TARBAWY
Indonesian Journal of Islamic Education
Vol. 11 No. 1 May 2024 (pp. 55 - 68) DOI: 10.17509/t.v11i1.58533
ISSN : 2580-6181 (Print), 2599-2481 (Online)
Journal homepage: https://ejournal.upi.edu/index.php/tarbawy/index

Analysis of Mid-Semester Assessment Items for Islamic Religious Education (IRE) in State Senior High Schools

Amalia Nurlitasari & Maemunah

UIN Sunan Kalijaga Yogyakarta, Indonesia

Corresponding author:
Amalia Nurlitasari, E-mail: : amaliapascasarjana@gmail.com

Abstract Teachers' knowledge and skills in conducting item analysis are still relatively low. Teachers still tend to create questions that are in accordance with core competencies and basic competencies without considering the quality of the test. This study aims to see the quality of question items and examples of questions that are feasible and those that are not feasible through the Mid-Semester Assessment (PTS) of class X Islamic Religious Education (PAI) at SMAN in the city of Yogyakarta. This study is a type of quantitative research and uses the JASP Program version 16.0. The results of the study show that Reliability is in the "High" category. The mean p value is highest in the "Medium" category of 56%. The highest D value is in the "Enough" category of 44% and the most distractions are in the "Acceptable" category of 128 out of 200 questions. Through this calculation, the characteristics of feasible and unfeasible questions can be seen.

Keywords: *Evaluation, Item Analysis, Islamic Religious Educator (IRE), Senior High School*

Abstrak. Pengetahuan dan keterampilan guru dalam melakukan analisis butir soal masih tergolong rendah. Guru masih cenderung membuat soal sesuai dengan kompetensi inti dan kompetensi dasar tanpa mempertimbangkan kualitas tes. Penelitian ini bertujuan melihat kualitas butir soal dan contoh soal yang layak dengan tidak layak melalui Penilaian Tengah Semester (PTS) kelas X mata pelajaran Pendidikan Agama Islam (PAI) di SMAN di kota Yogyakarta. Penelitian ini merupakan jenis penelitian kuantitatif dan menggunakan Program JASP versi 16.0. Adapun hasil penelitian menunjukan bahwa Reliabilitas pada kategori "Tinggi". Nilai mean p paling banyak pada kategori "Sedang" sebesar 56%. Nilai D paling banyak pada kategori "Cukup" sebesar 44% dan distractor paling banyak pada kategori "Diterima" sebesar 128 dari 200 soal. Melalui perhitungan tersebut maka dapat dilihat karakteristik soal yang layak dan tidak layak.

Kata Kunci*: Evaluasi, Analisis Butir, Pendidikam Agama Islam, Sekolah Menengah Atas*

Article history:

Received:
05 January 2024

Revised:
25 February 2024

Accepted:
10 April 2024

Available online:
06 May 2024

Introduction

The assessment of learning outcomes carried out by teachers has been explained in the Indonesian Ministry of Education and Culture Regulation No. 104 of 2014 concerning the Assessment of Learning Outcomes by Educators and Educational Units in Basic Education and Junior High Education Article 1, the assessment of learning outcomes by educators is the collection of information about students' learning achievements in the context of spiritual and social attitudes, knowledge, and skills obtained from learning (Ministry of Education and Culture Regulation No. 104, 2014). In measuring current learning outcomes, the learning curriculum designs assessments or assessments into three categories, namely diagnostic assessments, formal assessments, and summative assessments (Maulinda, 2022). The assessment of learning outcomes that is still in line or dominantly carried out until now is in the aspect of summative assessment (Anisah, 2022). Summative assessment is used as a reference to see the success of students' Teaching and Learning Activities (KBM) either after approximately 3 months of learning or referred to as the Mid-Semester Assessment (PTS) or after one learning semester referred to as the End of Year Assessment (PAT) (Barokah, 2019).

Seeing the importance of summative assessment in measuring student learning achievement, SMAN 10 Yogyakarta, which is one of the state schools in Yogyakarta, uses summative assessment as a process of assessing student learning outcomes. One of the assessments carried out is the Multiple Choice Assessment (PTS) which is carried out after 3 months of the teaching and learning process. The researcher conducted an analysis of the social items of the Multiple Choice Assessment (PTS) because when the research was conducted, SMA N 10 Yogyakarta would conduct a PTS in the form of multiple choices totaling 40 social items. Students are required to work on the PTS social exams created and compiled by a team of subject teachers. Therefore, the social tests made by teachers must have good quality and be able to provide accurate results as a means of carrying out the assessment process in order to know the abilities of students well (Himawan & Nurgiyantoro, 2022).

Ideally, the assessment prepared by the teacher is able to provide a lot of information about students' abilities (Setiawan et al., 2020). However, social problems arise when teachers compile social items, there are still many weaknesses in the compilation process so that they produce tests that do not have valid measuring strength. The instrument that does not have valid measuring strength will not provide any information regarding the abilities of students (Aripiani, 2014). Analysis of social item instruments is very important for teachers to do because sometimes social items are made too difficult or too easy, making it difficult for teachers to distinguish student characteristics from student abilities and success (Nur & Palobo, 2018).

In addition, teachers' knowledge and skills in conducting social item analysis are still lacking (Prihatin, 2022). Several previous research results regarding social item analysis show that teachers only create social items in accordance with core competencies and basic competencies without analyzing the quality of each social item item in detail as follows: First, from the results of research interviews, it was found that teachers have not been able to balance their abilities in compiling good thesis. Teachers only take it from textbooks, social banks, and the internet, then after it is finished, the teacher immediately tests it on students (Rokhyati, 2011). Second, teachers rarely conduct social item analysis because of their lack of understanding in analyzing each social

item. Teachers conduct social item analysis using manual methods through eixceil (Mia, 2014). Third, the information obtained through interviews was that the social PTS Geinap class V was tested by subject teachers without conducting qualitative or quantitative analysis (Ardianto, 2019).

This was also confirmed by the Deputy Head of PAI Curriculum at SMA N 10 Yogyakarta based on an interview on May 3, 2023 that in making social, teachers have carried out social studies before being distributed to students in the form of social information and social conformity with the grids or materials taught. However, teachers have not conducted an analysis of the social items used in making social PAT or PTS. The inability and lack of insight into this social item analysis causes teachers to be unable to obtain reciprocal information to differentiate student characteristics in terms of student abilities and success. Thus, it is certain that student success in teaching and learning activities certainly cannot be measured accurately.

In order for the assessment to be measured accurately, social item analysis is needed. Social item analysis is interpreted as a step in reviewing the question items of each social item to be tested in order to obtain quality social tests (Sudjana, 2016). In analyzing social items, there are two approaches in analyzing social test test data, namely using classical theory and using item reispoint theory (Nur & Palobo, 2018). Classical theory generally uses qualitative techniques through the examination of judgment to see the suitability between social items and learning materials in the curriculum which are then poured into the social grid. While the theory of item review uses quantitative techniques that contain social reliability, difficulty index, differentiation index, and distractor effectiveness (Rokhyati, 2011).

Choosing Islamic Religious Education as the subject to be studied is certainly not without reason. Islamic religious education in schools has existed since a person started going to school, both in kindergarten and college to teach various aspects such as the Qur'an, Hadith, Aqidah, Morals, Fiqh, and Tarih (Taufik & Halimah, 2019). Therefore, it is very important to conduct an analysis of social items at the upper elementary level that can distinguish between good and bad so that teachers get information about the extent of the religious abilities of students.

Through quantitative techniques, teachers will get feedback so that it can be used as a reference in making social decisions that will be tested. The results of the analysis of social items with quantitative techniques are related to reliability, difficulty index (ITK), differential strength index (IDB) and distractor effectiveness. The results of social reliability function to determine whether the PTS social class X can be used concurrently in different times (Hanifah, 2014). Reliability is seen based on the Alpha Croinbach score with a classification of 0.00-0.20 (very low); 0.21-0.40 (low); 0.41-0.70 (average); 0.71-0.90 (high); 0.91-1.00 (very high) (Arikuntoi, 2016). The number that indicates the difficulty or ease of a social is called the difficulty index (ITK). The difficulty index is derived from the JASP program calculation score with a classification of 0.00-0.30 (difficult); 0.31-0.70 (medium); and 0.71-1.00 (easy) (Arikunto, 2016). The results of the social difficulty level index are used to see the quality of the PTS social item items (Sudjana, 2016).

The differentiation strength index (IDB) is used to determine the differences in the abilities of students who have been able to master the material and those who have not mastered the material. can be declared to meet the requirements if the differentiation strength index is above 0.20 (Hanifah, 2014). Distraction effectiveness is a choice of distractors in each multiple-

choice social item. In the Final Assessment (PTS) of grade X PAI subjects at SMA N 10 Yogyakarta, 4 distractors and one key are used. In meeting the criteria for distractor effectiveness, 5% of the distractors are needed (Subali, 2014). After carrying out several social item analysis processes, the social items that fall into the feasible category can be used directly, social items in the less feasible category can be revised, while social items in the unfeasible category can be dropped so that they must be deleted or replaced with other social items (Mania et al., 2020).

Based on previous assessment evidence where teachers are more inclined to conduct social analysis through qualitative techniques, the researcher carries out research innovations by emphasizing more on quantitative techniques for Islamic Religious Education (PAI) subjects. Therefore, this research has a contribution to research in the field of Islamic Religious Education (PAI) which discusses the analysis of social items of the Middle Level Assessment so that the social quality used to obtain information about the characteristics of students can be known. using the JASP program and the results of this research study can be used as a reference for further research.

Methods

The research uses quantitative approach which is carried out through empirical analysis with quantitative descriptive methodology. According to Hadad Nawawi in (Rotama et al., 2020), quantitative descriptive analysis is a way of describing a research condition. Quantitative descriptive is used to describe the results of the calculation of social item analysis in the JASP program version 16.0 assisted by Microisoift Excel.

The data collection technique uses interviews with informant subjects, Deputy Head of Curriculum, Mrs. Mar'atul Allamah, S.Ag. and the social documentation of the Grade X Islamic Religious Education (PAI) Assessment (PTS) of 40 socials made by the subject teachers. However, during the analysis calculation, 20 socials were dropped because the value showed negative so that they did not meet the requirements. This study was located at SMA N 10 Yogyakarta, precisely at Jl. Gadeian No. 5, Ngupasan, Keic. Goindoimanan, Yogyakarta City.

The population in this study was the PTS social of the PAI subject of grade X. The sample used was the answers of grade X IPS 1 students totaling 36 students. The researcher chose class X IPS 1 because this class is a high or superior class.

Results and Discussion

Reliability Analysis

Social reliability is a very important thing in measurement. The consistency of the results obtained by the same person when retested even with different time conditions is a reference for the belief of the theist (Arikunto, 2016). According to Zainal Arifin, a social is considered to have high reliability if it provides consistency. The more reliable a social test is, the more the test results will have the same results when used again (Arifin, 2016). Reliability is seen based on the Alpha Croinbach score with a classification of 0.00-0.20 (very low); 0.20-0.40 (low); 0.40-0.70 (fair); 0.70-0.90 (high); 0.90-1.00 (very high) (Arikunto, 2016). The following are the results of the calculation of social reliability obtained through the JASP program:

Figure 1. Reliability Results

Frequentist Scale Reliability Statistics

Estimate	Cronbach's α	Guttman's λ2	Average interitem correlation
Point estimate	0.864	0.880	0.149
95% CI lower bound	0.787	0.733	0.069
95% CI upper bound	0.918	0.942	0.269

Note. Variables V9 and V10 correlated perfectly.

Based on Figure 1. The results of reliability through the JASP program, researchers use two formulas to see social reliability, namely Croinbach's with a value of 0.864 and Guttman with a value of 0.880. Both show results that are not too different and are both above 0.85. This means that the stability of the social test results according to the reliability classification can be said to be in high criteria because the Alpha value is 0.864. So that whenever the social test is used, it will produce relatively the same results. Good test items will have a high level of reliability, producing consistent results, therefore if an unreliable social is found, it can be immediately replaced with a new social.

Difficulty Level Index Analysis

Through the Difficulty Level Index (ITK), teachers are able to find out how easy or difficult the social is that is being tested on students (Burhan, 2020). The difficulty index in the JASP calculation results is seen through the Mean value. In this study, to find out the difficulty index, in addition to using JASP, it can also be done through the Microsoft Excel with the "If" formula. The results that appear also show the same results. According to Arikunto, the difficulty index used in this study is divided into three classifications as follows: 0.00-0.30 (difficult); 0.31-0.70 (medium); and 0.71-1.00 (easy). If a social item has a difficulty number of 0.00, then the social item is very difficult because no students answered correctly. Likewise, if a social item has a difficulty number of 1.00, then the social item is considered easy because all test participants were able to answer correctly. A good social item is one that is neither too easy nor too difficult. A social item that is too easy does not stimulate students to solve it. Conversely, if the social item is too difficult, it will cause students to become discouraged and not have the enthusiasm to try to do it again (Arikunto, 2016). The following are the results of the calculation of the Difficulty Level Index analysis using JASP:

Figure 2. Difficulty Level Index Results

Frequentist Individual Item Reliability Statistics

	If item dropped			
Item	**Cronbach's α**	**Guttman's λ2**	**Item-rest correlation**	**mean**
V1	0.868	0.884	0.010	0.222
V2	0.863	0.880	0.250	0.361
V3	0.866	0.882	0.076	0.167
V4	0.854	0.872	0.703	0.861
V5	0.856	0.874	0.550	0.722
V6	0.862	0.879	0.305	0.972
V7	0.858	0.875	0.557	0.917
V8	0.855	0.873	0.573	0.750
V9	0.859	0.877	0.435	0.889
V10	0.859	0.877	0.435	0.889
V11	0.859	0.877	0.420	0.889
V12	0.859	0.877	0.402	0.833
V13	0.859	0.876	0.506	0.917
V14	0.859	0.876	0.450	0.889
V15	0.857	0.874	0.529	0.833
V16	0.861	0.878	0.327	0.778
V17	0.860	0.877	0.379	0.611
V18	0.856	0.874	0.518	0.722
V19	0.863	0.880	0.243	0.806
V21	0.855	0.872	0.570	0.583
V22	0.865	0.881	0.120	0.861
V23	0.855	0.872	0.594	0.722
V24	0.853	0.871	0.639	0.639
V25	0.866	0.882	0.043	0.111
V26	0.863	0.880	0.278	0.639
V27	0.861	0.878	0.331	0.889
V28	0.865	0.881	0.115	0.111
V29	0.861	0.878	0.342	0.556
V30	0.862	0.879	0.278	0.806
V31	0.864	0.880	0.213	0.750
V32	0.860	0.876	0.386	0.639
V33	0.857	0.874	0.496	0.639
V34	0.868	0.884	0.054	0.333
V35	0.857	0.874	0.511	0.694
V36	0.858	0.875	0.557	0.917
V37	0.868	0.883	0.064	0.694
V38	0.869	0.884	0.020	0.639
V39	0.861	0.878	0.372	0.917
V40	0.861	0.878	0.403	0.944

Note. The following items were reverse scaled: V1, V22, V28, V37.

Based on Figure 2. The level of social difficulty of the Middle Level Assessment of Grade X of Islamic Religious Education Subjects can be seen in the Mean P table and then interpreted through the classification of the level of social difficulty. Judging from the classification of the level of social difficulty, those included in the social category "Difficult" are at numbers 3, 22, 25 or 8%. The social category "Medium" is at numbers 1, 2, 5, 8, 16, 17, 18, 19, 21, 23, 24, 26, 29, 30, 31, 32, 33, 34, 35, 37 and 38 or 56%. The category "Easy" is at number 4, 6, 9, 10, 11, 12, 13, 14, 15, 27, 28, 36, 39, 40 or about 36%. If social is difficult, it means the level of student ability is low, in other words, if social is tested on students with low ability, then the level of social difficulty will be high. On the other hand, if social is easy, it means that the level of student ability is high, in other words, the level of social difficulty will be high when social is tested on high-ability questions, both levels of social difficulty are defined as the proportion of students in the group who answer social questions correctly (Nurcahyo, 2017). Social that has a balanced level of difficulty index can be said to be good social (Arifin, 2016). This is further explained by Sudjana who said that, "The comparison between easy-medium-difficult social can be made 3-4-3, which means, 30% of social is categorized as easy, 40% of social is categorized as medium, and 30% of social is categorized as difficult". Based on the explanation, it can be seen that easy or difficult social does not mean it is not worth using. The social should be followed up by the social maker. Difficult social items are re-examined so that it can be known why the social is included in the easy or difficult category (Sudjana, 2016). The results of the analysis of social items of PTS grade X at SMA N 10

Yogyakarta showed that they had not yet reached 3-4-3. The difficult category received the least quota or did not meet 30% compared to the easy category which was above 30%. While for the medium category, it received a quota of up to 50%.

Social items that are difficult to answer by students can be caused by several things. A review of the answer key needs to be done. If the answer key is correct, the difficulty of the social item is possible because the material being asked has not been taught or the learning has not been completed, so that the minimum competency that must be mastered by students has not been achieved. Another factor that has an influence is that the material being measured is not appropriate to be asked using the social form that is being used and the social statements or sentences are too long (Rokhyati, 2011).

Differential Strength Index Analysis

According to Sudaryoinoi in Fitrianawati, the discriminatory strength of theistic items is the ability of a social item to distinguish between theistic students who have mastered the material being asked and theistic students who have not mastered the material being asked (Fitrianawati, 2017). The discriminatory strength index is able to distinguish which students are smart (upper polynomial) and which are less smart (lower polynomial) from the total score (Zainal Arifin, 2016). Smart students will answer more correctly, while students who are less smart will answer incorrectly. The number that shows the magnitude of the discriminatory strength is called the discrimination index or D (Anas Sudijoinoi, 2016). In the JASP program, discrimination strength can be seen through the Item-reist coirreilatioin. In this study, the researcher used the discrimination strength index according to Arikuntoi who classifies the social discrimination strength into five categories, namely 0.00-0.20 (Bad); 0.21-0.40 (Enough); 0.41-0.70 (Good); above 0.70 (Very Good). The discrimination index number ranges between Good discrimination strength indexes ranging from 0.40-0.70 in the category (Good) (Arikunto, 2016). In addition to using JASP to determine discrimination strength, this study also uses the help of Microisoif Excel with the "If" formula. Through Microisoift Eixceil, the analysis results show that it is easier to see which students are included in the upper and lower categories. Some experts use a prediction of 27% of each category (Anas Sudijoinoi, 2016). The results that appear also show the same results. The results of the IDB calculation using the JASP program are as follows:

Figure 3. Results of the Different Strength Index

Frequentist Individual Item Reliability Statistics

	If item dropped			
Item	Cronbach's α	Guttman's λ2	Item-rest correlation	mean
V1	0.868	0.884	0.010	0.222
V2	0.863	0.880	0.250	0.361
V3	0.866	0.882	0.076	0.167
V4	0.854	0.872	0.703	0.861
V5	0.856	0.874	0.550	0.722
V6	0.862	0.879	0.305	0.972
V7	0.858	0.875	0.557	0.917
V8	0.855	0.873	0.573	0.750
V9	0.859	0.877	0.435	0.889
V10	0.859	0.877	0.435	0.889
V11	0.859	0.877	0.420	0.889
V12	0.859	0.877	0.402	0.833
V13	0.859	0.876	0.506	0.917
V14	0.859	0.876	0.450	0.889
V15	0.857	0.874	0.529	0.833
V16	0.861	0.878	0.327	0.778
V17	0.860	0.877	0.379	0.611
V18	0.856	0.874	0.518	0.722
V19	0.863	0.880	0.243	0.806
V21	0.855	0.872	0.570	0.583
V22	0.865	0.881	0.120	0.861
V23	0.855	0.872	0.594	0.722
V24	0.853	0.871	0.639	0.639
V25	0.866	0.882	0.043	0.111
V26	0.863	0.880	0.278	0.639
V27	0.861	0.878	0.331	0.889
V28	0.865	0.881	0.115	0.111
V29	0.861	0.878	0.342	0.556
V30	0.862	0.879	0.278	0.806
V31	0.864	0.880	0.213	0.750
V32	0.860	0.876	0.386	0.639
V33	0.857	0.874	0.496	0.639
V34	0.868	0.884	0.054	0.333
V35	0.857	0.874	0.511	0.694
V36	0.858	0.875	0.557	0.917
V37	0.868	0.883	0.064	0.694
V38	0.869	0.884	0.020	0.639
V39	0.861	0.878	0.372	0.917
V40	0.861	0.878	0.403	0.944

Note. The following items were reverse scaled: V1, V22, V28, V37.

Based on Figure 3. The results of the differential strength of the Middle Level Assessment of Grade X of Islamic Religious Education Subjects can be seen in the Item-reist correlation table and then interpreted through the differential strength index classification. Judging from the differential strength classification, those included in the "Poor" social category are at numbers 3, 6, 9, 10, 22, 25, 28, 37, 38, 40 or 26%. The "Fair" social category is at numbers 1, 7, 11, 12, 13, 14, 16, 17, 18, 19, 26, 27, 30, 31, 34, 36, 39 or 44%. The category "Good" is in the number 2, 4, 5, 15, 29, 32, 35 or 18%. The category "Very Good" is in the number 8, 21, 23, 24, 33 or 13%. The most social is in the sufficient category. However, it is a concern for subject teachers as social creators to look back at the social characteristics of the bad category because the bad social perception is quite high, namely 26%.

Distractor Effectiveness Analysis

A good social must have a relatively homoigein question so that it is not easily guessed by students (Ratnaningsih et al., 2013). The question functions well if it has great appeal to students who do not master and understand the material. In fact, when the question is not chosen at all by students, it means that the question is stupid, too misleading (Sudaryoinoi, 2011). So in analyzing the first step that must be taken is to count the number of students who choose each answer option. Then use the formula = (Number of students who choose answer A, B, C, D, or Ei/total number of students who work) to calculate the level of distractor effectiveness/distractor effectiveness as in Figure 4.

Figure 4. Results of distractor effectiveness

NO	A	B	C	D	E
1	0,0%	19,4%	0,0%	72,2%	2,8%
2	2,8%	11,1%	16,7%	36,1%	22,2%
3	44,4%	8,3%	16,7%	16,7%	13,9%
4	86,1%	8,3%	0,0%	2,8%	2,8%
5	11,1%	11,1%	5,6%	63,9%	0,0%
6	0,0%	0,0%	91,7%	2,8%	0,0%
7	91,7%	2,8%	5,6%	0,0%	0,0%
8	8,3%	72,2%	5,6%	2,8%	5,6%
9	5,6%	0,0%	83,3%	2,8%	2,8%
10	88,9%	0,0%	5,6%	2,8%	2,8%
11	88,9%	5,6%	2,8%	0,0%	0,0%
12	5,6%	8,3%	80,6%	0,0%	2,8%
13	8,3%	88,9%	0,0%	0,0%	0,0%
14	88,9%	11,1%	0,0%	0,0%	0,0%
15	2,8%	0,0%	13,9%	75,0%	0,0%
16	77,8%	5,6%	11,1%	0,0%	5,6%
17	2,8%	16,7%	11,1%	55,6%	5,6%
18	2,8%	69,4%	22,2%	0,0%	2,8%
19	11,1%	0,0%	5,6%	2,8%	69,4%
20	13,9%	25,0%	27,8%	22,2%	5,6%
21	19,4%	2,8%	11,1%	52,8%	8,3%
22	30,6%	19,4%	11,1%	13,9%	22,2%
23	11,1%	5,6%	66,7%	11,1%	0,0%
24	16,7%	61,1%	8,3%	2,8%	8,3%
25	11,1%	5,6%	5,6%	55,6%	13,9%
26	2,8%	8,3%	58,3%	11,1%	13,9%
27	0,0%	0,0%	11,1%	80,6%	0,0%
28	88,9%	11,1%	0,0%	0,0%	0,0%
29	19,4%	52,8%	22,2%	2,8%	0,0%
30	0,0%	5,6%	11,1%	2,8%	72,2%
31	13,9%	2,8%	5,6%	2,8%	66,7%
32	8,3%	2,8%	58,3%	22,2%	2,8%
33	8,3%	0,0%	19,4%	55,6%	5,6%
34	33,3%	44,4%	8,3%	5,6%	2,8%
35	5,6%	13,9%	63,9%	11,1%	0,0%
36	5,6%	88,9%	0,0%	0,0%	2,8%
37	8,3%	19,4%	30,6%	19,4%	19,4%
38	11,1%	11,1%	13,9%	55,6%	0,0%
39	91,7%	2,8%	0,0%	5,6%	0,0%
40	0,0%	0,0%	5,6%	86,1%	0,0%

Based on Figure 4. The yellow coil is the answer key. The effectiveness of distractors is very much needed in social situations where students make mistakes in choosing answers. The distractors should be adjusted to the learning materials (Fitrianawati, 2017). The last step in analyzing the effectiveness of distractors is to use the "if" formula according to the predetermined index, namely D> 5% then accepted, 5%> D> 0% then revised, and D = 0% then rejected. The following are the results of the distractor effectiveness analysis using Microisoift Excel:

Figure 5. Distraction effectiveness description

NO	A	B	C	D	E
1	Ditolak	Diterima	Ditolak	Diterima	Revisi
2	Revisi	Diterima	Diterima	Diterima	Diterima
3	Diterima	Diterima	Diterima	Diterima	Diterima
4	Diterima	Diterima	Ditolak	Revisi	Revisi
5	Diterima	Diterima	Diterima	Diterima	Ditolak
6	Ditolak	Ditolak	Diterima	Revisi	Ditolak
7	Diterima	Revisi	Diterima	Ditolak	Ditolak
8	Diterima	Diterima	Diterima	Revisi	Diterima
9	Diterima	Ditolak	Diterima	Revisi	Revisi
10	Diterima	Ditolak	Diterima	Revisi	Revisi
11	Diterima	Diterima	Revisi	Ditolak	Ditolak
12	Diterima	Diterima	Diterima	Ditolak	Revisi
13	Diterima	Diterima	Ditolak	Ditolak	Ditolak
14	Diterima	Diterima	Ditolak	Ditolak	Ditolak
15	Revisi	Ditolak	Diterima	Diterima	Ditolak
16	Diterima	Diterima	Diterima	Ditolak	Diterima
17	Revisi	Diterima	Diterima	Diterima	Diterima
18	Revisi	Diterima	Diterima	Ditolak	Revisi
19	Diterima	Ditolak	Diterima	Revisi	Diterima
20	Diterima	Diterima	Diterima	Diterima	Diterima
21	Diterima	Revisi	Diterima	Diterima	Diterima
22	Diterima	Diterima	Diterima	Diterima	Diterima
23	Diterima	Diterima	Diterima	Diterima	Ditolak
24	Diterima	Diterima	Diterima	Revisi	Diterima
25	Diterima	Diterima	Diterima	Diterima	Diterima
26	Revisi	Diterima	Diterima	Diterima	Diterima
27	Ditolak	Ditolak	Diterima	Diterima	Ditolak
28	Diterima	Diterima	Ditolak	Ditolak	Ditolak
29	Diterima	Diterima	Diterima	Revisi	Ditolak
30	Ditolak	Diterima	Diterima	Revisi	Diterima
31	Diterima	Revisi	Diterima	Revisi	Diterima
32	Diterima	Revisi	Diterima	Diterima	Revisi
33	Diterima	Ditolak	Diterima	Diterima	Diterima
34	Diterima	Diterima	Diterima	Diterima	Revisi
35	Diterima	Diterima	Diterima	Diterima	Ditolak
36	Diterima	Diterima	Ditolak	Ditolak	Revisi
37	Diterima	Diterima	Diterima	Diterima	Diterima
38	Diterima	Diterima	Diterima	Diterima	Ditolak
39	Diterima	Revisi	Ditolak	Diterima	Ditolak
40	Ditolak	Ditolak	Diterima	Diterima	Ditolak

Based on Figure 5. Rejected distractors mean that no one chooses the answer choice so that it cannot function effectively as distractors. Revised distractors mean that they are at a level of less than 5% but above 0%, so the teacher should replace or re-evaluate the answer choices. Then if the distractors are accepted, it means that the distractor answer choices function well.

Based on the explanation, it can be concluded that distractors are answer choices that are not answer keys. Although the answer choices are wrong options, the wrong answer choices must be able to be considered according to their effectiveness, either as a disruptor, a deterrent, or a distractor for some examinees (Himawan & Nurgiyantoiro, 2022). Quantitative social item analysis can also be done using a commutative program. However, due to the limitations of the JASP program in analyzing the effectiveness of distractors, the researcher used Microisoift Excel to determine its effectiveness.

Example of PTS questions for class X at SMA N 10 Yogyakarta
with feasible and infeasible categories.

22.	Pernyataan dibawah ini yang **bukan** merupakan contoh dari maukuf alaih atau nazir adalah…
	A. Ibu ani menyerahkan sebidang tanah kepada bapak ali B. H. Hasan merupakan ketua panitia pembangunan masjid al ikhlash C. Iuran jariah dari jamaah untuk sarana majlis ta'lim diterima hj. Aisyah D. Sumbangan wakaf untuk klinik kesehatan diterima oleh h. Amir E. Panitia pembangunan madrasah tsanawiyah al-mahbubiyah

Based on the results of the social analysis, social noi. 22 in the Difficulty Level Index shows a figure of 0.14 which is included in the "difficult" category. The strength index is 0.12 which means it is difficult. The majority of the upper class who answered correctly were only 1 and the lower class who answered correctly were only 2 students, causing the social to be unable to distinguish students who understand the material with inappropriate wishes so that it enters the "unworthy" category. However, this means that the social has good distractor effectiveness on the answer choices. Distraction effectiveness on noi. 22 is in the "Accepted" category which shows that each answer choice has been chosen by students above 5%. If viewed from the social, then the possibility that causes the social is not feasible, the five answer choices use ambiguous word choices, do not use words that are easy to understand so that there is only distractor effectiveness but the ITK and IDB aspects are not fulfilled.

21.	Adanya wakif, nazir, mauquf bih, dan sighat merupakan…
	A. Hukum wakaf B. Sunnah wakaf C. Wajib wakaf D. Rukun wakaf E. Bentuk wakaf

Question no. 21 is said to be in the feasible category because the ITK analysis result of 0.72 is included in the "moderate" category. This means that the social is not too difficult and not too easy. The differentiation strength index of 0.80 is included in the "very good" category, which means that the social is able to differentiate students who understand the material from those who do not. However, in relation to the effectiveness of distractors, answer choices A, C, D, E are included in the "acceptable" category while option B indicates the "revised" category. The revised answer choices do not mean that they have to be discarded. Revision means that they can still be used, but teachers must re-evaluate the choice of words in question B that are in accordance with the scope of question choices A, C, D, and Ei.

Conclusion

Based on the analysis of social items in terms of Reliability, Difficulty Level Index (ITK), Strength Index (IDB), and Distraction Effectiveness, the following results were obtained:
Judging from the reliability classification, it can be said to be in the "High" criteria because the Alpha value is 0.864 so that whenever the social test is used, it will produce relatively the same results. Judging from the classification of the level of social difficulty, those included in the "Difficult" social category are at numbers 3, 22, 25 or 8%. The social category "Medium" is in numbers 1, 2, 5, 8, 16, 17, 18, 19, 21, 23, 24, 26, 29, 30, 31, 32, 33, 34, 35, 37 and 38 or about 56%. The category "Easy" is in numbers 4, 6, 9, 10, 11, 12, 13, 14, 15, 27, 28, 36, 39, 40 or about 36%. 3. Judging from the classification of strength, those included in the "Bad" social category are at numbers 3, 6, 9, 10, 22, 25, 28, 37, 38, 40 or approximately 26%. The "Fair" social category is at numbers 1, 7, 11, 12, 13, 14, 16, 17, 18, 19, 26, 27, 30, 31, 34, 36, 39 or approximately 44%. The "Good" category is at numbers 2, 4, 5, 15, 29, 32, 35 or approximately 18%. The category "Very Good" is in the category 8, 21, 23, 24, 33 or 13%. The effectiveness of distractors shows a total of 42 out of 200 answer choices fall into the category "rejected" and the rest are in the category accepted and revised. This is a concern for teachers to replace the rejected social so that the rejected word choices in a number can be adjusted to the appropriate context. Questions that have entered the category "revised" as many as 30 out of 200 can be re-examined and those that have entered the category "accepted" as many as 128 out of 200 can be maintained.

References

Anas Sudijono. (2016). Anas Sudijono, Pengantar Evaluasi Pendidikan. *RajaGrafindo Persada*, *Cet. Ke-16*.

Anisah, G. (2022). Kerangka konsep assessment of learning, assessment for learning, dan assessment as learning serta penerapannya pada pembelajaran. *Al-aufa: jurnal pendidikan dan kajian keislaman*, *3*(2). https://doi.org/10.32665/alaufa.v3i2.1201

Ardianto, N. (2019). *Analisis butir soal penilaian tengah semester genap kelas v tema 6 panas dan perpindahannya sd dabin iii kecamatan tegal barat kota tegal.* Universitas Negeri Semarang.

Arikunto, S. (2016). *Dasar-dasar evaluasi pendidikan* (Revisi, Cet. Ke-8). Bumi Aksara.

Aripiani, T. D. (2014). *Analisis Alat Penilaian Pembelajaran Ujian Akhir Semester Mata Pelajaran Produktif Keahlian Kelas X Tata Boga di SMK Shandy Putra Bandung*. Universitas Pendidikan Indonesia.

Barokah, M. (2019). Manajemen penilaian sumatif pada ranah kognitif pembelajaran PAI kelas X semester ganjil di SMA Negri 2 Pontianak tahun pelajaran 2017/ 2018. *Al-Idarah: Jurnal Kependidikan Islam*, *9*(2).

Burhan, N. (2020). Penilaian Pembelajaran Bahasa Berbasis Kompetensi. *Didaktika: Jurnal Kependidikan*, *9*(1).

Fitrianawati, M. (2017). Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru Dan Hasil Belajar Peserta Didik. *JPT : Jurnal Pendidikan Tematik*, *2*(3).

Hanifah, N. (2014). Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi. *SOSIO E-KONS*, *6*(1).

Himawan, R., & Nurgiyantoro, B. (2022). Analisis butir soal latihan penilaian akhir semester ganjil mata pelajaran bahasa Indonesia kelas VIII SMPN 1 Bambanglipuro Bantul menggunakan program ITEMAN. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, Dan Pengajarannya*, *8*(1).

Mania, S., Fitriani, F., Majid, A. F., Ichiana, N. N., & Abrar, A. I. P. (2020). Analisis Butir Soal Ujian Akhir Sekolah. *Al Asma : Journal of Islamic Education*, *2*(2). https://doi.org/10.24252/asma.v2i2.16569

Maulinda, U. (2022). Pengembangan Modul Ajar Berbasis Kurikulum Merdeka. *Tarbawi*, *5*(2).

Mia, A. I. (2014). *Analisis butir soal dalam buku ajar ilmu pengetahuan alam kelas v di madrasah ibtidaiyah al maarif 02 Palang Sukorejo Pasuruan*. UIN Maulana Malik Ibrahim.

Nur, A. S., & Palobo, M. (2018). Pelatihan Analisis Butir Soal Berbasis Komputerisasi Pada Guru SD. *MATAPPA: Jurnal Pengabdian Kepada Masyarakat*, *1*(1). https://doi.org/10.31100/matappa.v1i1.79

Nurcahyo, F. A. (2017). Aplikasi IRT dalam Analisis Aitem Tes Kognitif. *Buletin Psikologi*, *24*(2). https://doi.org/10.22146/buletinpsikologi.25218

Permendikbud No 104. (2014). Salinan Lampiran Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 104 Tahun 2014 Tentang Penilaian Hasil Belajar oleh Pendidik Pada Pendidikan Dasar dan Pendidikan Menengah. *Pedoman Evaluasi Kurikulum*.

Prihatin, R. P. (2022). *Analisis Soal SKI Kelas X MAN 2 Bantul Yogyakarta Pada Penilaian Akhir Tahun 2020/2021*. UIN Sunan Kalijaga .

Ratnaningsih, D. J., Isfarudi, I., & Soleiman, N. (2013). Analisis Butir Soal Pilihan Ganda Ujian Akhir Semester Mahasiswa Di Universitas Terbuka Dengan Pendekatan Teori Tes Klasik. *Jurnal Pendidikan Terbuka Dan Jarak Jauh*, *12*(2).

Rokhyati. (2011). *Karakteristik Secara Kualitatif Dan Kuantitatif Soal Ulangan Akhir Semester Genap Bahasa Indonesia Kelas Xii Sma Negeri Di Kabupaten Purbalingga Tahun Pelajaran 2010/2011*. Universitas Negeri Yogyakarta.

Rotama, A. D., Budiutomo, T. W., & Bowo, A. N. A. (2020). Analisis Butir Soal Penilaian Tengah Semester Mata Pelajaran PPKN Kelas VII di SMP Muhammadiyah 7 Yogyakarta. *Academy of Education Journal*, *11*(01). https://doi.org/10.47200/aoej.v11i01.314

Setiawan, M. A., Susongko, P., & Hayati, M. N. (2020). Pendeteksian DIF pada Perangkat Tes Objektif Penilaian Akhir Semester IPA dengan Mengunakan Permodelan Rasch. *PSEJ (Pancasakti Science Education Journal)*, *5*(2).

Subali, B. (2014). Analisis Soal Baik Kualitatif Maupun Kuantitatif. *Workshop Item Development Dosen Poltekes Kebidanan Politeknik Kesehatan Surakarta*, *5*.

Sudaryono. (2011). Implementasi Teori Responsi Butir (Item Response Theory) pada Penilaian Hasil Belajar Akhir di Sekolah. *Jurnal Pendidikan Dan Kebudayaan*, *17*(6).

Sudjana, N. (2016). Penilaian Hasil Proses Belajar Mengajar. In *Penilaian dan Hasil Belajar Mengajar*.

Taufik, A., & Halimah, I. (2019). Buku Guru Pendidikan Agama Islam dan Budi Pekerti Kelas X SMA/SMK. In *Direktorat Jenderal Pendidikan Agama Islam Kementerian Agama Indonesia*.

Zainal Arifin. (2016). Evaluasi Pembelajaran. *Jurnal Pendidikan*, *03*.