The influence of assessment types on students' performance in Machine Learning Education

An analysis of students' learning gain in k-means clustering

Madeline El Aissati[1]

Supervisor: Gosia Migut[1]

[1]**EEMCS, Delft University of Technology, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 23, 2024

Name of the student: Madeline El Aissati
Final project course: CSE3000 Research Project
Thesis committee: Gosia Migut, Mark Neerincx

Abstract

With the increasing influence of Machine Learning (ML) on our lives, the need for education on this topic is growing. A key component of education is assessment and improving this aspect could lead to better student learning performance. This study aimed to investigate the influence of different assessment methods on students' learning performance in k-means clustering. Two different assessment methods were used: a closed-book problem-based assignment and an open-book short answer exam. Participants were notified of their assessment method, after which they were instructed to watch a video lecture and take the assessment. Results show a significantly improved learning gain when using the open-book assessment, where learning gain was defined as the difference in score between the pre- and post-test. Between these two methods the open-book assessment is therefore favourable. However, future research is needed to develop a validated concept inventory for k-means clustering and identify other possible assessment techniques.

I Introduction

While demand for Artificial Intelligence (AI) expertise is growing [1], Ko [2] highlights the fact that we have little knowledge of what students need to know and what knowledge is necessary to teach Machine Learning. This includes methods to assess knowledge of concepts in the field of Machine Learning.

Assessment plays a crucial role in improving education methods and understanding how students learn [3], by determining to what extent learning objectives have been met. Consequently, it provides a means to measure student learning [4] [5] [6]. However, little research exists on the relation between assessment methods and students' performance [7]. Islam *et al.* [7, p. 1] argues that "A good understanding of this association would lead to mature assessment methods that could evidently enhance student-learning process".

In [8], methods for assessing computational thinking (CT) have been discussed, aiming to identify the main features of CT assessment methods. While these attempts provide insight into different assessment techniques, research on the assessment of Machine Learning is scarce [6].

This study aims to investigate the influence of different assessment methods on students' learning gain in k-means clustering, as measured by their obtained grade. By comparing two different assessment methods in a controlled experiment on university students, the goal is to find one specific method which results in a higher learning gain of k-means clustering. According to a survey conducted among 105 engineering students, problem-based assessments were considered as the best way for student learning [9], followed by open-book examination. While closed-book examination could encourage cramming and diminish CT skills [10], the problem-based assessment was made closed-book, to ensure students would need to apply their knowledge, rather than recall.

The main contribution of this paper is the presentation of a method for analysing students' learning gain in k-means clustering. Results show a significantly higher learning gain when using an open-book short-answer exam, as opposed to a closed-book problem-based assignment. Further research is needed to investigate the validity and reliability of the assessments used, as well as identifying other possible assessment techniques for k-means clustering.

The structure of this paper is as follows: first, relevant literature is discussed in the Background section, after which the Method describes the participant pool and the experiment in four steps. Next, the section on Responsible Research discusses ethical implications of the study and mitigation strategies. Results are presented in section V. The paper concludes with a discussion of limitations and suggestions for future work on the topic at hand.

II Background

The Teaching-Learning-Assessment Cycle

In *Designing better engineering education through assessment* [11], Suskie [3] defines assessment in terms of the four steps which need to be taken according to the Teaching-Learning-Assessment Cycle:

1. Formulate unambiguous statements of the expected learning outcomes.
2. Design learning experiences to guide students towards achieving these outcomes.
3. Implement methods to measure student achievement of the learning outcomes.
4. Use the results to improve teaching and learning.

The first step can be achieved using Bloom's taxonomy [12], which can be seen in Figure 1, a framework for categorizing learning objectives. Assigning each learning objective a level of complexity will provide a basis for designing the learning experience in the second step.

The third step requires knowledge of different assessment types. Typically, assessment methods are divided into two main categories: formative and summative assessment. Formative assessment is often used during an instruction to provide immediate feedback to the educator [13], whereas summative assessment is used to measure learning performance after completing an instruction. Assessments can also be categorized into direct and indirect: direct involves judging students' work directly, while indirect makes use of students' opinions. The Accreditation Board for Engineering and Technology (ABET) emphasizes the importance of using direct assessment [3].

The final step can provide insight into changes which need to be made, such as reformulating learning outcomes or improving the learning experience.

Assessment and student performance

The study of Islam et al. [7] is among the few to delve into the relation between assessment methods and students' performance. The effect of four different assessment formats is investigated in the context of a networking course. It was found that collective assessment formats provide students with the

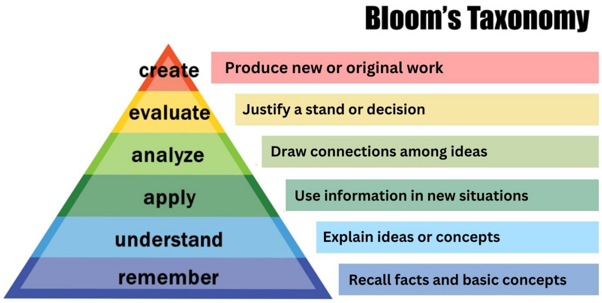


Figure 1: Bloom's revised taxonomy. Adapted from [14]

opportunity to demonstrate their knowledge effectively. A notable flaw in this research is the lack of a pre-test: the learning performance of students may have been influenced by their prior knowledge. Additionally, the different assessment formats were not substantiated and assessed different topics across the course, making it difficult to single out a suitable assessment for a specific topic.

The impact of assessment environments on student learning responses has also been investigated by Jimaa [15], developing a method to characterize assessment environments. By presenting a survey on students' experiences with different assessment environments, the paper concludes that a high volume of formative assessment and feedback results in students being more positive.

Assessment in ML-related fields

In the field of engineering education, assessment techniques have been subject to different studies [11][9][16]. A survey conducted amongst 105 engineering students at Queensland University of Technology [9], found that students considered problem-based assessments as the best way for student learning. The next two popular assessment methods were open-book in-class solving and open-book final examination.

The field of Computational Thinking (CT) can help acquire certain Science, Technology, Engineering and Mathematics (STEM) skills [8], such as mathematics, statistical analysis and programming algorithms, skills which are frequently used in ML. While the assessment of CT has been the subject of numerous studies [8][17], a single study was found on the assessment of ML [6]. The study created an overview of the different methods used in assessing the learning of ML in K-12, but found the assessments lacking methodology, constructive feedback, validity and reliability.

Concept Inventories

The scarcity of research on ML assessment is also visible in the field of Concept Inventories: useful tools to measure students' conceptual understanding of a topic. First developed in physics by David Heskenes and his colleagues in 1992, CIs have since been adopted in many STEM fields [18]. While CIs are abundant in fields such as physics and biology, no widely used CIs have been developed for computer science [19]. This is visible in the field of ML as well: only one CI was found on the understanding of AI for middle school students [20]. Existing CIs on computer-science related topics such as calculus and statistics could be used as a basis for developing CIs in computer science [19].

III Method

This section describes the procedure used to teach k-means clustering, select participants and to quantitatively assess participants' performance in k-means clustering.

A Teaching k-means clustering

Following the "Four Steps of the Teaching-Learning-Assessment Cycle" as described by Suskie [3], the first two steps consisted of defining the learning objectives and designing the learning experience to achieve these objectives.

Step 1. Define the learning objectives

Participants received a lesson on k-means clustering: a method used to group unlabelled data into different clusters, with k specifying the amount of clusters. This topic was chosen as it is the most popular clustering method and has been widely studied [21]. Three different lectures on k-means [22] [23] [24] were used to prepare a list of learning objectives as defined in Table 1. The objectives focus on the lower levels of Bloom's taxonomy (Figure 1), as the goal of the explanation is to provide the participants with a basic understanding of k-means. Due to the varying backgrounds and skills of the prospective participants, the implementation of k-means was not included in the learning objectives.

Table 1: Learning objectives for k-means clustering.

ID	Learning Objective	Level
LO1	Explain the difference between supervised and unsupervised learning	Remember
LO2	Name two pros and two cons of the k-means algorithm	Remember
LO3	Name three different convergence criteria for the k-means algorithm	Remember
LO4	Explain what clustering is and name two applications	Understand
LO5	Explain the four main steps of the k-means clustering algorithm	Understand
LO6	Calculate Euclidean distance between two given 2D points A and B	Apply
LO7	Calculate the sum of squared errors for a given set of points and centroids.	Apply

Step 2. Design learning experience to achieve objectives

This step involved deciding on the amount of time for the explanation of k-means and the amount of time participants would receive to study. Current teaching practices for k-means were reviewed [22] [23], the average duration of these explanations was approximately 30 minutes. It was therefore decided to design an instruction of this length. To ensure each participant received the exact same instruction, a video recording was used. It should be noted that at this stage, participants were notified of the type of assessment they would receive. Participants watched the video and were allowed to take notes; pausing the video was not allowed. After the video, participants were given 20 minutes to study.

Step 3. Choose appropriate assessment methods
The third step consisted of deciding on assessment techniques. As this assessment takes place after completing the instruction of k-means clustering, summative assessment was deemed most appropriate. Jimaa [15] however encourages the use of formative assessment and feedback, therefore formative elements were included in the instruction, such as multiple choice questions. Had the study been conducted over a whole course, students would have received feedback on their mistakes as well.

In order to compare performance across dissimilar assessment formats, a pre-test and a post-test must be conducted [6]. To this end, a CI on k-means clustering was developed, using the CI in [20] as inspiration (see Appendix). The topics per question can be found in Table 2. The process of validation has been left for a future study as the focus was laid on ensuring the questions cover as many learning objectives as possible.

Table 2: Topics of questions in pre- and post-test with their corresponding learning objectives

ID	Topic	Learning Objective
1	Supervised or unsupervised learning	LO1
2	Supervised or unsupervised learning	LO1
3	True/false clustering statements	LO4
4	Convergence criteria of k-means	LO3
5	Calculating the sum of squared errors	LO7
6	Disadvantages of k-means	LO2
7	Conducting one iteration of k-means	LO5, LO6

Following the survey conducted at Queensland University of Technology [9], the two assessment methods will be a closed-book problem-based assignment, and an open-book short-answer exam (see Appendix).

While these assessment methods are important for the learning process of the participants, they will primarily serve to simulate the process of a course: the results of these assessment methods will be analysed for common misconceptions, but they are not used to measure participants' learning performance. This is done using the pre- and post-test results of the participants.

Measuring learning performance
The learning performance, defined as the difference between the pre- and post-test scores, will be measured using the average of gains adapted from Hake's formula [25], which is the standard measure used for reporting scores on concept inventories. The normalized gain for each student is calculated individually using the pre- and post-test scores, with 7 being the highest possible score:

$$\frac{Post_i - Pre_i}{7 - Pre_i} * 100\% \tag{1}$$

These individual scores are then used to calculate the average of gains, as seen in equation 2.

$$\text{Average of Gains} = \frac{\sum_{i=1}^{n} \text{Gain}_i}{n} \tag{2}$$

where: n is the total number of gain values.

Lastly, to account for participants choosing the blank option, the point system will be as follows:

- 1 point for each correct answer
- 0 points for each blank answer
- -1 points for each incorrect answer

B Experimental set-up

20 participants took part in this study (M = 23.4 years, SD 2.7). The experiment used a between-subject design and exclusion and inclusion criteria were as follows:

- The study was performed on people with no prior knowledge of k-means clustering. This was determined through a pre-test: participants were asked to confirm no prior knowledge of k-means clustering and a pre-test score above 50% resulted in exclusion.
- Eligible participants were currently studying or had completed a bachelor's degree, or were pursuing a master's degree.

Participants were divided into two groups at random, while maintaining the female:male ratio. Each group received a different assessment (see Figure 2). In a larger study, it would be appropriate and useful to add a control group which is not notified of the assessment method, to verify the assumption that students will adjust their study techniques if they know how they will be assessed. This experiment did not include this control group, because in general students are always aware of the final assessment format of a course. Additionally, the limited amount of time would not have allowed for enough participants.

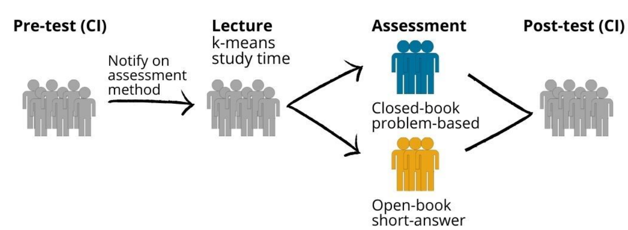


Figure 2: Experiment set-up

Since participants will be notified of the assessment method in step 2, the expectation is that students will adjust their study techniques depending on the assessment [15], which could eventually affect their learning performance. The hypothesis for this experiment is therefore as follows: between closed-book problem-based and open-book short-answer assessment, one method will result in a significantly higher learning gain of k-means clustering.

IV Responsible Research

Before the experiment was conducted, a request for approval was submitted to the Ethical Committee of TU Delft. Additionally, ethical implications of the study were taken into account. The Netherlands Code of Conduct for Research Integrity [26] defines five principles to ensure integrity of your study. Each principle will be laid out in this section, along with two common forms of data bias.

Honesty refers to accurately reporting your process and possible margins of uncertainty. This was considered throughout the whole process: no unfounded claims were made and results were not changed.

Scrupulousness highlights the importance of using scientific methods and designing the research method with care, while Independence refers to not letting the research process be affected by non-scientific considerations. These were achieved by using relevant literature to substantiate each design choice. Additionally, the choices were discussed with fellow researchers to minimize the influence of personal biases in the study.

Transparency was enforced by laying out the steps in the research process completely and focusing on their reproducibility for the audience. The reproducibility of the study was ensured by publishing the complete dataset on the TU Delft repository. Additionally, the set-up of the study was motivated and the reason for leaving out a control group was explained.

The final principle is Responsibility, which refers to conducting relevant research and acknowledging that a research does not operate in isolation. To ensure the relevance of this study, different papers on assessment methods were collected and a research gap was identified: the lack of research on students' performance in relation to assessment methods. Since in [7] it was argued that understanding this relation could improve student learning, this study can be deemed socially and scientifically relevant.

It is also important to identify biases which may affect this research and how these can be mitigated. Selection bias, occurring in non-representative samples, was minimized by gathering participants from different universities and education levels. Additionally, the majority of participants identified as female, possibly due to recruiting through personal connections. However, the ratio of men to women was equal between the participant groups.

Lastly, outlier bias was of concern, considering the frequent use of averages to convey results. To avoid this, boxplots were used to depict the range of scores, as the median is visible in this plot.

V Results

The last step of Suskie's assessment cycle [3], using assessment results to improve teaching and learning, is laid out in this section.

Table 3 shows the different scores per assessment type[1]. One participant was removed as they exhibited a pre-test score which was too high (65%). Figure 3 shows the distribution of the participants' normalized learning gain per assessment type. As the pretest scores were not normally distributed, a Mann-Whitney U test was conducted, showing no significant difference between the groups (z = -0.87, p = 0.384). The learning gains were then verified for normality using the Shapiro-Wilk test and for equal variance using Bartlett's test. To test the hypothesis that one assessment method would result in a significantly higher learning gain, a two-sample t test was used. The difference between the closed-book learning gain and the open-book learning gain was found to be significant (t(20) = -3.273; p = 0.004)

Table 3: Participants' scores per assessment method

Assessment	N	Pre-test mean	Post-test mean	Average of Gains (%)
Closed-book	10	1.3	4.4	53
Open-book	10	2	6.0	81

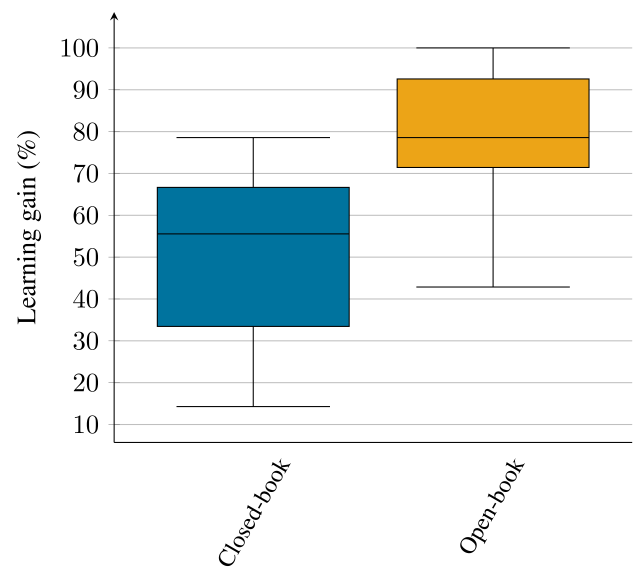


Figure 3: Learning gain in the closed-book and open-book groups, with a total of 20 participants

Additionally, participants' results per question from the pre- and post-test were documented in Figures 4 and 5. Participants were awarded 1 point for each correct answer, 0 points for a blank answer and -1 for each incorrect answer.
The most notable observations are:

- In the pre-test, only 1 out of 20 participants got question 3 correct. Out of the 10 participants who attempted to answer the question, 7 believed clustering to be a supervised learning method.
- In the post-test, 6 out of 10 participants in the closed-book group answered question 4 incorrectly. In the open-book group this was only 1 participant. This question was about convergence criteria for k-means clustering.

[1]The complete dataset can be found on the 4TU data repository: https://doi.org/10.4121/48e59b4b-de29-4007-9963-019d72c2a00e

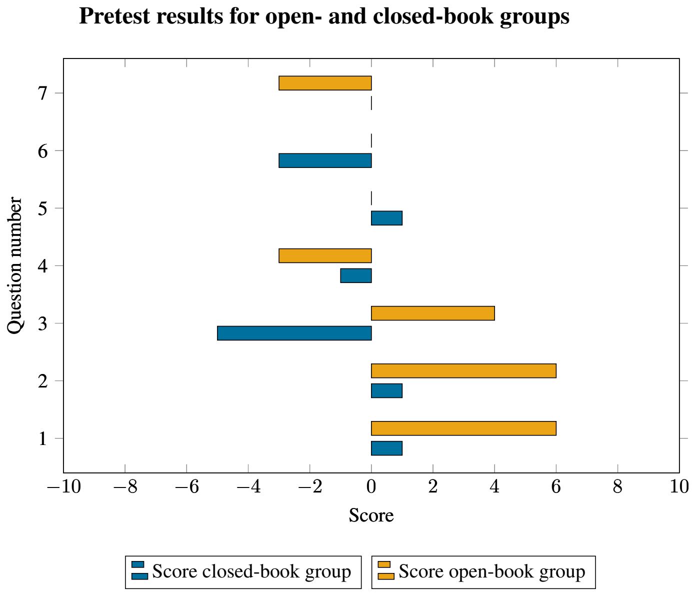


Figure 4: Participants' results per question in the pre-test

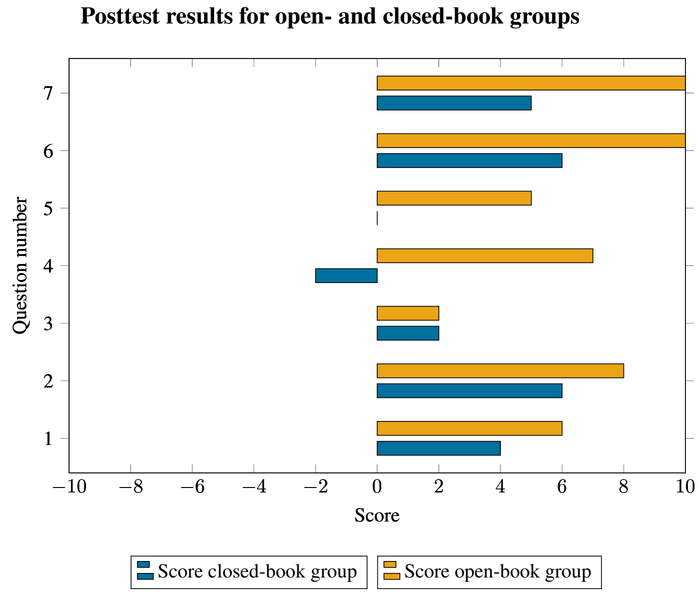


Figure 5: Participants' results per question in the post-test

- In the open-book group, question 7 was done correctly by each participant, whereas 5 out of 10 participants left this question blank in the closed-book group. The question required conducting a complete iteration of k-means by-hand.

VI Discussion

Although the results support the claim that an open-book short-answer assessment results in a higher learning gain for k-means clustering, it is important to recognise several potential limitations.

A first concern was the brief time-frame of the study. In the words of Linda Suskie [3, p. 8], "the fastest way to kill any assessment effort is to base a major decision on the results of only one assessment." To observe the long-term effects of the instruction and assessment methods used, the study would have to be conducted over a longer period of time.

The second concern was the difference between locations, as the experiment was not conducted in one sitting. The locations differed in noise, light and equipment, which led to some participants being able to concentrate better than others.

The third concern was the company in which participants completed the experiment: while most participants were in a one-on-one setting with the researcher, the experiment was also done with multiple participants at once. A notable consequence was that participants would communicate with each other, as they watched the video-lecture together. In one specific group of three participants who took the open-book assessment, none of the participants made use of the allocated study time. A logical explanation for this would be the fact that these participants were affected by each other's choices. Despite their choices, one of the participants scored a perfect score on the post-test, which showed that the lack of study time did not negatively affect their understanding.

The fourth concern relates to the participants' educational backgrounds. While the pre-test ensured no prior knowledge of k-means clustering, some participants (e.g. Aeronautical engineering, Mechanical Engineering) had more prior knowledge of mathematics than others (e.g. Industrial Design). This could have possibly lead to an advantage in the mathematical questions of the assessment.

Relating to the questions of the assessment, the fifth concern was the level of the questions. As there was no option to test the assessment for example on an expert in k-means clustering, the questions could have been too easy or too difficult. The most labour-intensive question was Question 7, involving a complete iteration of k-means (see Figure 6). This question was done correctly by all participants in the open-book group, but only by 50% of the participants in the closed-book group. A possible reason for this gap could be the fact that participants were allowed to ask the researcher questions during the allocated study time. Not all participants made use of this option, which may have led to bigger gaps in information. The most common question (n=4) related to centroids: participants did not understand the necessity of centroids, how to calculate them or which points to include. All of these questions came from participants of the closed-book group, suggesting they may have been confused, leading to a lower score on Question 7.

VII Conclusions and Future Work

In this paper, the influence of the type of assessment on students' learning gain in k-means clustering was investigated. Two different assessment methods were used: a closed-book problem-based assignment and an open-book short-answer exam. A concept inventory on k-means clustering was used as a pre- and post-test. Results show a significantly higher learning gain for participants who took the open-book assessment ($t(20) = -3.27$; $p = 0.004$). Additionally, all of the open-

Q. 7

- Given are the following points: $(1, 4)$, $(2, 2)$, $(5, 5)$ and $(4, 6)$.
- We have two centroids given: $\mu_1(1, 2)$ and $\mu_2(6, 6)$
- Conduct one iteration of the k-means algorithm (here $k = 2$).
- What are the coordinates of the new centroid μ_{1new} for the cluster which contains our original μ_1?

○ a. $\mu_{1new}(\frac{4}{3}, \frac{8}{3})$
○ b. $\mu_{1new}(5, 5\frac{2}{3})$
○ c. $\mu_{1new}(\frac{8}{3}, \frac{4}{3})$
○ d. $\mu_{1new}(4, 4\frac{1}{4})$
○

Figure 6: Question 7 of the pre-/post-test

book participants were able to conduct a full iteration of k-means clustering in the post-test, whereas only 50% of the closed-book participants managed to do this. The concept inventory (CI) was not validated however, making it difficult to draw conclusive evidence from participants' results. All of the assessments were developed from scratch, due to the lack of assessment techniques in the field of Machine Learning. This lack warrants the need for development of validated Machine Learning CIs and research on different suitable assessment techniques for Machine Learning. To observe the long-term effects of certain assessment techniques, the study would have to span an entire course.

References

[1] E. Gibney, "AI talent grab sparks excitement and concern," *Nature*, vol. 532, pp. 422–423, Apr. 2016.

[2] A. J. Ko, "We need to learn how to teach machine learning," *Bits and Behavior*, 2022.

[3] L. Suskie, "UNDERSTANDING THE NATURE AND PURPOSE OF ASSESSMENT," in *Designing Better Engineering Education Through Assessment: A Practical Resource for Faculty and Department Chairs on Using Assessment and ABET Criteria to Improve Student Learning*, pp. 3–22, 2023.

[4] G. R. Morrison, S. M. Ross, J. R. Morrison, and H. K. Kalman, *Designing effective instruction*. Hoboken, NJ: Wiley, eighth edition ed., 2019.

[5] L. Chen, "N. M. Seel, T. Lehmann, P. Blumschein, O. A. Podolskiy (2017): Instructional Design for Learning: Theoretical Foundations: Sense Publishers, Rotterdam, DOI:10.1007/978-94-6300-941-6," *Technology, Knowledge and Learning*, vol. 24, pp. 519–522, Sept. 2019.

[6] M. F. Rauber and C. Gresse Von Wangenheim, "Assessing the Learning of Machine Learning in K-12: A Ten-Year Systematic Mapping," *Informatics in Education*, May 2022.

[7] K. Islam, P. Ahmadi, and S. Yousaf, "Assessment Formats and Student Learning Performance: What is the Relation?,"

[8] C. Lu, R. Macdonald, B. Odell, V. Kokhan, C. Demmans Epp, and M. Cutumisu, "A scoping review of computational thinking assessments in higher education," *Journal of Computing in Higher Education*, vol. 34, pp. 416–461, Aug. 2022.

[9] A. Karim, S. Fawzia, and M. M. Islam, "Factors affecting deep learning of engineering students," in *Proceedings of the 26th Annual Conference of the Australasian Association for Engineering Education (AAEE2015)* (A. Oo, A. Patel, T. Hilditch, and S. Chandran, eds.), pp. 1–8, Australia: School of Engineering, Deakin University, 2015.

[10] B. Johanns, A. Dinkens, and J. Moore, "A systematic review comparing open-book and closed-book examinations: Evaluating effects on development of critical thinking skills," *Nurse Education in Practice*, vol. 27, pp. 89–94, 2017.

[11] J. Spurlin, S. A. Rajala, and J. P. Lavelle, eds., *Designing Better Engineering Education Through Assessment: A Practical Resource for Faculty and Department Chairs on Using Assessment and ABET Criteria to Improve Student Learning*. Routledge, 1st ed., 2008.

[12] B. S. Bloom, *Taxonomy of Educational Objectives: The Classification of Educational Goals*. New York, NY: Longmans, Green, 1956.

[13] B. Olds and R. Miller, *Using Formative Assessment for Program Improvement*, pp. 266–284. 06 2023.

[14] P. Armstrong, "Bloom's taxonomy," 2010.

[15] S. Jimaa, "The impact of assessment on students learning," *Procedia - Social and Behavioral Sciences*, vol. 28, pp. 718–721, 2011.

[16] J. Perez, C. Vizcarro, J. Garcia, A. Bermudez, and R. Cobos, "Development of Procedures to Assess Problem-Solving Competence in Computing Engineering," *IEEE Transactions on Education*, vol. 60, pp. 22–28, Feb. 2017.

[17] A. Espinal, C. Vieira, and A. Magana, "Professional Development in Computational Thinking: A Systematic Literature Review," *ACM Transactions on Computing Education*, vol. 24, no. 2, 2024.

[18] H. H. S. J. David Sands, Mark Parker and R. Galloway, "Using concept inventories to measure understanding," *Higher Education Pedagogies*, vol. 3, no. 1, pp. 173–182, 2018.

[19] L. P. K. W. C. L. C. Taylor, D. Zingaro and M. Clancy, "Computer science concept inventories: past and future," *Computer Science Education*, vol. 24, no. 4, pp. 253–276, 2014.

[20] H. Zhang, A. Perry, and I. Lee, "Developing and validating the artificial intelligence literacy concept inventory: an instrument to assess artificial intelligence literacy among middle school students," *International Journal of Artificial Intelligence in Education*, 2024.

[21] K. P. Sinaga and M.-S. Yang, "Unsupervised k-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020.

[22] A. Ng, "The k-means clustering algorithm." Lecture Notes, 2019. CS229: Machine Learning.

[23] E. Grimson, J. Guttag, and A. Bell, "The k-means clustering algorithm." Lecture Notes, 2016. Machine Learning.

[24] A. Elnouty, "Unsupervised learning." Lecture Notes, 2021. Machine Learning.

[25] R. R. Hake, "Evaluating conceptual gains in mechanics: A six thousand student survey of test data," in *The Changing Role of Physics Departments in Modern Universities: Proceedings of ICUPE*, vol. 399, (College Park, Maryland (USA)), pp. 595–604, AIP Conference Proceedings, AIP, March 1997.

[26] KNAW, NFU, NWO, TO2-Federatie, Vereniging Hogescholen, and VSNU, "Nederlandse gedragscode wetenschappelijke integriteit," 2018.

Appendix A

Concept Inventory on k-means clustering

Concept inventory k-means clustering

Gender: m/f/non-binary **Age:** ________ **TIME: 15 MINUTES**

!!! For each question: the last option is blank, use this if you have no idea what the answer is!

Q. 1 A machine is given a dataset of thousands of audio clips (containing sounds only) and generates a new audio clip similar to the dataset. This is an example of ...

○ a. Unsupervised learning
○ b. Supervised learning
○

Q. 2 A machine is given a dataset of images of flowers with the identification of whether the image is of a “tulip” or “not tulip” and is trained to identify when an image contains a tulip. This is an example of ...

○ a. Unsupervised learning
○ b. Supervised learning
○

Q. 3 Which of the following statements is/are **true**? **(Select all that apply)**

○ a. Conducting k-means clustering repeatedly on the same dataset will always yield the same result
○ b. Clustering makes use of unlabeled data
○ c. Clustering can be used for finding similar groups of customers
○ d. Clustering is a method in the field of supervised learning
○

Q. 4 Which of the following **does not** describe a possible convergence criterium for k-means clustering?

○ a. No or minimum decrease in the Euclidean distance between data points
○ b. No or minimum change of centroids
○ c. No or minimum reassignments of data points to different clusters
○ d. No or minimum decrease in the sum of squared errors
○

Continue on next page!

Appendix A

Concept Inventory on k-means clustering

Q. 5 Calculate the sum of squared errors for the given cluster and round your answer to two decimals.

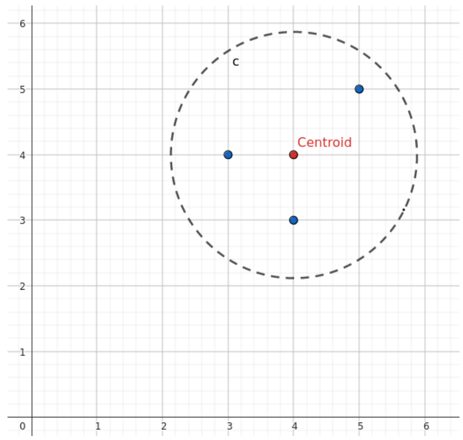


Figure 1: Cluster

- ◯ 2.00
- ◯ 3.00
- ◯ 3.41
- ◯ 4.00
- ◯

Q. 6 Which of the following **is not** a disadvantage of the k-means clustering algorithm?

- ◯ a. k-means is sensitive to initialization
- ◯ b. k-means can get stuck in a local optimum
- ◯ c. k-means can only find round-shaped clusters
- ◯ d. k-means is expensive to run
- ◯

Q. 7 Given are the following points: (1, 4), (2, 2), (5, 5) and (4, 6). We have two centroids given: $\mu_1(1, 2)$ and $\mu_2(6, 6)$ Conduct one iteration of the k-means algorithm (here k = 2). What are the coordinates of the new centroid μ_{1new} for the cluster which contains our original μ_1?

- ◯ a. $\mu_{1new}(\frac{4}{3}, \frac{8}{3})$
- ◯ b. $\mu_{1new}(5, 5\frac{2}{3})$
- ◯ c. $\mu_{1new}(\frac{8}{3}, \frac{4}{3})$
- ◯ d. $\mu_{1new}(4, 4\frac{1}{4})$
- ◯

Appendix B

Closed-book problem-based assignment on k-means clustering

Closed-book problems k-means clustering

Gender: m/f/non-binary **Age:** ______ **TIME: 15 MINUTES**

For each question: Use the boxes to show your calculation/reasoning.

- Given the set of points in Figure 1 and 2, conduct k-means on it with k = 2. Cluster 1 has initial centroid point 5 (9, 7) and cluster 2 has initial centroid point 6 (6, 8). Your stopping criterium is: no reassignments of data points to new clusters.

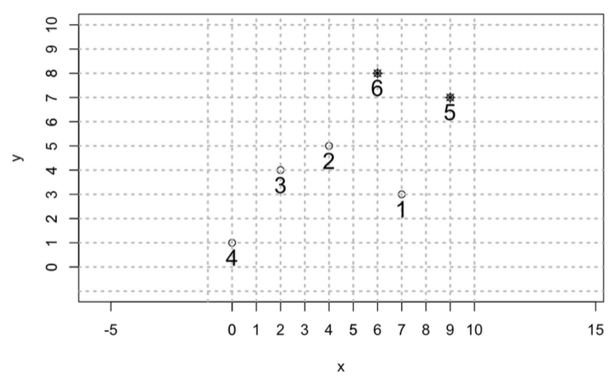


Figure 1: Set of points

point	x	y
1	7	3
2	4	5
3	2	4
4	0	1
5	9	7
6	6	8

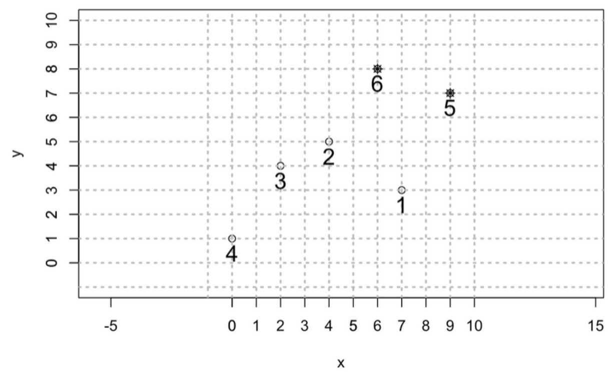
Figure 2: Coordinates

Q. 1 How many iterations are needed until the stopping criterium is met? Show your work for each iteration.

Appendix B

Closed-book problem-based assignment on k-means clustering

NB: these figures are the same as the previous page, copied for your convenience.



point	x	y
1	7	3
2	4	5
3	2	4
4	0	1
5	9	7
6	6	8

Q. 2 Which points are in the final cluster 1?

Q. 3 Which points are in the final cluster 2?

End of quiz

Appendix C

Open-book short-answer assignment on k-means clustering

K-means clustering open-book assignment

Gender: m/f/non-binary **Age:** ______ **TIME: 15 minutes**

- For each question, use the space to show your calculations or explain your answer!

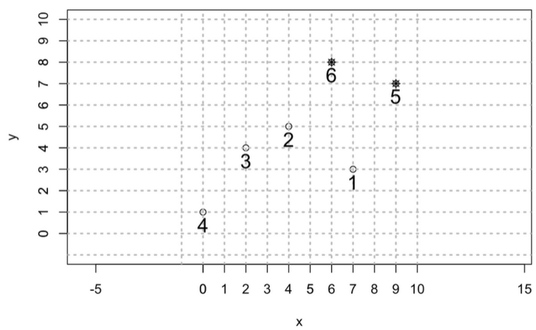


Figure 1: Set of points

point	x	y
1	7	3
2	4	5
3	2	4
4	0	1
5	9	7
6	6	8

Figure 2: Coordinates

Question 1. Iteration one of k-means algorithm, we use k=2. Use Figure 1 and 2 for this question.

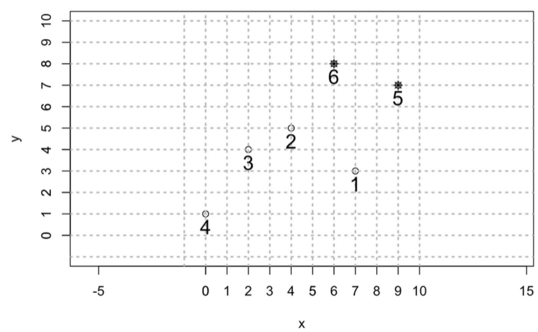
(a) Suppose we use point 5 (9, 7) as initial centroid for cluster 1 and point 6 (6, 8) as initial centroid for cluster 2. Which points will be in cluster 1?

(b) Do the same for cluster 2: which points will be in cluster 2?

Appendix C

Open-book short-answer assignment on k-means clustering

(c) Now you have the cluster memberships. Calculate the new centroids for cluster 1 and cluster 2.



point	x	y
1	7	3
2	4	5
3	2	4
4	0	1
5	9	7
6	6	8

NB: These figures are the same as before, copied for your convenience.

Question 2. Suppose we use as stopping criterium: no re-assignments of data points to new clusters. How many iterations do we need until we reach this criterium? Show your calculations.

—————————————- **End of quiz** —————————————-