Validation of the Case Method Performance Assessment Evaluation System in Presenting Learning Outcome Products

Satria Mihardi[1*], Derlina[2], Wawan Bunawan[3], Alkhafi Maas Siregar[4], Tuti Hardianti[5], Muhammad Khairi Dahlan[6], Puji Nawawi[7], Nurlita[8], Andreas Bintang[9], Agnes Viola Manik[10]

{mihardi@unimed.ac.id[1], derlina@unimed.ac.id[2], wawanbunawan@unimed.ac.id[3], alkhafi@unimed.ac.id[4], tuti.hardianti@unimed.ac.id[5], khairidahlan23@gmail.com[6], pujinawawi@mhs.unimed.ac.id[7], litaa.4232421010@mhs.unimed.ac.id[8], bintangandreas001@mhs.unimed.ac.id[9], agnesviolamanik.4233121012@mhs.unimed.ac.id[10]}

Physics Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Medan, Medan, Indonesia[1,2,3,4,5], Physics Education Study Program-Physics Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Medan, Medan, Indonesia[6,7,8,9,10]

Abstract. Outcome-Based Education (OBE) aims to foster a long-lasting learning process by employing innovative, interactive, and efficient techniques. OBE transforms all elements of the educational system, from curriculum development and setting learning goals to crafting teaching methods, designing assessment processes, and shaping the educational environment. The case method means students carry out scientific research or social research. Then, students conclude. The case method is the result of reporting descriptive information regarding a trial or experiment, event or analysis, or project. The case method can also be interpreted as the result of intensive and systematic investigations carried out by individuals, groups communities, or units to examine data that is related to several variables. The research was aimed at validating the evaluation system used in assessing performance at the Case method stage instructed to students. The method is based on the Cohen Kappa vs Pearson's Correlation validation test. The testing aims to compare the two validations in describing the feasibility of performance assessment. Statistical testing using IBM Software. The results show a very small level of comparison in the Case Method stage indicators. The comparisons shown interpret the suitability of the validation tests. Feasibility of high reliability makes it a good level of suitability to be applied and used as a basis for learning assessment with the Case Method.

Keywords: Assessment; Case Method; Evaluation; Performance; Validation

ICIESC 2024, September 17, Medan, Indonesia

DOI 10.4108/eai.17-9-2024.2353066

1 Introduction

The OBE curriculum is an educational approach that emphasizes expected learning outcomes or goals. In the context of OBE, clear and measurable learning objectives are the main focus, and the learning process is designed to achieve these goals[1–3]. OBE provides a framework that allows educational institutions to measure student achievement more accurately, increase accountability, and ensure that graduates are ready to face real-world demands Outcome-Based Education (OBE) focuses on creating a sustainable learning process through innovative, interactive, and effective methods. OBE shapes every aspect of education, including curriculum design, setting learning objectives and outcomes, developing educational strategies, designing learning methods, assessing procedures, and crafting the educational environment.[4–8]. Education is a key element in community development in improving the quality of human resources. In an effort to improve the quality of education, curriculum design has a very important role. The curriculum is a design that includes planning regarding targets, content, and learning materials as a guide in implementing learning activities to achieve specific educational goals. The development of the curriculum is influenced by various factors, such as student needs, norms, values of the education system, community demands, direction of education programs, and developments in science and technology. One approach to designing a curriculum that has received widespread attention is Outcome-Based Education (OBE)[8–11].

Achievement of university-level learning outcomes is an important factor as an index of OBE achievement in the college curriculum. Achievement through assignment activities in the form of case methods and team-based projects. A case is a narrative that contains information and requires analysis. Participants are required to make decisions or assessments using the given information. Cases entail scenarios where choices need to be made, problems addressed, or policies and practices evaluated or re-examined. Effective cases generally stem from actual events but can span present circumstances, historical events, or even distant past occurrences. Essentially, a case is a narrative, scenario, selected dataset, or statement that poses an unresolved and thought-provocative problem, situation, or question. Cases include information but do not provide analysis[12–14]. Cases contain complex and unstructured problems that may include relevant or irrelevant information and often do not include all the information an analyst wants. Cases contain many contextual ways to introduce new material and create opportunities for application to relevant case resolution. The case method is the result of reporting descriptive information about an experiment or experiment, event analysis, or project. The case method can also be interpreted as the result of intensive and systematic investigations carried out by individuals, community groups, or units to study data related to several variables.

Factor validity is measured when an item is constructed using multiple factors (there is the similarity between one factor and another). The evaluate the validity of this factor, you need to compare the factor score (the number of items in the factor) with the total factor score (the sum of all factors). The effectiveness of an item is shown by how well it correlates with or contributes to the overall score, which is calculated by comparing the item score with the total item score. The combine using multiple elements, it means that the validity of the item is tested by correlating the item score with the element score[15–17]. The selectivity of the method is determined by comparing the results of the analysis of samples containing contaminants, degradation products, similar compounds, other foreign substances, or placebo carriers with the

results of the analysis of samples without the addition of these substances. The difference in results if any is the difference between the results of the two tests.

2 Methods

Analysis testing is used by comparing the results of the two testing systems which are inter-rater vs construct validity. This analysis uses Cohen Kappa and Pearson's Correlation validation test. Both analyses describe the profile of the proposed instrument development through item analysis. Information on the level of feasibility of using the instrument against the rater's assessment is the reason for developing the instrument used as a learning evaluation[18]. Testing is done using IBM Software.

Table 1. Interpretation Agreement

Value of K	Strength of Agreement
< 0.20	Poor
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Good
0.81 - 1.00	Very Good

3 Results and Discussions

Instrument development through expert testing stages with Cohen Kappa analysis. The results are shown in Table 2 with an assessment of 2 (two) Raters. The development of the instrument used for assessment consists of 20 items. The resulting feasibility shows the amount of perceptual conformity of the two raters to the Instrument. The comparison of Rater perceptions cannot be used as a standard basis for stating the feasibility of the instrument[18–20]. The development stage is an evaluation action of assessments from various parties and methods in testing the feasibility of the instrument. The results of the analysis presented are an important key as information that is worthy of being accounted for as valid data.

Table 2. Crosstabulation Raters

		Rater2		
		Invalid	Valid	Total
Rater1	Invalid	6	3	9
	Valid	5	6	11
Total		11	9	20

Table 3 displays a Cohen's kappa value of 0.208, which indicates the degree of agreement between the two raters, beyond what could occur by chance. In this case, p-value greater than 0.001, this kappa value is statistically significant from zero, demonstrating a significant level of agreement. This fair level of agreement is highlighted by the kappa value of 0.208, resulting from the comparison of the raters' assessments. This value of kappa is significantly similar to zero (κ=0.208, p>0.001). The suitability between the two raters in providing an assessment of the feasibility of the resulting instrument is the basis for continuing to the testing stage[21–23].

This testing stage was carried out with a sample of 27 people. The test uses 20 items that must be completed with stages by the case method that has been trained. Training on the stages of the case method to reduce the possibility of assessment points from each item. Table 4 shows the results of the validation test from testing items on 27 sample people.

Table 3. Symmetric Kappa

Symmetric Measures					
		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	.208	.214	.949	.343
N of Valid Cases		20			

a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.

Table 4. Validity Test

		No.2	No.5	No.8	No.9	No.10	No.14	No.15	No.16
Total	Pearson Correlation	.393*	.469*	.402*	.410*	.634**	.403*	.467*	.410*
	Sig. (2-tailed)	.043	.014	.037	.034	.000	.037	.014	.034
	N	27	27	27	27	27	27	27	27

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

The validation results show that only 8 items are considered valid, while the other 12 are invalid. Validity tests on the 8 valid items indicate a moderate level of reliability (Table 5). The level of reliability makes the assessment of instrument development feasible by considering the formation of instrument types based on the development of the 8-item criteria. This criterion shows the alignment between items that have almost the same level of reliability as seen in Table 6.

Table 5. Reliability Test

Cronbach's Alpha	N of Items
.592	8

Table 6. Corrected Reliability

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Item2	230.74	2622.507	.233	.580
Item5	225.56	2594.872	.233	.581
Item8	238.89	2817.949	.261	.569
Item9	231.85	2484.900	.431	.517
Item10	239.63	2480.627	.587	.489
Item14	221.48	2574.644	.218	.589
Item15	241.85	2484.900	.377	.532
Item16	226.30	2901.140	.131	.602

Comparison of test results from Table 2 and Table 4 shows the quantity of feasibility and suitability of the assessment rates vs instrument testing. Results that are by levels that are not too far from the raters' predictions indicate that the results of item development are running according to the planned development[16, 19–21]. The resulting instrument can show and describe the conditions that are targeted for instrument development. Further development makes the instrument a standard by considering the criteria of valid items[24–27]. Further testing requires a larger number of samples with various variations without reducing the criteria for valid tests. Instruments are very suitable as standards for testing the ability to train the stages of the case method that students have in developing abilities[10, 17, 28, 29].

4 Conclusion

The results show a very small level of comparison in the Case Method stage indicators. The comparisons shown interpret the suitability of the validation tests. Feasibility of high reliability makes it a good level of suitability to be applied and used as a basis for learning assessment with the Case Method. Further development was implemented by comparing a wider number of samples and varying the number of items to test the feasibility of the effectiveness of the resulting instrument.

References

[1] Qiao, Y., Fu, H.: Study on OBE Teaching Concept in the Context of Deep Learning for the Construction of University Mathematics Microcourses. Comput. Intell. Neurosci. (2022). https://doi.org/10.1155/2022/6860842.

[2]. Rahayu, N., Suharti, D.S., Wigati, F.A., Taufanawati, E.: Investigating The Components Of Outcome Based Education In Efl Classroom: A Lesson Plan Analysis. English Rev. J. English Educ. (2021). https://doi.org/10.25134/erjee.v9i2.4419.

[3]. Paul Leong, D.C.: Outcome-Based Education in Open Distance Learning: A Study on Its Implementation Amidst the Pandemic. Malaysian J. Soc. Sci. Humanit. (2022). https://doi.org/10.47405/mjssh.v7i9.1747.

[4]. Khan, M.S.H., Salele, N., Hasan, M., Abdou, B.O.: Factors affecting student readiness towards OBE implementation in engineering education: Evidence from a developing country. Heliyon. (2023). https://doi.org/10.1016/j.heliyon.2023.e20905.

[5]. Zamir, M.Z., Abid, M.I., Fazal, M.R., Qazi, M.A.A.R., Kamran, M.: Switching to Outcome-Based Education (OBE) System, a Paradigm Shift in Engineering Education. IEEE Trans. Educ. (2022). https://doi.org/10.1109/TE.2022.3169184.

[6]. Yang, F., Fan, J.: Construction of OBE Concept Autonomous Learning Mode in University Teaching Based on the Internet. J. Cases Inf. Technol. (2022). https://doi.org/10.4018/JCIT.295250.

[7]. Yong, W., Kharabsheh, R., Bester, F.: Quality model construction and continuous improvement mechanism design in OBE education. Appl. Math. Nonlinear Sci. (2021). https://doi.org/10.2478/amns.2021.2.00102.

[8]. Hapinat, H.L.: Practices on the outcomes-based education (OBE) implementation in select HEI

graduate school programs in the Philippines as input to institutionalizing mandatory accreditation. Int. J. Adv. Appl. Sci. (2023). https://doi.org/10.21833/ijaas.2023.03.021.
[9]. Lumius, L.D., Asli, M.F.: OBEInsights: Visual Analytics Design for Predictive OBE Knowledge Generation. Int. J. Adv. Comput. Sci. Appl. (2021). https://doi.org/10.14569/IJACSA.2021.01212108.
[10]. Rahate, V., Metre, S., Ambad, R., Bhirange, S.: Impact of outcome based education (OBE) on teaching effectiveness of faculty members of professional program. Indian J. Forensic Med. Toxicol. (2020). https://doi.org/10.37506/ijfmt.v14i4.12731.
[11]. Muzakir, M.I., Susanto: Implementasi Kurikulum Outcome Based Education (Obe) Dalam Sistem Pendidikan Tinggi Di Era Revolusi Industri 4.0. Edukasiana J. Islam. Educ. (2023). https://doi.org/10.61159/edukasiana.v2i1.86.
[12]. Chumak, M., Nekrasov, S., Hrychanyk, N., Prylypko, V., Mykhalchuk, V.: Applying Case Method in the Training of Future Specialists. J. Curric. Teach. (2022). https://doi.org/10.5430/jct.v11n1p235.
[13]. Gómez García, L.D., Alba Cabañas, M.: Teaching with the case method: opportunities and problems since the COVID-19 pivot to online. Account. Res. J. (2022). https://doi.org/10.1108/ARJ-09-2020-0298.
[14]. Huang, C.H., Hsiao, L.H.C., Ko, S.L.: Effect of applying case method to anti-corruption education on learning motivation and learning effectiveness. Rev. Cercet. si Interv. Soc. (2021). https://doi.org/10.33788/rcis.73.17.
[15]. SÜRÜCÜ, L., MASLAKÇI, A.: VALIDITY AND RELIABILITY IN QUANTITATIVE RESEARCH. Bus. Manag. Stud. An Int. J. (2020). https://doi.org/10.15295/bmij.v8i3.1540.
[16]. Almanasreh, E., Moles, R., Chen, T.F.: Evaluation of methods used for estimating content validity, (2019). https://doi.org/10.1016/j.sapharm.2018.03.066.
[17]. Mohajan, H.K.: TWO CRITERIA FOR GOOD MEASUREMENTS IN RESEARCH: VALIDITY AND RELIABILITY. Ann. Spiru Haret Univ. Econ. Ser. (2017). https://doi.org/10.26458/1746.
[18]. McHugh, M.L.: Interrater reliability: The kappa statistic. Biochem. Medica. (2012). https://doi.org/10.11613/bm.2012.031.
[19]. Foody, G.M.: Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. Remote Sens. Environ. (2020). https://doi.org/10.1016/j.rse.2019.111630.
[20]. O'Shea, D., Nathan, R., Sharma, A., Wasko, C.: Improved Extreme Rainfall Frequency Analysis Using a Two-Step Kappa Approach. Water Resour. Res. (2023). https://doi.org/10.1029/2021WR031854.
[21]. Yanto, D.T.P., Hastuti, Zaswita, H., Kabatiah, M., Sukardi, Ambiyar: Validity Test Analysis of Virtual Laboratory-Based Job Sheet for Power Electronics Course. Int. J. Inf. Educ. Technol. (2023). https://doi.org/10.18178/ijiet.2023.13.9.1951.
[22]. Di Vico, I.A., Stone, J., Mcwhirter, L., Riello, M., Zanolin, M.E., Colombari, M., Fiorio, M., Tinazzi, M.: Performance validity tests in nonlitigant patients with functional motor disorder. Eur. J. Neurol. (2023). https://doi.org/10.1111/ene.15703.
[23]. Sukadarin, E.H., Md Deros, B., Ghani, J.A., Ismail, A.R., Mohd Nawi, N.S., Abdull, N.: Validity test for Simple Ergonomics Risk Assessment (SERA) method. Malaysian J. Public Heal. Med. (2016).
[24]. Strelow, E.L., Gerisch, A., Lang, J., Pfetsch, M.E.: Physics informed neural networks: A case study for gas transport problems. J. Comput. Phys. (2023). https://doi.org/10.1016/j.jcp.2023.112041.
[25]. Mizukoshi, K.: Case Study Research and Applications. Japan Mark. J. (2023). https://doi.org/10.7222/marketing.2023.045.

[26]. Alemu, M.: Improving secondary school students physics achievement using reciprocal peer tutoring: A multi-level quasi-experimental study. Eurasia J. Math. Sci. Technol. Educ. (2020). https://doi.org/10.29333/ejmste/115164.
[27]. Japkowicz, N.: Performance Evaluation. In: Machine and Deep Learning in Oncology, Medical Physics and Radiology, Second Edition (2022). https://doi.org/10.1007/978-3-030-83047-2_6.
[28]. Aulia, R., Michael Johan Sulistiawan, Wagiran, W., Deby Luriawati Naryatmojo: Rekontruksi Asesmen Formatif Modul Ajar Teks Deskripsi Bagi Siswa SMP di Era Merdeka Belajar. J. Onoma Pendidikan, Bahasa, dan Sastra. (2024). https://doi.org/10.30605/onoma.v10i1.3188.
[29]. Schmalz, U., Spinler, S., Ringbeck, J.: Lessons Learned from a Two-Round Delphi-based Scenario Study. MethodsX. (2021). https://doi.org/10.1016/j.mex.2020.101179.