The Content Validity of Digital Test Items for Evaluation Courses Based on Superitem-Wondershare Using Aiken’s Calculation

D G H Divayana¹, A Adiarta², I G Sudirtha³

¹Department of Informatics Education, Universitas Pendidikan Ganesha
²Department of Electrical Education, Universitas Pendidikan Ganesha
³Department of Vocational Education, Universitas Pendidikan Ganesha

e-mail: hendra.divayana@undiksha.ac.id

Abstract. The digital test item for evaluation courses that is designed with tiered arrangements use super item concept and is showed digitally use the Wondershare application makes it easy for students to practice the quality of their critical thinking in working on test questions whenever and wherever they are. But to get the good quality test items, then it is still necessary to test the item validity properly and accurately. The method that can be done to determine the valid test items is through calculating the validity of the items uses the Aiken model. Based on those matters, the purpose of this study was to show the calculation process of the test items validity uses the Aiken model. Subjects that were involved in the calculated process the validation of each test item were four experts, including two experts in the field of evaluation and two experts in the field of informatics engineering. The tool was used to assess the test items was a questionnaire consisting of 42 question items. The method used to analyze the test items was by comparing the results of the test item validation used the Aiken model with the validation classification according to Guilford. The study results were showed that from 42 test items tested; there were two invalid items and 40 valid items.

1. Introduction

The cognitive abilities of students in following a learning process in evaluation courses in universities can be measured through tests. A good test to use is a valid test and can accurately measure the cognitive abilities of students from easy to challenging levels. One form of test that can be used to meet the intended test criteria is a Superitem-based digital format test.

At present, superitem-based digital formatted test items are essential and necessary in making it easier for lecturers to make test questions that arranged in stages from easy to challenging levels with accurate value calculations and easily accessible. One application that can be used to create digital-form test items with a tiered arrangement from easy to difficult is Wondershare. Although the application of Wondershare, in general, is very suitable to be used as a facility to make digital-form test items in several universities, however, test items that were imported into the Wondershare application that was implemented in the Informatics Engineering Departement, Universitas Pendidikan Ganesha wasn’t known certainly it content validity yet. For that, so it is necessary to test the content validity of the test items by using an appropriate calculation model. One of the formula used to test the content validity of an instrument is the Aiken formula.

Based on that situation, so the problem statement of this research was how the level of content validity of digital test items based on Superitem-Wondershare that was used in the Evaluation course at Informatics Education Department, Universitas Pendidikan Ganesha with the Aiken formula.
calculation process? The purpose of this research showed the process of calculating the Aiken formula used in measuring the level of content validity of digital test items for evaluation courses.

Some of the studies that underlie this research, including research were conducted in 2016 by Balan, and Djuniadi [1] showed the use of the Wondershare quiz creator application to make digital tests in English subjects for vocational students. Other research was conducted in 2017 by Gaol, Khumaidi, and Masrukan [2], showed the results of calculating the content validity of the test items using the Aiken formula that was conducted by three experts.

The research that was conducted in 2018 by Ningdyah, Greenwood, and Kidd [3], showed the use of Aiken formula in determining the coefficient of content validity for measuring instruments characteristic of educational models in Indonesia. The research that was conducted in 2018 by Ikhsanudin, and Subali [4] showed that two quantitative approaches could be used to measure the content validity test, namely Aiken’s V and Lawshe’s CVR. The research that was conducted in 2018 by Apawu, Owusu-Ansah, and Akayuure [5] showed the use of superitem tests to analyze and monitor students’ level of thinking in learning algebra. Research that was conducted in 2019 by Bahri and Supahar [6] showed a quantitative approach using the Aiken formula in calculating the content validity of religious test items integrated with science. Research that was conducted in 2019 by Rahmawati, Priatna, and Juandi [7] showed the use of superitem test models in exploring students’ critical thinking abilities by expanding abstract level thinking into thinking with algebraic characters.

2. Method
The approach that was used in this research was the development of instruments. The stages of instrument development consist of 6 stages [8], including 1) determination of variables, 2) determination of indicators, 3) determination of instrument items, 4) instrument trials, 5) instrument analysis, and 6) determination of final instruments. At the stage of determination of variables carried out the determination of the chapters that were used as the main component of test questions. At the stage of determination of indicators carried out by doing the arranged of chapters into sub-chapters. At the stage of determination of instrument, items were carried out making items that were sourced from the sub-chapter. At the trial stage, the instrument was tested for content validity and instrument reliability. However, specifically in this study, the instrument trials focused on content validity testing only. Subjects were involved in conducting the content validity testing were four experts consisting of two experts in the field of informatics education and two experts in the field of evaluation. The tool was used by experts to assess instrument items was a questionnaire consisting of 42 question items. At the instrument analysis stage, was carried out an analysis of the test results of the instrument's content validity. Analysis of the results of the instrument content validation was done by interpreting the results of the content validation (which had been obtained using the Aiken formula) based on the Guilford content validation reference. The Aiken formula [9] can be shown through equation (1), while the categorization of the Guilford’s content validation can be shown in Table 1[10].

$$V = \frac{\sum S}{n(c-1)}$$

Notes:
- $V$: Expert agreement index regarding item validity
- $n$: number of experts
- $r_n$: score assessment that given by expert-$n$ on item-$n$
- $S$: $r_n - I_0$
- $I_0$: lowest score in the scoring category
- $c$: number of choice scores

| Validity Score Range | Category               |
|----------------------|------------------------|
| 0.80 < $r_{xy}$ ≤ 1.00 | Very high validity     |
| 0.60 < $r_{xy}$ ≤ 0.80 | High validity          |
| 0.40 < $r_{xy}$ ≤ 0.60 | Medium validity        |
| 0.20 < $r_{xy}$ ≤ 0.40 | Low validity           |
| 0.00 < $r_{xy}$ ≤ 0.20 | Very low validity      |
| $r_{xy}$ ≤ 0.00       | Invalid                |

Table 1. The Categorization of Content Validation Refers to Guilford
3. Results and Discussion
Some of the results were obtained in this research was based on the completion of the instrument development stages. The results of this research can be explained in full as follows.

a) Variable Determination Results
The variables in the digital test based on Superitem-Wondershare for evaluation course were sourced from the chapters listed in the book entitled “Program Evaluation” authored by Dewa Gede Hendra Divayana. The variables can be shown in full in Table 2.

| Chapter | Variable                                                                 |
|---------|--------------------------------------------------------------------------|
| I       | The basic concept of program evaluation                                  |
| II      | Types of evaluation models                                               |
| III     | Strengths and weaknesses of program evaluation models                    |
| IV      | Research methodology in program evaluation                               |
| V       | Instruments in program evaluation                                        |
| VI      | Decision-making steps in program evaluation                              |

b) Indicator Determination Results
The indicators in a digital test based on Superitem-Wondershare for evaluation courses are sourced from sub-chapters of each variable previously was described in Table 2. Those indicators that intended can be shown in Table 3.

| Chapter | Indicator                                                                                                               |
|---------|--------------------------------------------------------------------------------------------------------------------------|
| I       | (1.1) Definition of Evaluation; (1.2) Definition of Program Evaluation; (1.3) Purpose of Program Evaluation; (1.4) Function of Program Evaluation; (1.5) Program Evaluator; (1.6) Requirements of Program Evaluator; (1.7) Elements of Program Evaluation; (1.8) Scope of Program Evaluation |
| II      | (2.1) Goal Oriented Evaluation Model; (2.2) Goal Free Evaluation Model; (2.3) Formative Sumatif Evaluation Model; (2.4) Countenance Evaluation Model; (2.5) Responsive Evaluation Model; (2.6) CSE-UCLA Evaluation Model; (2.7) Context, Input, Process, Product Evaluation Model; (2.8) Discrepancy Model |
| III     | (3.1) Strengths and Weaknesses of the Goal Oriented Evaluation Model; (3.2) Strengths and Weaknesses of the Free Goal Evaluation Model; (3.3) Strengths and Weaknesses of the Formative Sumatif Evaluation Model; (3.4) Strengths and Weaknesses of the Countenance Evaluation Model; (3.5) Strengths and Weaknesses of the Responsive Evaluation Model; (3.6) Strengths and Weaknesses of the CSE-UCLA Evaluation Model; (3.7) Strengths and Weaknesses of the CIPP Evaluation Model; (3.8) Strengths and Weaknesses of the Discrepancy Evaluation Model |
| IV      | (4.1) Research Approach; (4.2) Research Methods; (4.3) Research Design; (4.4) Research Instruments; (4.5) Research Subjects; (4.6) Data Collection Techniques and Procedures; (4.7) Data Analysis Techniques |
| V       | (5.1) Closed Instruments; (5.2) Open Instruments                                                                         |
| VI      | (6.1) Setting of the Evaluation Standards; (6.2) Calculation of Average Scores on Each Evaluation Component and Performing a Conversion to a Scale of Five; (6.3) Changing the Score that Obtained into the Standard Score (Z-Score); (6.4) Changing Z-Score into T-Score; (6.5) Interpretation of T-Score into the Glickman Quadrant; (6.6) Comparing Evaluation Results with Evaluation Standards |

c) Results of Determination of Instrument Items
Instrument items in digital tests based on super item-Wondershare for evaluation courses were sourced from questions on each of the indicators were described earlier in Table 3. Those instrument items intended can be shown in Table 4.
**Table 4. Instrument Items in Digital Test Based on Superitem-Wondershare for Evaluation Courses**

| Chapter | Indicator |
|---------|-----------|
| I       | 1. Explain the evaluation means!  
         | 2. Explain what the program evaluation means!  
         | 3. Explain the purpose of program evaluation!  
         | 4. Explain the function of program evaluation!  
         | 5. Explain the program evaluator means!  
         | 6. Explain the requirements to be a program evaluator!  
         | 7. Explain the elements of program evaluation!  
         | 8. Explain the scope of program evaluation! |
| II      | 9. Explain about the **Goal-Oriented Evaluation Model**!  
         | 10. Explain about the **Goal Free Evaluation Model**!  
         | 11. Explain about the **Formative Sumatif Evaluation Model**!  
         | 12. Explain about the **Countenance Evaluation Model**!  
         | 13. Explain about the **Responsive Evaluation Model**!  
         | 14. Explain about the **CSE-UCLA Evaluation Model**!  
         | 15. Explain about the **Context, Input, Process, Product Evaluation Model**!  
         | 16. Explain about the **Discrepancy Model**! |
| III     | 17. Explain the strengths and weaknesses of the **Goal-Oriented Evaluation Model**!  
         | 18. Explain the strengths and weaknesses of the **Free Goal Evaluation Model**!  
         | 19. Explain the strengths and weaknesses of the **Formative Sumatif Evaluation Model**!  
         | 20. Explain the strengths and weaknesses of the **Countenance Evaluation Model**!  
         | 21. Explain the strengths and weaknesses of the **Responsive Evaluation Model**!  
         | 22. Explain the strengths and weaknesses of the **CSE-UCLA Evaluation Model**!  
         | 23. Explain the strengths and weaknesses of the **CIPP Evaluation Model**!  
         | 24. Explain the strengths and weaknesses of the **Discrepancy Evaluation Model**! |
| IV      | 25. Explain the approach is used in evaluation research!  
         | 26. Explain the method is used in evaluation research!  
         | 27. Explain the design is used in evaluation research!  
         | 28. Explain closed and open instruments in evaluation research!  
         | 29. Explain the forms of instruments is used in evaluation research!  
         | 30. Explain the steps in determining the research subject in evaluation research!  
         | 31. Explain the techniques and procedures for collecting data in evaluation research!  
         | 32. Explain data analysis techniques in evaluation research! |
| V       | 33. Explain the form of closed instruments (questionnaire forms) is used in evaluation research!  
         | 34. Explain the form of the open instruments (forms of interview guidelines, observation guidelines, and documentation) is used in evaluation research! |
| VI      | 35. Explain the stages in setting evaluation standards!  
         | 36. Explain the calculation of average scores on each evaluation component and convert to a scale of five!  
         | 37. Explain the calculation of average scores in each evaluation component!  
         | 38. Explain how to convert the calculation results of the average score to a scale of five!  
         | 39. Explain how to change the score that obtained into the standard score (Z-Score)!  
         | 40. Explain how to change Z-Score into T-Score!  
         | 41. Explain how to interpret the T-Score into the Glickman quadrant!  
         | 42. Explain how to compare evaluation results with evaluation standards! |

**d) Instrument Trial Results**

The instrument trials were carried out by four experts. The results of the trial instrument can be seen in full in Table 5.

**Table 5. Trial Results of the Digital Test Based on Superitem-Wondershare for Evaluation Courses**

| Items | Expert | I | II | III | IV | S1 | S2 | S3 | S4 | ΣS | V  | Category of Validity |
|-------|--------|---|----|-----|----|----|----|----|----|----|----|---------------------|
| 1.    | 4      | 5 | 5  | 4   | 4  | 3  | 4  | 4  | 3  | 14 | 0.875  | Very High           |
| 2.    | 4      | 5 | 5  | 4   | 4  | 3  | 4  | 4  | 3  | 14 | 0.875  | Very High           |
| 3.    | 5      | 5 | 4  | 4   | 4  | 3  | 3  | 3  | 4  | 14 | 0.875  | Very High           |
| 4.    | 4      | 5 | 5  | 4   | 4  | 3  | 4  | 4  | 3  | 14 | 0.875  | Very High           |
| 5.    | 5      | 4 | 4  | 5   | 4  | 3  | 3  | 3  | 4  | 14 | 0.875  | Very High           |
| 6.    | 4      | 4 | 4  | 5   | 3  | 3  | 3  | 3  | 4  | 13 | 0.813  | Very High           |
| 7.    | 4      | 4 | 5  | 5   | 3  | 3  | 4  | 4  | 4  | 14 | 0.875  | Very High           |
| 8.    | 4      | 5 | 4  | 5   | 3  | 4  | 3  | 3  | 4  | 14 | 0.875  | Very High           |
| 9.    | 4      | 4 | 5  | 5   | 3  | 3  | 4  | 4  | 4  | 14 | 0.875  | Very High           |
Based on the results of instrument analysis, there were 40 valid instrument items and two invalid instrument items. Invalid items were item-28 and item-36, while valid items were all other items besides those two items. The valid items were entered into the Wondershare application. The display of the final instruments of digital test based on Superitem-Wondershare can be shown in Figure 1.

**Figure 1.** The Display of Final Instruments Entered into the Wondershare Application
The results of this study have answered the limitations of the research of Gaol, Khumaedi, and Masrukan which only involved three experts in conducting the validity test of the instrument, whereas in this study it had involved four experts. The results of this study have also answered the constraints of Apawu, Owusu-Ansah, and Akayuure research which only showed the use of superitem-based tests for learning algebra, but through this study also demonstrated the use of superitem-based tests for evaluation courses. The obstacle that is still found in this study is the source of the test questions, which is still limited sourced from just one book.

4. Conclusions
The content validity of the digital test items based on Superitem-Wondershare was included in the very high category based on the categorization of the content validation of Guilford, so making it suitable for measuring the student’s knowledge in understanding the evaluation courses. The solution as problem-solving to the obstacle in this research can be done by creating more varied test questions sourced from e-books, journals, and other documents related to evaluation.

5. Acknowledgment
The authors would like to express their highest thanks to the Head of the Research and Community Service Institute of Universitas Pendidikan Ganesha that had provided the opportunities and had given the funding in this research through research grant contract number 365/UN48.16/LT/2019.

References
[1] Balan Y A and Djuniadi 2016 Implementing media of computer-based test for english language proficiency of vocational high school students at the borderline between indonesia and Timor Leste SENIT 2016 pp 79–88
[2] Gaol P L, Khumaedi M, and Masrukan 2017 Development of instrument for self-confidence character assessment in mathematics subjects at middle school J. Edu. Research & Eval. 6(1) pp 63–70
[3] Ningdyah A E M, Greenwood K M, and Kidd G 2018 A Training-model scale’s validity and reliability coefficients: expert evaluation in Indonesian professional psychology programs Makara Hum. Behav. Stud. in Asia 22(1) pp 56–66
[4] Ikhsanudin, and Subali B 2018 Algebraic thinking characteristics of eighth grade junior high school students based on Superitem Test of SOLO model The 5th International Conference on Research, Implementation, & Education of Mathematics and Sciences, IOP Conf. Ser.: J. Physics: Conf. Ser. 1097 pp 1–9
[5] Apawu J, Owusu-Ansah N A, and Akayuure P 2018 A study on the algebraic working processes of senior high school students in Ghana Euro. J. Scie. & Math. Educ. 6(2) pp 62–68
[6] Bahri M F, and Supahar 2019 Content validity and reliability analysis of integrated islamic-science test instrument to measure the student’s critical thinking ability J. Islam. & Soc. Stud. 5(1) pp 39–48
[7] Rahmawati D I, Priatna N, and Juandi D 2019 Algebraic thinking characteristics of eighth grade junior high school students based on Superitem Test of SOLO model International Conference on Mathematics and Science Education (ICMScE 2018), IOP Conf. Ser.: J. Physics: Conf. Ser. 1157 pp. 1–7
[8] Sugiharni G A D, Setiasih N W, Mahendra I W E, Ardana I M and Divayana D G H 2018 Development of alkin model instruments as evaluation tools of blended learning implementation in discrete mathematics course on stikom Bali J. Theor. Appl. Inf. Technol. 96(17) pp 5803–5818
[9] Retnawati H 2016 Proving content validity of self-regulated learning scale (the comparison of aiken index and expanded gregory index) Research & Eval.in Educ. 2(2) pp 155–164
[10] Suyasa P W A, Kurniawan P S, Ariawan I P W, Sugandini W, Adnyawati NDMS, Budhyani IDAM, and Divayana D G H 2018 Empowerment of CSE-UCLA model based on glickman quadrant aided by visual application to evaluate the blended learning program on SMA Negeri 1 Ubud J. Theor. Appl. Inf. Technol. 96(18) pp 6203–6219