

Jupyter Notebook-Based Diagnostic E-Assessment Model for Novice Programmer

¹Panyahuti, ²Ganefri, ³Ambiyar

¹Padang State University a student of the Doctoral Program, Padang State University, Indonesia, Vocational Technology Education Study Program. His research deals with computer-based assessments. Jl. Prof. Dr. Hamka, Air Tawar Barat., North Padang Sub-district, Padang City, West Sumatra

²Padang State University an Indonesian technical lecturer and academician. He is the Rector of Padang State University (UNP) for the period 2020-2024.

³Padang State University a lecturer and academician, an expert in evaluation and assessment. He teaches at the Technological and Vocational Education Doctoral Program, Padang State University

¹p.matondang@gmail.com

Article History Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract: Programming is one of the most difficult subjects for new students, and an important effort to improve these skills is by adjusting their cognitive abilities as well as learning styles. Therefore, this study aimed to improve student's learning outcomes through the use of Diagnostic E-Assessment Model. It determines the effectiveness of the model in improving student learning outcomes. Furthermore, programming teachers need to relate authentic information about unlearned topics to the students since their skills are grouped at the SOLO taxonomy level. Data grouping was conducted using the K-means clustering method which is often used in data mining analysis. It measures students' programming skills using an AMS application in line with the Jupyter Notebook engine. The data was obtained twice, after the completion of learning and after the teacher followed up on the assessment results. Also, the differences in the first and second assessments were analyzed using the N-Gains method. The results showed a significant increase in programming skills. This was indicated by the significant movement of numbers from the low to the highest level group such as Relational. Therefore, this model can also be used for other lessons as well.

Keywords: E-Assessment, Assessment, Programming, K-Means, Taxonomy.

1. Introduction

Currently, the implementation of learning assessments needs to be adapted to the evolution of learning styles and should be able to promote better learning [1]. Newly emerging e-learning platforms have developed a variety of assessment tools. Many of these have simply changed the way assessments are presented by moving questions from paper to monitor and changing evaluators from teacher to machine. Computer-based assessment is not only concerned with the process but also the security [2]. In online assessment, modern network technologies can be used to quickly transfer and analyze large amounts of data. This results into simple and efficient assessment, test distribution, and reporting [3]. This model saves time, human labor, expenses, and other costs in various aspects compared to traditional assessments. However, the ultimate goal should be to make learning better and more effective [1].

Numerous literature and study have shown that programming is one of the most difficult subjects for new students. Several previous studies stated that it is a difficult subject, especially for beginners [4]; [5]; [6]; [7] [8]. They are consistent with studies experienced considerable difficulty in teaching basic programming to vocational high school students. A study on computer learning emphasizes the need to examine students' cognitive maturity and learning styles. It was reported that student academic success in computer programming was closely related to cognitive maturity and learning styles [9].

In the process of improving performance, conducting assessments, analyzing student successes and difficulties will be the first steps. This is because they highlight key aspects of the learning process which should be developed [10] [11] [12]. They stated that less capable students should not only be assisted to overcome their mistakes but also need good and correct learning strategies.

In online learning, the assessment of computer programming skills cannot be conducted optimally by teachers. The assessment is conducted directly in the computer lab, where the teacher sees students writing program code directly. This is the best way to assess a person's programming skills [13] and it is different from their cognitive ability where assessments can be conducted by online exam tools [14].

Several models of programming skill assessment have been made, such as on the website Aspiring and [13], simulator coding test [15]. Some important questions can be answered by this simulator application, such as: do prospective programmers know how to write code? Can they compile good programming code? Are they efficient in making program code?. To recruit the best prospective programmers, they compare the quality of the code with

the time spent in writing [16]. Unfortunately, this application is devoted to testing the programming skills of prospective workforce not for students.

In education, Progtest forms the models of computer-based assessment skills. Progtest provides feedback to test participants [17] but teachers cannot use this application to make real-time assessments. To anticipate plagiarism of program code, [18] an assessment model specifically for detecting plagiarism in program code was developed. Furthermore, [17] [19] another model was developed to provide feedback to students specifically for analyzing programming skills, especially in branching codes.

This model is like a programming virtual lab, where all program code written by students is seen in real-time on the teacher's computer monitor. The student's program lines can be checked while writing the program code. The teacher can also run program code on the computer since the assessment of programming skills is based on the SOLO (Taxonomic Structure of Observed Learning Outcomes) domain. This domain proposed by Biggs [20] is used as a taxonomy of general education, learning environment, and programming [21] [22] [23]. It consists of 5 levels, namely Prestructural, Unistructural, Multistructural, Relational and Extended Abstract.

This study aimed to determine the effectiveness of the developed model in improving student learning outcomes. Programming teachers are expected to have authentic data on subject matter that fails to be mastered by students. Furthermore, teachers are also expected to obtain data on student skills in groups at the taxonomy level used.

2. Assessment of Programming Skills with Jupyter Notebook

This assessment was designed using a Jupyter Notebook and will be used for students that are beginners, especially those studying the Python or C programming language. Jupyter Notebook is a web application used to write programs with a console-based approach, interactive computing, accessible from a client computer [24]. This notebook can be accessed through a modern web browser. It makes it very practical to be used on a local computer such as a desktop application or accessed on a server that is located far from the client. When using a remote server, the only software needed is the web browser, which can easily provide access to students [25]. The notebook files created are in a simple and documented JSON format, with the extension '.ipynb' and can be easily modified for special purposes [26] [27].

Jupyter Notebook is run from an iframe [28], packaged in an application called AMS (Assessment Management System). iframe can be used to display other pages in a website or display other HTML documents [29]. The AMS developed is described as follows:

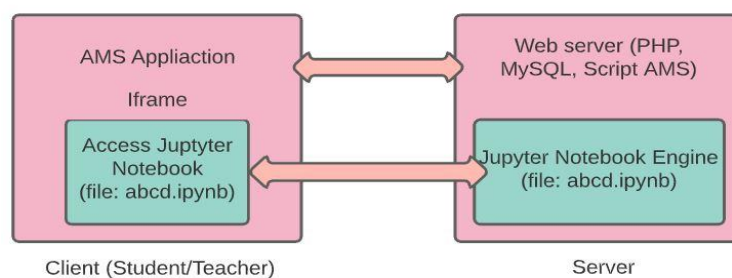


Figure 1. Programming Skill Assessment Application

The AMS Assessment Application will call up the Jupyter Notebook on an Iframe web page. Programs are written on this page by students and the file is named abcd.ipynb (Figure 1). It can be modified and executed and in a few minutes (according to the configuration), it will be stored in the MySQL database as a record and stored in a log file [30]. This is because all events during writing and running the program can be seen and assessed at other times.

To determine the skills of the study subjects, data were grouped using the K-Means method. The group-level is adjusted to the SOLO taxonomy level used in creating the questions. The K-means method was chosen because of its simple nature and commonly used for forming data clusters [31]. For numerical data with a large number of dimensions, the method is very efficient for classifying similar data into the same cluster [32].

3. Method

Diagnostic E-Assessment Model was used in implementing AMS to measure programming skills. This model is designed to align teaching and learning activities with assessment criteria [33]. The modified syntax of the model is as follows:

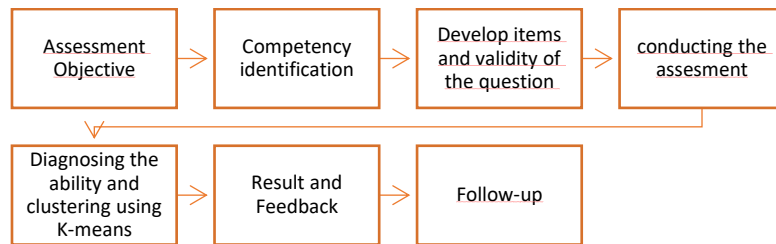


Figure 2. Diagnostic E-Assessment Model of Programming Skills

SOLO taxonomy is used as in table 1 below to develop the difficulty level of the question items used in the trial.

Table 1. Programming skill level based on SOLO taxonomy

Solo Level	Programming Algorithm Design
Prestructural (P)	Lack of basic knowledge in using program syntax and rules of writing programs
Unistructural (U)	Combines program syntax into a simple program and produces correct output
Multistructural (M)	Combines program syntax of several blocks into a single unit to produce correct output
Relational (R)	Combines and integrates two or more building blocks in more complex ways to complete a given programming task

In this study trial, 4 (four) items were developed to test the students' programming skills. The subjects were class X students (beginners) at a 3-year vocational school, the Computer and Informatics study program at the State Vocational High School 1 Lahat. The number of test subjects was 30 and the questions developed consisted of one moderate and one difficult item. The programming language tested was Python and the design of the questions to be made is described in table 2 below:

Table 2. SOLO taxonomy level in essay question writing

Taxonomy Level	Competency	Subject matter	Difficulty
Unistructural (U)	Write program code according to programming language rules and syntax. Create program code with data types, variables, constants, and expressions	Program writing rules, Data types, variables, operators, and expressions	Easy
Multi structural (M)	Create program code with arithmetic and logic operations	Arithmetic operators, create programs for numerical calculations	Moderate
Multistructural (M)	Create program code of branching	If – else, While loop, For loop	Moderate

		structure and looping structure		
Relational (R)		Create program code of branching structure and looping structure	If – Elif - Else, while loop, For loop	Difficult

The assessment will be conducted on the internet network, and each computer is connected to a server. The AMS application and Jupyter Notebook are installed on the server. The assessment participants will receive the same number of questions as the randomized items.

The results will be stored per unit of time and the teacher can conduct a step-by-step assessment of the program code written by students. The program code is given a value based on the grouping of the tested material and grouped according to the ability to answer the material. The results will be used as feedback to improve learning.

The diagnostic assessment is conducted twice when all the subject matter has been delivered and after the teacher repeats the lesson on the unfinished material. The effectiveness of the learning outcomes is tested by looking at the differences in the assessment results. This is conducted before repeating the material and after receiving feedback from the diagnostic assessment.

To obtain the difference between the first and the second assessment, the N-Gain test was used. The formula [34] is:

$$N - Gain = \frac{S2 - S1}{100 - S1} \quad (1)$$

Where S1 = average 1st assessment, S2 = average 2nd assessment.

To form data clusters, the K-means formula is used [35]:

$$d(x, y) = \sqrt{\sum_{i=1}^n (xi - yi)^2} \quad (2)$$

d is the distance between x and y, x is the cluster center data, y is data on the attribute, i is each data, n is the number of data, xi is data at the Ith cluster center, and yi is data on each ith data. In determining the new centroid point, the following formula is used:

$$Ci = \frac{\sum_{i=1}^n xi \in si}{n} \quad (3)$$

Ci is the ith centroid, si is the ith object, xi is the value of the ith object, and n is the number of data for each group.

4. Results and Discussion

The AMS application is developed for assessing programming skills in the domain of student skills. This development was conducted due to the demands of online learning in the Covid-19 pandemic, and can run on Windows, Linux, and MacOS operating systems. For the application to run properly, it is necessary to install a web server on an operating system, which should be able to run PHP scripts and MySQL server database. Furthermore, it is necessary to install the Jupyter Notebook application that is capable of running the Python kernel, preferably using Anaconda. To start the test, a service that runs the Python kernel needs to be used. The Windows 2019 server operating system with installed Anaconda is used. In addition, shortcuts to run the service is available on the desktop, such as the Shell prompt (Anaconda).

The assessment was conducted in class X (ten) at the Vocational High School 1 Lahat, South Sumatra province. All 30 students of class X Software Engineering Skills Competency were used as the test subjects. The type of test question used was essay. Furthermore, students wrote answers in the form of program code on the AMS application provided and the teacher monitored their activities through the computer screen. Initially, they log in to the application in order to choose the test provided.

The questions and answers were provided on one page. The questions were at the top while the answers were right below them. Subsequently, students wrote the program in the window provided. On this page, they were able to test, save or edit the program, which has being created. The answer page for writing program code is shown in Figure 3.

4. Question

Cup A contains coffee, cup B contains milk, cup C contains tea, and cup D contains water. The content of cup A and cup C are exchanged for each other, so do Cup B and Cup D.

Make a python program to exchange each of those cups. Cup A, B, C, D are the variables. Coffee, Milk, Tea, and Water are the data.

Use how do you exchange the content of the cups in real life as the analogy of the exchange of the variables. Print the variables.

Answer

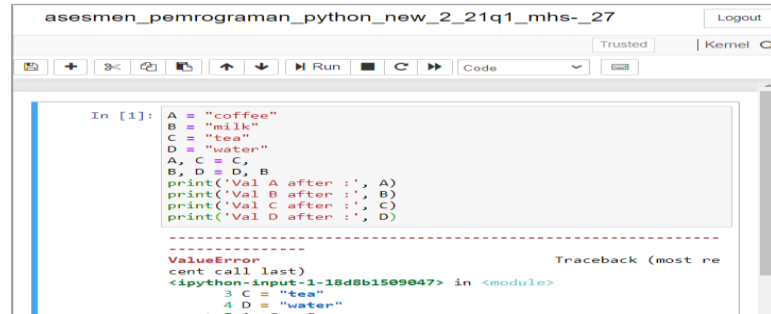


Figure 3. Student answer log files

The results were stored in the answer log file as shown in Figure 3. The AMS application created and saved the answer file per unit of time (10 minutes) and can be adjusted as needed. Furthermore, mistakes made while making the program are well recorded and documented. The documentation will serve as a reference for the teacher to provide scores. This enables them to analyze the student's arrangements in creating the program, the lines that have an error, the output, the algorithm used, and other assessments. In addition, a button is provided to aid in continuing or moving to another question. The students are allowed to move to other questions and make changes to the program code provided the allotted time is still available. However, when it is up, the AMS application will automatically close the question and answer window then log out on the user.

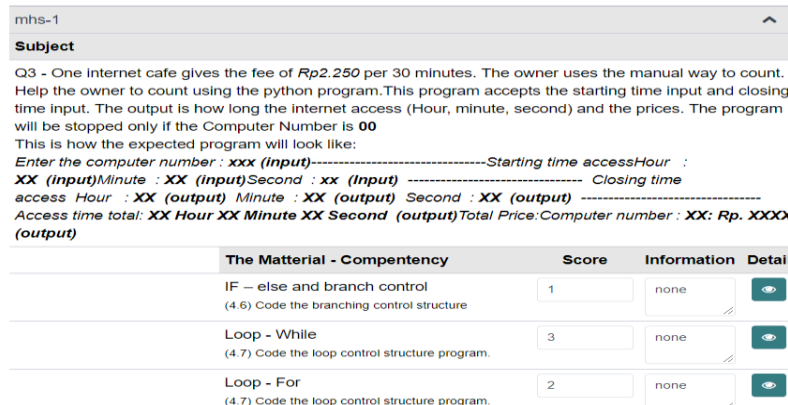


Figure 4. Teacher's scoring form

Figure 4 showed that the score for each material tested each given on a scale of 0 to 4. The number 1 is given to students with the category of "less capable", 2 for "capable enough", 3 for "capable", and 4 for "very capable". The analysis results are shown in Table 3.

Table 3 Unsuccessful Programming Skills Report
Python Programming Download: 2021-03-18 08:53:20

Num	Lesson Materials	Unsuccessful (%)
1	Program writing rules	7
2	Data types	13
3	Variables	27
4	Arithmetic operators	40
5	Translating programs for numerical calculations	47

6	IF – else and branch control	80
7	Loop – While	83
8	Loop – For	90
9	IF – else nested branch control	93
11	Access list data type	80

Table 3 is the teacher's reference for improving the learning process in the follow-up steps of the Diagnostic model. This includes teaching their weaknesses in the subject matter.

To perform the analysis using the K-means algorithm, the answer scores per question item was used as the data source. The results of the first and second assessments after learning were presented in Table 4 and 5.

Table 4. Student programming skills assessment score

Num	Q1	Q2	Q3	Q4	Num	Q1	Q2	Q3	Q4
1	2	1	1	1	16	4	2	1	4
2	1	3	1	2	17	3	2	3	3
3	3	2	2	3	18	1	4	3	3
4	3	1	3	3	19	4	2	1	3
5	4	1	1	4	20	4	1	4	1
6	3	1	3	1	21	4	1	4	1
7	3	1	3	1	22	3	1	2	4
8	3	3	1	2	23	4	4	4	1
9	4	3	1	1	24	4	2	4	4
10	2	1	1	4	25	4	4	1	4
11	4	1	4	4	26	4	4	4	2
12	4	4	1	1	27	1	2	1	1
13	1	2	3	4	28	2	1	1	2
14	4	1	3	4	29	2	1	1	2
15	2	1	4	3	30	1	1	1	2

Table 5. Student programming skills assessment score

Num	Q1	Q2	Q3	Q4	Num	Q1	Q2	Q3	Q4
1	2	1	1	1	16	4	2	1	4
2	1	3	1	2	17	3	2	3	3
3	3	2	2	3	18	1	4	3	3
4	3	1	3	3	19	4	2	1	3
5	4	1	1	4	20	4	1	4	1
6	3	1	3	1	21	4	1	4	1
7	3	1	3	1	22	3	1	2	4
8	3	3	1	2	23	4	4	4	1
9	4	3	1	1	24	4	2	4	4
10	2	1	1	4	25	4	4	1	4
11	4	1	4	4	26	4	4	4	2
12	4	4	1	1	27	1	2	1	1
13	1	2	3	4	28	2	1	1	2
14	4	1	3	4	29	2	1	1	2
15	2	1	4	3	30	1	1	1	2

The number of questions tested was Q1, Q2, Q3, and Q4 and the students' scores were between 1 and 4. To maintain the quality of the test in this second assessment, the questions used were not similar to the first. The context was different from the first, but the level of difficulty and the form of the questions were similar.

The first step in grouping data using K-means is to determine the initial centroid randomly, with the number of groups as needed [36]. This determination cannot be conducted arbitrarily. However, it is necessary to consider the suitability of the data at the desired cluster level. The number of groups made is adjusted to the skill level of the SOLO taxonomy. Student's skills were grouped into 4 (four), namely Prestructural (C1), Unistructural (C2),

Multistructural (C3), and Relational (C4). The initial centroid data is presented in table 6 below following the consideration for the suitability of the item scores against the group level:

Table 6. Initial Centroid

Level	QC1	QC2	QC3	QC4
C1	1	1	2	1
C2	1	2	2	2
C3	3	3	3	4
C4	3	4	4	4

The results of K-means analysis showed that the centroid did not change in the 5th and 6th iterations, and this is presented in table 7 below:

Table 7. The final centroid in the 5th and 6th iterations

Level	QC1	QC2	QC3	QC4
C1	2.444444	1.111111	2.111111	1.333333
C2	2.833333	3.5	1.333333	2.166667
C3	3.153846	1.461538	2.461538	3.615385
C4	4	4	4	1.5

Table 8. The final result of the K-means analysis in the 6th iteration

N um	C 1	C 2	C 3	C 4	M in	L evel	N um	C 1	C 2	C 3	C 4	M in	L evel
1	1	2	3	4	1	C	1	3	2	1	4	1	C
	.25	.9	.24	.72	.25	1	6	.4	.66	.81	.39	.81	3
2	2	1	3	4	1	C	1	2	2	0	2	0	C
	.71	.94	.43	.39	.94	2	7	.16	.4	.99	.87	.99	3
3	1	1	0	3	0	C	1	3	2	3	3	2	C
	.97	.85	.95	.35	.95	3	8	.74	.66	.43	.5	.66	2
4	1	3	0	3	0	C	1	2	2	1	3	1	C
	.97	.12	.95	.64	.95	3	9	.69	.1	.88	.91	.88	3
5	3	3	1	4	1	C	2	2	4	3	3	2	C
	.28	.33	.79	.92	.79	3	0	.47	.01	.18	.04	.47	1
6	1	3	2	3	1	C	2	2	4	3	3	2	C
	.11	.23	.71	.35	.11	1	1	.47	.01	.18	.04	.47	1
7	1	3	2	3	1	C	2	2	3	0	4	0	C
	.11	.23	.71	.35	.11	1	2	.73	.18	.77	.5	.77	3
8	2	0	2	3	0	C	2	3	3	4	0	0	C
	.36	.65	.67	.35	.65	2	3	.8	.18	.05	.5	.5	4
9	2	1	3	3	1	C	2	3	3	1	3	1	C
	.71	.76	.47	.2	.76	2	4	.73	.75	.88	.2	.88	3
10	2	3	1	5	1	C	2	4	2	3	3	2	C
	.92	.23	.96	.32	.96	3	5	.37	.25	.07	.91	.25	2
11	3	4	1	3	1	C	2	3	2	3	0	0	C
	.62	.25	.86	.91	.86	3	6	.84	.96	.48	.5	.5	4
12	3	1	4	3	1	C	2	2	2	3	4	2	C
	.48	.76	.02	.04	.76	2	7	.05	.66	.73	.72	.05	1
13	3	3	2	4	2	C	2	1	2	2	4	1	C
	.28	.43	.32	.5	.32	3	8	.37	.66	.51	.72	.37	1
14	3	3	1	4	1	C	2	1	2	2	4	1	C
	.21	.71	.17	.03	.17	3	9	.37	.66	.51	.72	.37	1
15	2	3	2	3	2	C	3	1	3	3	5	1	C
	.56	.84	.07	.91	.07	3	0	.94	.12	.1	.22	.94	1

The value between each group is calculated to determine groups C1, C2, C3, or C4 in table 7. The result with the smallest value is the criteria for being a group, and in table 7 on row 1, C1 = 1.25, C2 = 2.9, C3 = 3.24, C4 = 4.72. Therefore, the smallest value is 1.25 in column C1, and it represents the level..

Table 8 showed that the number of levels calculated were Prestructural (C1) 9, Unistructural (C2) 6, Multistructural (C3) 13, and Relational (C4) 2. This data is presented in Figure 7.

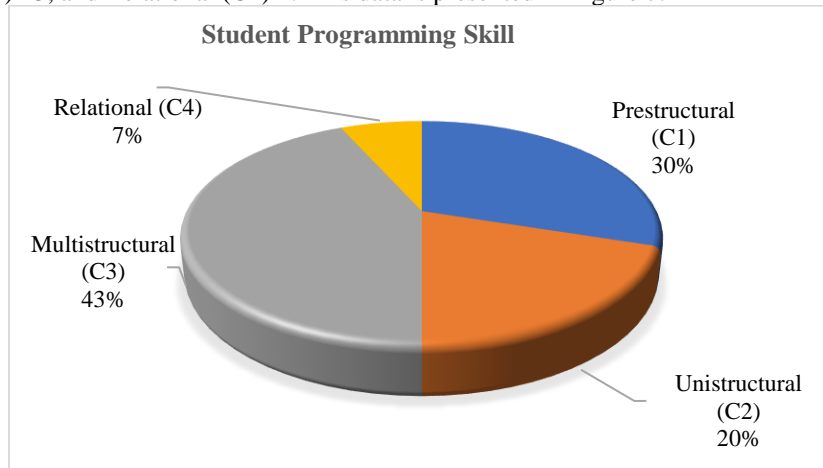


Figure 8. Student Programming Skill Assessment-1

With the same steps, the analysis was conducted on the second assessment data. Initial centroid used data in the first assessment analysis represented in table 6. The final results of this second analysis are described in Figure 9 as follows:

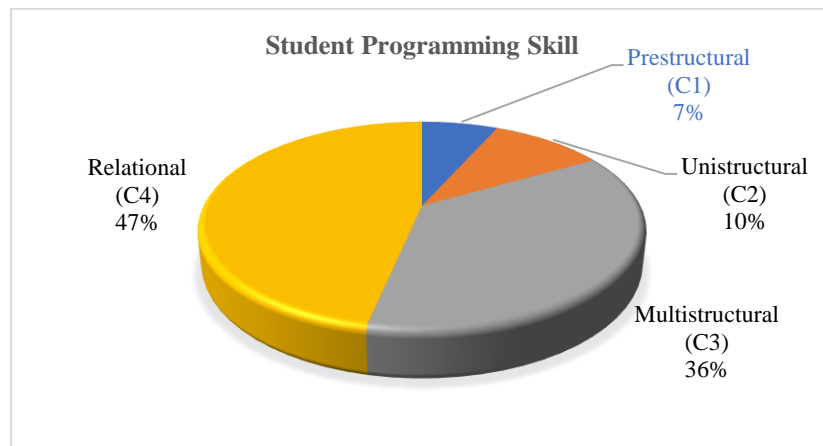


Figure 9. Student Programming Skill Assessment-2

The effectiveness of the assessment model for improving student learning outcomes is calculated using N-Gains as presented in table 8 below:

Table 9. N-Gain of each student's skill level

Level	Assessment-1 (%)	Assessment-2 (%)	N-Gains
Prestructural (C1)	30	7	-0.247311828
Unistructural (C2)	20	10	-0.111111111
Multistructural (C3)	43	36	-0.109375
Relational (C4)	7	47	0.754716981

A negative number was obtained at levels C1, C2, and C3. Therefore, there is a decrease in the number of students at that level. This is consistent with the standard, where the student group switches to a higher level

such as Relational (C4). There was a very good increase of 0.754 in the N-Gain number at the Relational level. When compared to the above N-Gain, 0.7 is in the high category [34].

The graphic in Figure 8 showed that programming lessons are difficult for students. This is because it requires guidance, direction, and experience from a teacher. Furthermore, it is a cognitive activity that requires reasoning [37], understanding, and technical skills [38]. Programming skills are very useful to learn, even though the lessons are quite difficult. Skills are one of the factors that directly affect student performance. It is the ability to use individual knowledge effectively and readily in completing programming tasks. Similarly, student's abilities can be determined through motivation and knowledge [39], [40].

5. Conclusion

From the analysis, it is reasonable to conclude that the implementation of the diagnostic E-Assessment model is effective in improving the learning outcomes of computer programmers. The results of student failures in the material tested are very useful for improving the learning process in the follow-up steps. Furthermore, the difficulty of programming lessons can be overcome by teaching the weaknesses of students.

This model can also be used in other subjects, but certainly with a different assessment application. An assessment application that can be used for the realm of knowledge in all subjects have also been built.

References

- A. H. Su, 'Educational Assessment of the Post-Pandemic Age: Chinese Experiences and Trends Based on Large-Scale Online Learning', *Educ. Meas. Issues Pract.*, vol. 39, no. 3, pp. 37–40, 2020.
- B. W. Camara, 'Never Let a Crisis Go to Waste: Large-Scale Assessment and the Response to COVID-19', *Educ. Meas. Issues Pract.*, vol. 39, no. 3, pp. 10–18, 2020.
- C. P. Panyahuti, G. Ganefri, A. Ambiyar, and K. Suryani, 'Safe Exam Browser For Android Clients On Web Based Exams', *Edukasi J. Pendidik.*, vol. 17, no. 2, p. 212, 2019.
- D. M. De Raadt, 'A Review of Australasian Investigations into Problem Solving and the Novice Programmer A Review of Australasian Investigations into Problem Solving and the Novice Programmer', no. December 2014, pp. 37–41, 2008.
- E. K. M. Y. Law, V. C. S. Lee, and Y. T. Yu, 'Learning motivation in e-learning facilitated computer programming courses', *Comput. Educ.*, vol. 55, no. 1, pp. 218–228, 2010.
- F. A. Joseph, S. Hunsinger, and P. Wu, 'Information Systems Applied Research 2015 AITP Education Special Interest Group (EDSIG) Board of Directors', *J. Inf. Syst. Appl. Res.*, vol. 8, no. 2, pp. 1–59, 2015.
- G. X. M. Wang and G. J. Hwang, 'A problem posing-based practicing strategy for facilitating students' computer programming skills in the team-based learning mode', *Educ. Technol. Res. Dev.*, vol. 65, no. 6, pp. 1655–1671, 2017.
- H. Z. Güney, 'Four-Component Instructional Design (4C/ID) Model Approach for Teaching Programming Skills', *Int. J. Progress. Educ.*, vol. 15, no. 4, pp. 142–156, 2019.
- I. G. L. White and M. P. Sivitanides, 'A Theory of the Relationships between Cognitive Requirements of Computer Programming Languages and Programmers' Cognitive Characteristics', *J. Inf. Syst. Educ.*, vol. 13, no. 1, pp. 59–66, 2009.
- J. A. Wijaya, H. Retnawati, W. Setyaningrum, K. Aoyama, and Sugiman, 'Diagnosing students' learning difficulties in the eyes of Indonesian mathematics teachers', *J. Math. Educ.*, vol. 10, no. 3, pp. 357–364, 2019.
- K. Ambiyar and Panyahuti, *Computer-Based Learning Assessment and Android*, 1st ed. Jakarta: Kencana, 2020.
- L. D. Tall and M. R. Razali, 'Diagnosing students' difficulties in learning mathematics', *Int. J. Math. Educ. Sci. Technol.*, vol. 24, no. 2, pp. 209–222, 1993.
- M. Aspiring-Minds Team, 'Assessing Programming Skills of Coders Validly and Scientifically', 2018. [Online]. Available: <https://www.aspiringminds.com/blog/research-articles/assessing-programming-skills-of-coders-validly-and-scientifically-automated-simulated-programming-assessment/>. [Accessed: 25-Oct-2020].
- N. S. Gupta and A. Gupta, 'E-Assessment Tools for Programming Languages: A Review', *Proc. First Int. Conf. Inf. Technol. Knowl. Manag.*, vol. 14, pp. 65–70, 2018.
- O. S. Mishra, 'Easily Conduct Coding Assessment Tailored To Your Requirements', 2020. [Online]. Available: https://mettl.com/en-id/coding-signup/?utm_medium=website&utm_source=www.google.com. [Accessed: 25-Oct-2020].

- P. Glider, 'Online Coding Tests For Your Tech Recruitment Strategy To Build An Effective Team', 2020. [Online]. Available: <https://glider.ai/articles/online-coding-tests-for-your-tech-recruitment-strategy-to-build-an-effective-team/>. [Accessed: 25-Oct-2020].
- Q. D. M. De Souza, S. Isotani, and E. F. Barbosa, 'Teaching novice programmers using ProgTest', *Int. J. Knowl. Learn.*, vol. 10, no. 1, pp. 60–77, 2015.
- R. M. N. Ngo, 'Eliminating Plagiarism in Programming Courses through Assessment Design', *Int. J. Inf. Educ. Technol.*, vol. 6, no. 11, pp. 873–879, 2016.
- S. S. Grover, '" Systems of Assessments " for Deeper Learning of Computational Thinking in K-12', *Annu. Meet. Am. Educ. Res. Assoc.*, no. 650, pp. 15–20, 2015.
- T. J. B. Biggs and K. F. Collis, 'The Psychological Structure of Creative Writing', *Aust. J. Educ.*, vol. 26, no. 1, pp. 59–70, 1982.
- U. R. Lister, B. Simon, E. Thompson, J. L. Whalley, and C. Prasad, 'Not seeing the forest for the trees: Novice programmers and the SOLO taxonomy', *Work. Gr. Reports ITiCSE Innov. Technol. Comput. Sci. Educ. 2006*, pp. 118–122, 2006.
- V. C. Izu, A. Weerasinghe, and C. Pope, 'A study of code design skills in novice programmers using the SOLO taxonomy', *ICER 2016 - Proc. 2016 ACM Conf. Int. Comput. Educ. Res.*, pp. 251–259, 2016.
- W. F. E. V. Castro and K. Fisler, 'Designing a multi-faceted SOLO taxonomy to track program design skills through an entire course', *ACM Int. Conf. Proceeding Ser.*, pp. 10–19, 2017.
- X. J. Jupyter Team, 'The Jupyter Notebook Introduction', *Jupyter Team*, 2015. [Online]. Available: <https://jupyter-notebook.readthedocs.io/en/stable/notebook.html>.
- Y. B. M. Randles, I. V. Pasquetto, M. S. Golshan, and C. L. Borgman, 'Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study', *Proc. ACM/IEEE Jt. Conf. Digit. Libr.*, 2017.
- Z. T. Kluyver *et al.*, 'Jupyter Notebooks—a publishing format for reproducible computational workflows', *Position. Power Acad. Publ. Play. Agents Agendas - Proc. 20th Int. Conf. Electron. Publ. ELPUB 2016*, pp. 87–90, 2016.
- AA. F. Pérez and B. E. Granger, 'Project Jupyter : Computational Narratives as the Engine of Collaborative Data Science', *Retrieved Sept.*, no. April, pp. 1–24, 2015.
- BB. A. G. Archer, N. York, and M. A. Pawlowski, '(12) United States Patent (10) Patent No .':, vol. 2, no. 12, 2013.
- CC. A. R. Alhafidz, 'Examples of Use and Iframe Functions', 2015. [Online]. Available: <https://www.dumetschool.com/blog/Contoh-Penggunaan-dan-Fungsi-iframe>. [Accessed: 02-Nov-2020].
- DD. B. Gavin, 'What is a Log File?', 2018. [Online]. Available: <https://www.howtogeek.com/359463/what-is-a-log-file/>. [Accessed: 03-Nov-2020].
- EE. K. P. Sinaga and M. S. Yang, 'Unsupervised K-means clustering algorithm', *IEEE Access*, vol. 8, pp. 80716–80727, 2020.
- FF. S. S. Yu, S. W. Chu, C. M. Wang, Y. K. Chan, and T. C. Chang, 'Two improved k-means algorithms', *Appl. Soft Comput. J.*, vol. 68, pp. 747–755, 2018.
- GG. A. E. Guerrero-Roldán and I. Noguera, 'A model for aligning assessment with competences and learning activities in online courses', *Internet High. Educ.*, vol. 38, no. April, pp. 36–46, 2018.
- HH. R. Hake, *Analyzing Change/Gain Scores*. Dept. Of Physics, Indiana University, 1999.
- II. C. Yuan and H. Yang, 'Research on K-Value Selection Method of K-Means Clustering Algorithm', *J.*, vol. 2, no. 2, pp. 226–235, 2019.
- JJ. A. Prasatya, R. R. A. Siregar, and R. Arianto, 'Application of the K-Means and C4.5 Method for the Prediction of Diabetes Patients', *Petir*, vol. 13, no. 1, pp. 86–100, 2020.
- KK. R. J. Barker and E. A. Unger, 'A predictor for success in an introductory programming class based upon abstract reasoning development', *ACM SIGCSE Bull.*, vol. 15, no. 1, pp. 154–158, 1983.
- LL. Z. Zhou, J. Li, C. Deng, and Y. Li, 'Construction and Practice of Competency-Oriented Programming Practice Course Group', no. Isetbm, pp. 39–44, 2019.
- MM. G. R. Bergersen, D. I. K. Sjøberg, and T. Dyba, 'Construction and validation of an instrument for measuring programming skill', *IEEE Trans. Softw. Eng.*, vol. 40, no. 12, pp. 1163–1184, 2014.
- NN. E. Vrachnos and A. Jimoyiannis, 'Secondary Education Students' Difficulties in Algorithmic Problems with Arrays: An Analysis Using the SOLO Taxonomy.', *Themes Sci. Technol. Educ.*, vol. 10, no. 1, pp. 31–52, 2017.