Development of A Thematic-Based Performance Assessment Model for the Elementary School Exam in Indonesia

Ni Luh Sekar Astuti[1], Supahar[2], Edi Istiyono[3]

Graduate Program in Educational Research and Evaluation, Yogyakarta State University

ni0005pasca.2020@student.uny.ac.id, supahar@uny.ac.id, edi_istiyono@uny.ac.id

ABSTRACT

This development research aims to develop a thematic-based performance assessment model for elementary school level exams to measure students' skills at the end of the level. The development of the performance appraisal model is carried out by referring to the principle of developing an assessment instrument and adapting the ADDIE development model. The development of the performance appraisal model produces outputs in the form of performance appraisal instruments and Student Worksheets (LKPD) as a guide for students and teachers in carrying out performance appraisals. The product in this development has been tested for content validity, construct validity, and reliability, and field tested to measure students' ability to determine the instrument's quality. In addition, the performance appraisal model developed has also been analyzed for its implementation and effectiveness as a tool for measuring student skills. The results of this study include the following: (a) content validity above 0.80 (high); (b) all of the assessment items/indicators have a loading factor value above 0.60 and can be said to be valid in measuring the performance appraisal construct; (c) reliability of 0.941 (very high); (d) the difficulty or item location is in the range between -2.86 to 1.80 or is in the easy to a difficult category and is accurate for measuring the ability of students in the range -3 to +2; (e) discrimination is in the range of 1.33 to 38.21 with good to very good categories; (f) the overall implementation of the performance appraisal model reached an average value of 97.96% or was well implemented; and (g) the overall performance appraisal model developed in this study has been declared effective (valid/reliable/objective/systematic/practical) in measuring students' skills.

Keywords: performance assessment, thematic, school exam

INTRODUCTION

The Covid-19 pandemic has had a significant impact on the implementation of education in Indonesia. The Minister of Education and Culture (2020) found a decrease in student academic achievement when implementing distance learning compared to face-to-face learning. This is due to learning lost by students, so the government provides the option to adjust the implemented curriculum into an emergency curriculum with simplified basic competencies. In addition to curriculum changes, national or equivalent exams are also abolished and replaced with exams organized by the education unit as a determinant of student graduation at the end of the level.

The results of Mardiana's research (2021) show that educators experience obstacles in carrying out assessments both in the realm of knowledge, skills, and attitudes due to the low attendance of students in the learning process and the presence of students who do not collect assignments given by educators. Exams during the Covid-19 pandemic greatly impact the assessment of skills that students should possess in accordance with the established curriculum. This is because the teacher cannot fully observe the process of practicing skills, so the results obtained do not reflect the abilities of students (Prawanti & Sumarni. 2020).

The difficulties experienced by teachers in carrying out assessments during the pandemic are related to the low learning motivation of students due to the abolition of the National Examination, which coincides with distance learning during the Covid-19 pandemic. This is supported by the results of research by Ghani & Zharfa (2020), which found a significant effect between the abolition of the National Examination (UN) and the motivation of students to learn during the Covid-19 pandemic. It can be seen that the learning motivation of students has decreased from before the abolition of the National Examination to after the abolition of the National Examination by a difference of 12.81%, namely from an average of 70.60% before the abolition of the National Examination to 57.79% in the conditions after the abolition of the National Examination, with the level of student learning motivation is in the medium category.

Before the Covid-19 pandemic and the National Examination was still being held, when students stepped into class at the end of the level, as educators or parents, they would make various efforts to increase learning motivation through tutoring, group study, or studying with their parents, as well as with their parents which motivate students both physically and psychologically. However, in reality, when the National Examination is abolished, educators and parents reduce additional hours of study because they assume that the grades used as consideration for graduation only use the average value of report cards, so there is no need to prepare for school exams. However, this step has an impact on children's learning motivation, where children also tend not to study and have more free time, so they use it more to do other activities that do not support their cognitive, affective, and psychomotor achievements like playing games or going out with other friends (Ghani & Zharfa. 2020).

The assessment of student learning outcomes by eliminating the National Examination is one of the implementations of the Merdeka Belajar policy revolution in order to improve the quality of graduates in the realm of literacy and numeracy (Mendikbud. 2020). The root of the problem of the less-than-optimal assessment for determining graduation carried out by educational units during the Covid-19 pandemic lies in the inability of school exam instruments to motivate students. This is probably because school exams are not challenging and do not have a significant effect or impact on students' graduation, so they do not stimulate students' adrenaline to make optimal efforts to get the best results. Therefore, it is necessary to design a school exam assessment process that is challenging and impactful for students, both during the assessment process and after the assessment process is carried out.

Considering the assessment guide in the 2013 curriculum and several previous research results, the researcher sees an opportunity to increase students' learning motivation through authentic assessments. One of them is the assessment of student performance, either through

performance, projects, portfolios, etc. Chappuis et al. (2012) revealed that performance appraisal is one of the alternative assessments that can provide multidimensional assessments in real and authentic situations. The performance appraisal process is carried out by observing and evaluating a process that raises skills, attitudes, and products simultaneously. Performance appraisal is generally used to assess the skill domain because it shows the student's performance in doing something. It can also be used to measure aspects of attitudes and knowledge through the performance shown by students. Because basically, performance appraisal is an assessment that trains students in applying their knowledge and competence in constructing something.

Performance appraisal can be used as a strategy or alternative in assessing learning outcomes that train students' literacy and numeracy skills. With performance appraisals, teachers can directly observe the performance shown by students and make decisions about the quality of learning outcomes they show. Performance assessment can be used as a basis for evaluating the abilities displayed by students in the form of processes and results. Performance assessments can also improve the learning process because performance assessments help teachers make decisions during the learning process or in planning for further learning. Harida & Junanto's research (2018) found that project-based performance appraisals can also improve students' literacy skills, especially scientific literacy. This indicates that in order to improve the quality of education in terms of literacy and numeracy skills.

The results of Desi & Dadan's research (2017) show that performance appraisal can be attached to the learning process, in which the syntax or learning steps according to the applied learning model can be used to evaluate student performance. Therefore, in developing performance appraisal instruments, it is necessary to analyze or determine the learning model applied in students' learning process. Learning in elementary schools with the 2013 curriculum is done on a thematic basis, so the performance assessment needs to be adjusted to the thematic learning steps. The implementation of thematic learning is carried out in three stages, namely planning, implementation and assessment, and the success of thematic learning is influenced by the competence of the teacher, school/government support, and characteristics of students. Assessment of learning outcomes is carried out in accordance with the 2013 curriculum and the type of thematic learning used, can be arranged per subject characteristic, or contain several subjects in a single sub-theme or theme.

Based on these problems, this research was conducted with the aim of (a) designing a thematic-based performance appraisal model for elementary school examinations; (b) testing the validity and reliability of the assessment instrument; (c) analyzing the results of the measurement of learning outcomes with the developed performance appraisal model; (d) knowing the implementation of the performance appraisal model; and (e) knowing the effectiveness of the performance appraisal model to measure student learning outcomes.

METHOD

This study aims to develop a thematic-based performance appraisal model for elementary school examinations. The development of the performance appraisal model in this study adapts

the ADDIE development model (in Branch, 2009 & Dick et al., 2005), which consists of five stages of development, namely analysis, design, development, implementation, and evaluation.

The assessment model developed has been validated by experts and practitioners. LKPD validation is carried out by 1 (one) basic education expert/lecturer from public universities in Bali and 2 (two) elementary school practitioners/teachers. Meanwhile, the assessment instrument was validated by 2 (two) experts/lecturers and 2 (two) practitioners/primary school teachers. The results of expert validation were analyzed using Aiken's V equation. The content validity index uses the criteria of Gregory (Retnawati, 2016), which if the agreement index is less than 0.40, then it is said to be of low validity; between 0.40 to 0.80, it is said to be of moderate validity, and if it is more than 0.80 it is said high validity.

Subsequently, the performance appraisal model was tested in April until June 2022. The trial was conducted in nine (9) public and private elementary schools located in Badung Regency, Province of Bali, by considering the curriculum used and the status of the school as favorable or unfavorable. Respondents who participated in the trial of the performance appraisal instrument developed in this assessment were 371 people. Next, the test results of the instrument were analyzed using LISREL software through confirmatory factor analysis (CFA) to determine the validity, reliability, and suitability of the performance appraisal model developed in this study.

After the performance appraisal model is tested to be valid, reliable, and meet the goodness of fit criteria, the researchers tested the final product to measure student learning outcomes. Learning outcomes were measured for 2 (two) weeks on 22 students in one elementary school in Badung Regency, Bali Province, in July 2022. The measurement results are used to analyze the level of difficulty and discrimination of items, where data analysis was carried out using the Item Response Theory with the Generalized Partial Credit Model (GPCM) approach with R-Studio software.

In addition, the researchers also analyzed the implementation and effectiveness of the developed performance appraisal model. The implementation of the model is evaluated through the observation sheet of teacher and student activities in the assessment process carried out by the principal. The results of the observations were analyzed for implementation by calculating the percentage of implementation of the performance appraisal model. Meanwhile, the analysis of the effectiveness of the performance appraisal model developed was carried out by modifying the effectiveness of the appraisal model according to Sudiyatno et al. (2015). Validity and reliability are seen in the test results of the assessment instrument. Meanwhile, the objectivity, systematics, and practicality of the assessment model were assessed through the distribution of questionnaires to classroom teachers involved in the implementation of the trial.

RESULTS AND DISCUSSION

The product developed in this study is a thematic-based performance assessment model for elementary school-level examinations. The assessment model developed is focused on assessing student skills in Core Competency 4 Curriculum 2013, namely "menyajikan pengetahuan faktual dalam bahasa yang jelas dan logis dalam karya yang estetis, dalam gerakan

yang mencerminkan anak sehat, dan dalam tindakan yang mencerminkan perilaku anak beriman dan berakhlak mulia" (or, in English, "present factual knowledge in a clear and logical language in aesthetic works, in movements that reflect healthy children, and in actions that reflect the behavior of children with faith and noble character"). The performance appraisal model developed in this study has met the standards of content validity, construct validity, reliability, and model is fit and can be implemented well and effectively in measuring student learning outcomes.

Content Validity

Content validity was obtained through the validation of the performance appraisal model by several experts and related experts. The results of expert validation were analyzed using Aiken's V equation. The content validity index uses the criteria of Gregory (Retnawati, 2016), which if the agreement index is less than 0.40, then it is said to have low validity; between 0.40 to 0.80, it is said to be moderately valid, and if it is more than 0.80 it is said to be high. The results of content validity by experts (experts and practitioners) are presented in the following table.

Table 1. Content Validity of Assessment Instruments

Component	**Assessment Aspect**	**Validity per Item**														
		A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**B8**	**B9**	**B10**	**C11**	**D12**	**E13**	**E14**	**E15**
Psychomotor Aspect	Conformity of the indicator formula with the measurement dimensions	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	Conformity of criteria with the formulation of indicators in the grid	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Construction Aspect	Conformity of the sequence/systematic assessment	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94

Component	Assessment Aspect	Validity per Item														
		A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**B8**	**B9**	**B10**	**C11**	**D12**	**E13**	**E14**	**E15**
	with student activities															
	Assessment can be used in assessing the performance of elementary school students	0.88	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	Appropriateness of assessment time allocation	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
	Each criterion can be used to conduct an individual assessment	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Linguistic Aspect	The language used is communicative	0.94	1.00	1.00	0.94	0.94	0.94	0.94	0.94	0.94	1.00	0.94	0.94	0.94	0.94	1.00
	The language used is not ambiguous (double meaning).	1.00	1.00	1.00	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	The language used is standard in accordance with linguistic rules.	0.94	1.00	1.00	0.94	0.94	0.94	0.94	0.94	1.00	1.00	0.94	1.00	1.00	1.00	1.00

Component	Assessment Aspect	Validity per Item														
		A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**B8**	**B9**	**B10**	**C11**	**D12**	**E13**	**E14**	**E15**
	The sentence does not contain words relating to ethnicity, religion, race, and intergroup	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Average Validity per Item		**0.95**	**0.97**	**0.97**	**0.94**	**0.96**	**0.96**	**0.96**	**0.96**	**0.96**	**0.97**	**0.96**	**0.96**	**0.96**	**0.96**	**0.97**

Table 2. Content Validity of Student Worksheets (LKPD)

Component	Assessment Aspect	Validity per Component				
		Section A	Section B	Section C	Section D	Section E
		Planning	Implementation	Evaluation	Reflection	Reporting
Content/Material Aspect	Conformity of content/material with the formulation of indicators in the assessment grid	1.00	1.00	1.00	1.00	1.00
	Compatibility with the level of development of elementary school students	0.92	1.00	1.00	1.00	1.00
Construction Aspect	The suitability of each step/stage with the dimensions measured and the student learning process	1.00	1.00	1.00	1.00	1.00

Component	Assessment Aspect	Validity per Component				
		Section A	Section B	Section C	Section D	Section E
		Planning	Implementation	Evaluation	Reflection	Reporting
	The sequence of instructions/questions given	1.00	1.00	1.00	1.00	1.00
	Accuracy of the notation/symbol/term used	1.00	1.00	1.00	1.00	1.00
	Appropriateness of time allocation for LKS	0.92	0.92	1.00	1.00	0.92
Linguistic Aspect	The language used is communicative	0.75	0.83	0.92	0.92	0.83
	The language used is not ambiguous (double meaning)	0.92	0.92	0.92	0.92	0.92
	The language used is standard in accordance with linguistic rules	1.00	1.00	1.00	1.00	1.00
	The sentence does not contain words relating to ethnicity, religion, race, and intergroup	1.00	1.00	1.00	1.00	1.00
Typography Aspect	Size fit	1.00	1.00	1.00	1.00	1.00
	Harmony and consistency of layout	1.00	1.00	1.00	1.00	1.00
	Sufficient place to answer	1.00	1.00	1.00	1.00	0.92
	The use of a combination of fonts that is not excessive	1.00	1.00	1.00	1.00	1.00

Component	Assessment Aspect	Validity per Component				
		Section A	Section B	Section C	Section D	Section E
		Planning	Implementation	Evaluation	Reflection	Reporting
	The use of font variations that are not excessive	0.92	0.92	0.92	0.92	0.92
	Normal spacing between letters and lines	1.00	1.00	1.00	1.00	1.00
	Presentation of texts and images that are attractive and clear	0.92	1.00	1.00	1.00	1.00
Average Validity per Item		**0.96**	**0.98**	**0.99**	**0.99**	**0.97**

Based on the data in Table 1. and Table 2. above, it can be seen that all indicators in the developed assessment instrument can be said to be valid in the high category. Meanwhile, the components in the LKPD can also be said to be valid with high criteria, except for the linguistic planning component, which is quite communicative and the validity is moderate. Several experts and experts also provide recommendations for improvements to the LKPD and assessment instruments that have been developed. Recommendations from experts and practitioners are followed up by revising the LKPD and assessment instruments, after which confirmation is made to experts and practitioners, as well as supervisors, until the instruments and LKPD are declared suitable for use in performance assessments on thematic-based elementary school level exams.

Construct Validity and Reliability

The construct validity and reliability tests on the components were carried out using empirical data from the performance appraisal model test results, which consisted of 15 assessment indicators. The results of the validity test of the performance assessment instrument on the thematic-based elementary school exams developed in this study are as follows.

Table 3. Validity Test Results

Dimension	Item Number	Indicator	Loading Factor	Description
Planning	A1	Students are able to choose a study topic according to their field of interest.	0.76	Valid
	A2	Students are able to analyze problems that occur in the daily lives of people in the surrounding environment related to the field of interest.	0.75	Valid
	A3	Students are able to synthesize alternative solutions in solving problems using their multiple intelligences.	0.63	Valid
	A4	Students are able to design projects to realize alternative solutions to solving problems.	0.81	Valid
	A5	Students are able to prepare the necessary tools and project implementation materials independently.	0.80	Valid
	A6	Students are able to develop procedures or work steps in completing projects.	0.70	Valid
	A7	Students are able to develop a project implementation plan independently.	0.69	Valid
Implementation	B8	Students are able to carry out projects independently.	0.77	Valid
	B9	Students are able to compile reports on the progress of implementing alternative solutions on a regular basis.	0.91	Valid
	B10	Students are able to overcome the challenges faced in project implementation.	0.66	Valid
Evaluation	B11	Students are able to evaluate the process of implementing alternative solutions to address the root of the problem objectively.	1.00	Valid
Reflection	B12	Students are able to reflect on the process of implementing alternative solutions to address the root of the problem objectively.	1.00	Valid

Dimension	Item Number	Indicator	Loading Factor	Description
Reporting	B13	Students are able to report project implementation in the form of written or verbal reports (audio, video, or audio-visual).	0.82	Valid
	B14	Students are able to choose a study topic according to their field of interest.	0.70	Valid
	B15	Students are able to analyze problems that occur in the daily lives of people in the surrounding environment related to the field of interest.	0.69	Valid

Based on the validity test results, all assessment indicators have a loading factor value above 0.60, so it can be concluded that all items in the planning component are valid to be used as a tool to measure the construct of the variables (Retnawati, 2016).

Meanwhile, the reliability of the performance appraisal instrument developed in this study was tested using Cronbach's Alpha Formula with the following results.

Table 4. Reliability Test Results

Reliability Statistics	
Cronbach's Alpha	N of Items
0.941	15

Based on the reliability test results, referring to the reliability criteria, it can be seen that the reliability of the performance appraisal instrument on the planning component developed by the researcher is in the reliable category or 0.941 (Retnawati, 2016). Therefore, it can be concluded that the thematic-based performance assessment instrument on the thematic-based elementary school exams that have been developed has very high consistency in measuring aspects of student performance.

Goodness of Fit (GOF)

The test results are also used to test the suitability of the model with Structural Equation Model (SEM) analysis through the LISREL software tool. The initial model suitability test results showed that the model did not meet the GOF criteria, so modifications were needed. The suggested modification of the LISREL program is to add the error covariance value in the analysis. The results of the suitability test of the assessment model after being modified are as follows.

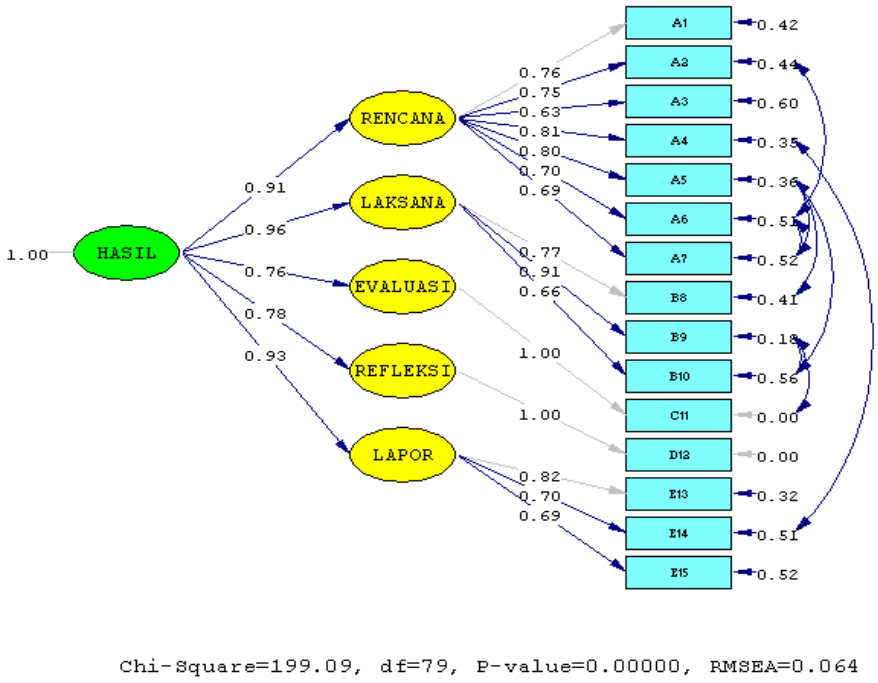

Figure 1. Fitness Test of Model Assessment After Modification

Based on the existing GOF criteria, where RMSEA (0.064) < 0.08; 0.95 < NFI (0.98) 1.00; 0.95 < NNFI (0.98) 1.00; 0.97<CFI (0.99) 1.00; IFI (0.99)>0.9; RFI (0.97)>0.9; and 0.90≤ AGFI (0.90), then GOF is met. It can be concluded that the model fits the data because the RMSEA of 0.064 is smaller than 0.08, and several other GOF criteria have been met, although the p-value (0.00000) is smaller than 0.05.

Further analysis of the correlation of the two variables indicated by the Pearson Correlation (R) value, it was found that the correlation coefficient between: (1) the planning dimension and the assessment results was 0.92; (2) the implementation dimension with an assessment result of 0.97; (3) the evaluation dimension of the assessment results is 0.74; (4) the dimension of reflection on learning outcomes is 0.77; and (5) reporting dimensions with learning outcomes of 0.94. It can be seen that the strongest correlation with performance appraisal results is the implementation dimension, with a correlation coefficient of 0.97. Meanwhile, the weakest correlation is the evaluation dimension, with a correlation coefficient of 0.74 (strong). However, the overall correlation of each assessment dimension with the assessment results is strong and very strong because it shows a correlation coefficient greater than 0.5 (strong) and 0.75 (very strong).

Analysis of Student Performance Measurement Results

The measurement results describe the quality of the assessment instrument as seen through the difficulty level and discrimination parameters. The results of the analysis of the quality of the performance appraisal instrument per assessment indicator developed are as follows.

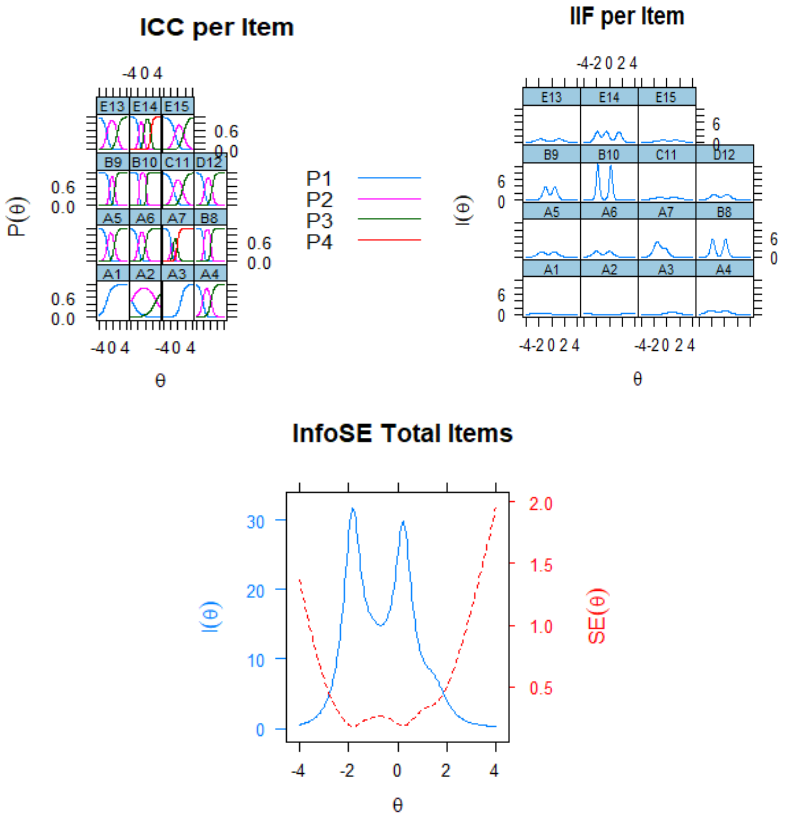

Figure 2. Results of the Analysis of the Characteristics of Items

The following is a recapitulation of the difficulty level parameter index and the overall distinguishing power of the performance appraisal indicators developed in this study.

Table 5. Recapitulation of Difficulty Levels and Distinguishing Power

Item Number	Discrimination (a)	Difficulty Level (b)	Difficulty Level		
			Category 2 (b_1)	Category 3 (b_2)	Category 4 (b_3)
1	1.93	-0.80	-2.69	-0.42	0.70
2	3.57	-0.56	-1.91	-0.36	0.60
3	38.21	-0.52	-1.91	-0.35	0.69
4	1.33	-1.01	-1.66	-1.59	0.21
5	2.08	-0.55	-1.91	-1.02	1.29
6	3.91	-0.52	-1.92	-0.37	0.73
7	1.47	-0.61	-1.10	-1.32	0.59
8	1.33	-0.97	-2.86	-0.78	0.73

Item Number	Discrimination (a)	Difficulty Level (b)	Difficulty Level		
			Category 2 ($b_{1)}$	Category 3 ($b_{2)}$	Category 4 ($b_{3)}$
9	33.74	-0.68	-2.48	-0.69	1.12
10	1.71	-0.40	-1.97	-0.57	1.33
11	2.01	-0.47	-2.68	-0.53	1.79
12	3.59	-0.48	-1.96	-0.21	0.72
13	2.02	-0.53	-2.64	-0.75	1.80
14	22.95	-0.61	-1.89	-0.70	0.77
15	33.74	-0.68	-2.48	-0.69	1.12

Description:

a : estimated value of discrimination parameter

b : the estimated value of the difficulty level parameter

b_1 : the estimated value of the difficulty level parameter to obtain a score of 2

b_2 : the estimated value of the difficulty level parameter to obtain a score of 3

b_3 : the estimated value of the difficulty level parameter to obtain a score of 4

The results of the discriminatory analysis showed that all items/indicators in the developed performance appraisal instrument were not found to have a discrimination index of less than 0.25. The discrimination index ranges from 1.33 to 38.21, with a good to very good category (Hambelton & Swaminathan, 1985). The highest discrimination index is located on indicator item number 3, and the lowest is on item 10. This shows that the assessment instrument developed can distinguish students from high and low groups on indicators of finding alternative solutions in solving problems related to the chosen topic/theme. Followed by item number 15 indicator, namely the ability of students to present performance in second place, and item number 14 indicator related to students' ability to compile portfolios in third place. Meanwhile, point 10, which is related to the ability to overcome challenges in project implementation, has the lowest discrimination index.

Looking at the estimated value of the parameter level of difficulty of the performance appraisal instrument above, it was found that the performance appraisal instrument developed in this study produced an accurate, valid, and reliable instrument to measure students' ability in the range of -3 to +2. The grain difficulty index ranged from -2.86 to 1.80 or is in the easy to difficult category (Hambelton & Swaminathan, 1985). The highest difficulty index is in the

indicator item number 10 (ability to overcome challenges in project implementation) and is followed by item indicator number 11 (ability to carry out evaluations). Meanwhile, the indicator with the lowest difficulty index is indicator number 4. namely the ability to design projects. This shows that item number 10 is less able to distinguish the ability of students in the high and low groups due to the high level of difficulty of the questions, causing students from both ability groups to not be able to achieve the maximum score.

The researcher also found that the further to the right of the data in the table, the higher the difficulty index value. This shows that it is increasingly difficult to obtain a higher score category or the higher the ability of students is needed to obtain the highest score category. This is in line with the concept of developing a scoring guide in which the highest score category represents the most optimal ability/performance of students, so it is natural that the difficulty index from category 2 to category 4 is higher.

Implementation of the Performance Appraisal Model

The results of the analysis of the implementation of the performance assessment model on the thematic-based elementary school exams developed in this study indicate that all indicators of the implementation of the assessment have been implemented well. The summary of the results of the analysis of the implementation of the assessment model per dimension can be seen in the following table.

Table 6. Results of Analysis of the Implementation of the Assessment Model

No.	Dimension	Percentage	Description
1	Planning	99.60%	well executed
2	Implementation	99.07%	well executed
3	Evaluation	100.00%	well executed
4	Reflection	100.00%	well executed
5	Reporting	91.67%	well executed
Average		97.96%	well executed

Overall, the performance appraisal model developed in this study has been implemented well. This indicates that the syntax/stages in the performance appraisal model can be determined to be implemented properly by the teacher since it is practical and easy to implement by the teacher. The lowest percentage of implementation on the reporting dimension is due to the low ability of students to write reports, create portfolios, and present reports, so students could not complete this stage properly.

Effectiveness of Assessment Model

The results of the analysis of the effectiveness of the performance appraisal model on the thematic-based elementary school examinations developed in this study indicate that all indicators of the implementation of the assessment are valid, reliable, objective, systematic, and practical. The summary of the results of the analysis of the effectiveness of the assessment model per criteria can be seen in the following table.

Table 7. Results of Analysis of the Effectiveness of the Assessment Model

No.	Dimension	Analysis Results	Description
1	Validity	loading factor > 0.60	Valid
2	Reliability	$\alpha = 0.941$	Reliable
3	Objectiveness	$\bar{x} = 4.09$	Objective
4	Systematic	$\bar{x} = 4.58$	Systematic
5	Practicality	$\bar{x} = 3.69$	Practical

The performance appraisal model developed in this study as a whole has been declared effective in measuring students' skills. Effectiveness, in this case, means that the performance appraisal model has a valid, reliable, and objective assessment instrument in providing a systematic and practical assessment in the implementation phase.

CONCLUSIONS AND SUGGESTIONS

Based on the final product review. The performance appraisal model developed in this study has been tested for quality, implementation, and effectiveness in measuring the skill domains of students at the end of the level. The assessment instrument developed has a very good discrimination index with a moderate to difficult level of difficulty. The assessment instrument developed is accurate in measuring the ability of students in the range of -3 to +2. The scoring guidelines that have been developed also have differences per category. The higher the score, the more difficult it is for students to fulfill. Besides that, the assessment model developed has also been proven to be well implemented by teachers and is effective in measuring the skills of students at the end of the level on a thematic basis.

Some limitations in the research conducted in accordance with the final product development results, among others as follows: (a) this study was only conducted in Badung Regency, Bali Province, so the results cannot be generalized to schools in other districts/provinces with different characteristics. ; (b) this research was conducted in schools that implement the 2013 curriculum so that modifications or adjustments need to be made if it will be applied in schools with other curricula; (c) this study only measures the skill domains of students at the end of elementary school level so that in calculating the final grades it needs to be integrated with cognitive and affective assessments that have been carried out by teachers in each educational

unit; (d) the analysis of the quality of the instrument in this study only focuses on the level of difficulty and distinguishing power with the assumption that the GPCM model is the most suitable so that in the future it is necessary to test the suitability of the model with the Item-Response Theory (IRT) theory; and (e) the analysis of the effectiveness of the model only examines the model in terms of validity, reliability, objectivity, systematics, and practicality of the model so that in the future it is necessary to analyze the effectiveness of the assessment model in improving learning outcomes or student motivation.

REFERENCES

1. Branch, R. M. (2009). Instructional Design: The ADDIE Approach. New York: Springer.
2. Chappuis, J., Stiggins, R., Chappuis, S., & Arther, J. (2012). Classroom Assessment for Student Learning, 2nd ed. Boston, MA: Pearson.
3. Dick, W., Carey, L., & Carey, J. O. (2015). The systematic design of instruction (8th ed.). Boston: Pearson.
4. Ghani, S., & Zharfa, M. (2020). Pengaruh Penghapusan Ujian Nasional Terhadap Motivasi Belajar Peserta Didik Di Masa Pandemi. Jurnal Pendidikan Tematik, 1(3), 184-196.
5. Hambleton, R. K., & Swaminathan, H. (1985). Item response theory principles and applications. New York: Kluwer-Nijhoff Publishing. https://doi.org/10.1007/978-94- 017-1988-9.
6. Mardiana. (2021). Identifikasi Hambatan-Hambatan Guru dalam Pelaksanaan Pembelajaran Daring pada Masa Pandemi Covid-19 di Sekolah Dasar. Skripsi, tidak diterbitkan. Universitas Jambi.
7. McMillan, J. H. (2018). Classroom Assessment; Principles and Practise that enhance Student Learning and Motivation. 7th ed. New York: Pearson Education.
8. Nitko, A. J. & Brookhart, S. M. (2014). Educational Assessment of Students, 6th ed. United States: Pearson.
9. Menteri Pendidikan dan Kebudayaan. (2020). Surat Edaran Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 4 Tahun 2020 tentang Pelaksanaan Kebijakan Pendidikan dalam Masa Darurat Penyebaran Coronavirus Desease (COVID-19).
10. Popham, W. J. (2017). Classroom Assessment: What Teachers Need to Know, Eight Edition. Los Angeles: Pearson.
11. Prawanti, L. T., & Sumarni, W. (2020). Kendala Pembelajaran Daring Selama Pandemic Covid-19. Seminar Nasional Pascasarjana 2020, UNNES, 286-291, ISSN: 2686 6404.
12. Purnamasari, R. & Purnomo, H. (2021). Implementasi Kurikulum 2013 Pada Pembelajaran Tematik-Integratif di Sekolah Dasar. Jurnal Ilmiah PGSD FKIP Universitas Mandiri, 7(1), 163-174, p-ISSN: 2477-5673, e-ISSN: 2614-722X.
13. Putri, A. C. (2015). Pelaksanaan Penilaian Autentik dalam Pembelajaran Tematik Pada Siswa Kelas IV A Sekolah Dasar Negeri 4 Wates Kecamatan Wates Kabuoaten Kulon Progo. Skripsi, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.
14. Realita. (2013). Kompetensi Guru MIN Sabang dalam Melaksanakan Evaluasi Pembelajaran; Telah atas Konstruksi Instrumen Penilaian Berbasi Kurikulum 2013. Junal Ilmia DIDAKTIKA, 19(2), 268-286.

15. Retnawati, H. (2016). Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti Umum, Mahasiswa, dan Psikometrian). Yogyakarta: Parama Publishing.
16. Sari, P. K., Khaerani, S., & Noordyana, M. A. (2017). Penerapan Pendekatan Proses Pada Pembelajaran Tematik Siswa Sekolah Dasar. HOLISTIKA: Jurnal Ilmiah PGSD, 1(2), 109-112, ISSN: 2579-6151.
17. Sari, N. A., Akbar, A. & Yuniastuti. (2018). Penerapan Pembelajaran Tematik Terpadu di Sekolah Dasar. Jurnal Pendidikan, 3(12), 1572-1582, e-ISSN: 2502-417X.
18. Sudiyatno, Badrun, K. & Muhyadi. (2015). Pengembangan Model Assessment As Learning Pembelajaran Akutansi di SMK. Jurnal Penelitian dan Evaluasi Pendidikan, 19(2), 189-201, http://journal.uny.ac.id/index.php/jpep.
19. Sundari. (2016). Model Pengembangan Asesmen Kinerja (Performance Assessment) Mata Pelajaran IPA berbasis Nilai Karakter di SMP Kota Ternate Maluku Utara. Jurnal EduBio Tropika, 2(1), 121-186.
20. Yuniawan, T. (2014). Model Penilaian Kinerja dalam Pembelajaran Membaca Berbasis Teks Narasi Bermuatan Pendidikan Karakter Cinta Budaya. Jurnal Pendidikan Karakter, 5(1), 61-72.
21. Zainul, A. & Nasution, N. (2001). Penilaian Hasil Belajar. Jakarta: Dirjen Dikti.