel perspective to assess students' course learning outcomes. This method measures what students remember. It also verifies how students apply what they know. It is consistent with the learning principles [15]. This method measures the capability of using skills or knowledge in different contexts, which is compatible with the existing research [10, 16].

We employed this method as a summative assessment to measure course learning outcomes from a novel view. It can help both instructors and students to understand course learning outcomes. However, instructors should not decide course learning outcomes based on this assessment alone. Different assessment approaches help us understand course learning outcomes better than a single method.

On the other hand, assessment should be a process to understand and improve student learning. The presented cross-validation method helped teachers understand student learning outcomes better than traditional methods. However, it is still muddy how to apply the cross-validation method to improve student learning outcomes.

Conclusion

We provided a cross-validation method to evaluate student learning outcomes in this paper. Two instructors were responsible for the same course but different classes. One of the two designed the test as the summative evaluation for all students. The results showed this novel method measured the student learning outcomes better than traditional ones. Traditional student learning outcomes assessment may not be accurate when an instructor teaches and evaluates the same students. The presented approach also provides a new aspect for instructors to view student learning outcomes.

However, this method did not improve student learning outcomes yet. In the future, we will go further and discover how this novel method can be applied in formative assessment to improve student course learning outcomes.

Funding

Yongbin Zhang appreciates the Beijing Municipal Education Commission for supporting undergraduate instructional reform (Grant No. 22150122001) and the Beijing Institute of Graphic Communication for funding the essential project of instructional reform (Grant No. 20231001). Hongxia Hu appreciates the Beijing Institute of Graphic Communication for supporting the reform of political and professional courses in communication universities.

Competing Interests

The author declare that he has no competing interests.

References

1. Douglass JA, Thomson G, Zhao CM (2012) The Learning Outcomes Race: The Value of Self-Reported Gains in Large Research Universities. Higher Education 64: 317-335.
2. Garcia-Martinez I, Fernandez-Batanero JM, Sanchiz DC, de la Rosa AL (2019) Using Mobile Devices for Improving Learning Outcomes and Teachers' Professionalization. Sustainability 11: 6917.
3. Gummineni M (2020) Implementing bloom's taxonomy tool for better learning outcomes of plc and robotics course. International Journal of Emerging Technologies in Learning 15: 184-192.
4. Nusche D (2008) Assessment of Learning Outcomes in Higher Education: a comparative review of selected practices", OECD Education Working Papers, No. 15, OECD Publishing, Paris.
5. Soares D, Carvalho P, Dias D (2020) Designing Learning Outcomes in Design Higher Education Curricula," International Journal of Art & Design Education 39: 392-404.
6. Caspersen J, Smeby JC, Olaf Aamodt P (2017) Measuring Learning Outcomes. European Journal of Education 52: 20-30.
7. Keshavarz M (2011) Measuring Course Learning Outcomes. Journal of Learning Design 4: 1-9.
8. Bloom BS (2001) A taxonomy for learning, teaching, and assessing a revision of Bloom's Taxonomy of educational objectives. New York; San Francisco; London (etc): Longman.
9. Dochy F, Segers M, Sluijsmans D (199) The Use of Self-, Peer and Co-assessment in Higher Education: a review. Studies in Higher Education 24: 331-350.
10. Barnett SM, Ceci SJ (2002) When and Where Do We Apply What We Learn? A Taxonomy for Far Transfer. Psychol Bull 128: 612.
11. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 15: 1929-1958.
12. Mutasa S, Sun S, Ha R (2020) Understanding artificial intelligence based radiology studies: What is overfitting? Clinical Imaging 65: 96-99.
13. Ying X (2019) An Overview of Overfitting and its Solutions. Journal of Physics: Conference Series 1168: 022022.
14. Krstajic D, Buturovic LJ, Leahy DE, Thomas S (2014) Cross-validation pitfalls when selecting and assessing regression and classification models. J Cheminform 6: 10.
15. Ambrose SA (2010) How learning works : seven research-based principles for smart teaching. San Francisco, CA: Jossey-Bass.
16. Cukurova M, Bennett J, Abrahams I (2018) Students' Knowledge Acquisition and Ability to Apply Knowledge into Different Science Contexts in Two Different Independent Learning Settings. Research in Science & Technological Education 36: 17-34.