Jurnal Pendidikan Agama Islam
ISSN: 1829-5746 | EISSN: 2502-2075
Vol. 18, No. 1, Juni 2021
http://ejournal.uin-suka.ac.id/tarbiyah/index.php/jpai/index

Analysis of Items with Item Response Theory (IRT) Approach on Final Assessment for Al-Quran Hadith Subjects

Ahmad Syafii
Yogyakarta State University, Yogyakarta, Indonesia
Email: ahmadsyafii.2020@student.uny.ac.id
Haryanto
Yogyakarta State University, Yogyakarta, Indonesia
Email: haryanto@uny.ac.id
Iqbal Faza Ahmad
Yogyakarta State University, Yogyakarta, Indonesia
Email: iqbalfaza.2020@student.uny.ac.id
Arifah Fauziah
Korea National University of Education, South Korea
Email: arifahfauziah@knue.ac.kr

DOI: https://doi.org/10.14421/jpai.2021.181-09

Abstract

Measurement in the field of education, especially the teaching and learning process can be done with measuring tools in the form of tests and non-tests. Islamic Religious Education is considered the same as other subjects. to realize student success is also measured through evaluation, which is a systematic process to obtain information about the effectiveness of teaching and learning activities. In addition, it can also assist teachers in achieving learning objectives and describe student achievement in accordance with predetermined criteria. This study aims to analyze the Year-End Assessment of MAN 2 Bantul for the 2020/2021 academic year consisting of 25 multiple choice questions. The items analyzed in the report have a total number of responses of 265 students. Item analysis using analysis with modern test theory (Item Response Theory). The results showed that the results of the model fit test showed that the instrument fit on the 2 PL model (Logistics parameter) with the lowest AIC value, namely 4874.85. The results of the parameter analysis of the level of difficulty indicate that there are 4 questions that are categorized as very easy, 4 easy questions, 15 moderate questions, 1 difficult question, and 1 very difficult question. This shows that the distribution of the difficulty level parameters is quite balanced. The results of the analysis of the different power parameters show that there are 23 good questions, 1 fairly good question, and 1 bad question. This shows that the different power level parameters are quite good. The results of the estimation of students' abilities with the MLE estimator showed that there were no students who had abilities below -4. There are 15 students with abilities above 2.00. There are 165 students who fall into the good category (ability -2 to 2 and 9 students who have abilities below -2. Based on the plot of the information function, it can be concluded that the optimal test if given to individuals with low abilities is around -1.2. Accurate questions to measure students' abilities with a range of -2.5 to 1.2.

Keywords: *Item Response Theory (IRT), Al-quran Hadith, Madrasah*

Received: 25 June 2021 | Revised: 30 June 2021 | Accepted: 30 June 2021
Pages: 167-194

DOI: https://doi.org/10.14421/jpai.2021.181-09

Abstrak

Pengukuran dalam bidang pendidikan khususnya proses belajar mengajar dapat dilakukan dengan alat ukur berupa tes maupun non tes. Pendidikan Agama Islam dinilai sama dengan mata pelajaran lainnya. untuk menyadari Keberhasilan siswa juga diukur melalui evaluasi, yaitu suatu proses sistematis untuk memperoleh informasi tentang efektivitas kegiatan belajar mengajar. Selain itu juga dapat membantu guru dalam mencapai tujuan pembelajaran dan menggambarkan prestasi siswa sesuai dengan kriteria yang telah ditetapkan. Penelitian ini bertujuan untuk menganalisis soal Penilaian Akhir Tahun MAN 2 Bantul Tahun Ajaran 2020/2021 terdiri dari 25 butir soal pilihan ganda. Butir soal yang dianalisis dalam pelaporan memiliki jumlah respons sebanyak 265 peserta didik. Analisis butir soal menggunakan analisis dengan teori tes modern (Item Respon Theory). Hasil penelitian menunjukkan bahwa hasil uji fit model menunjukkan instrumen soal fit pada model 2 PL (parameter Logistik) dengan nilai AIC terendah yaitu 4874,85. Hasil analisis parameter tingkat kesukaran menunjukkan bahwa terdapat 4 soal yang berkategori sangat mudah, 4 soal mudah, 15 soal sedang, 1 soal sukar, dan 1 soal sangat sukar. Hal tersebut menunjukkan ketersebaran parameter tingkat kesukaran sudah cukup berimbang. Hasil analisis parameter daya beda menunjukkan bahwa terdapat 23 soal baik, 1 soal cukup baik, dan 1 soal tidak baik. Hal tersebut menunjukkan parameter tingkat daya beda sudah cukup baik. Hasil estimasi kemampuan siswa dengan estimator MLE menunjukkan bahwa tidak ada siswa yang memiliki kemampuan di bawah -4. Terdapat 15 siswa dengan kemampuan di atas 2,00. Ada 165 siswa yang masuk dalam kategori baik (kemampuan -2 hingga 2 dan 9 siswa yang memiliki kemampuan di bawah -2. Berdasar plot fungsi informasi dapat ditarik kesimpulan bahwa tes optimal jika diberikan kepada individu dengan kemampuan rendah sekitar -1,2. Soal akurat untuk mengukur kemampuan siswa dengan range -2,5 hingga 1,2.
Kata kunci: Item Respons Theory (IRT), Al-quran Hadits, Madrasah

Introduction

A test is a task, or a series of tasks designed to obtain information about educational and psychological traits or attributes. The ability of students in the world of education can be determined through a series of tests (Parmaningsih and Saputro 2021) . The test is part of the assessment process. Assessment is the process of collecting some informations about students and classes for instructional decision-making purposes (Istiyono 2020) . Another related term is measurement. Measurement is the determination of a number against a certain characteristic possessed by certain people, objects or objects according to clear rules or formulations. The characteristics of measurement are the use of numbers or certain rules or formulas (Parmaningsih and Saputro 2021) . The relationship between tests, measurements, and assessments shows that the assessment of learning outcomes can only be carried out properly and correctly when using the information obtained through measuring learning outcomes

DOI: https://doi.org/10.14421/jpai.2021.181-09

(Matondang, Jalinus, and Ambiyar 2020) . The quality of the test items must meet the appropriate requirements (Mulyani, Efendi, and Ramalis 2021) . This is important because the test results are used as the basis for determining the "decision" of the examinees, for example passing or not passing, being accepted or rejected.

In conducting an assessment, an instrument is needed that can measure students' abilities accurately. In the case of summative or end of semester exams, it is usually done with multiple choice. The process of assigning a number to something or someone based on a rule is called measurement. The activity of systematically determining numbers for an object is defined as a measurement by Mardapi (Mardapi 2012) . Results must be as accurate as possible. Although measurement does not produce a value or determine whether something is good or bad, the results of these measurements can be used to make decisions.

There are at least two forms of assessment in educational practice, namely formative assessment and summative assessment. In context *assessment for learning*, summative assessment is important to do. If the learning process is complete, focus on the summative to describe the quality of students. The summative assessment at the end of the lesson focuses on learning and achieving the desired outcome. Summative assessment is used to track students' progress in the program by providing a summary of their experience.

Measurement in the field of education, especially the teaching and learning process can be done with measuring tools in the form of tests and non-tests. The measuring instrument or test instrument commonly used in evaluating student learning outcomes is a set of questions. A set of questions used to measure student learning outcomes must have good quality in order to measure the actual abilities of students (Sarea and Hadi 2015) .

It can be said that the measurement is quantitative, to determine the value quantitatively a measuring instrument is needed, one of which is a test. The measuring instrument or test instrument commonly used in evaluating student learning outcomes is a set of questions.

Measuring tools that can be used in the process of evaluating the learning process can be in the form of homework assignments, quizzes, midterm exams (UTS), and final semester exams (UAS). The test is one form of instrument used to measure a number of questions that have true or false

DOI: https://doi.org/10.14421/jpai.2021.181-09

answers, or all true or partially true with the aim of knowing the learning achievements or competencies that have been achieved by students in certain fields (Mardapi 2012) . In this report, the tests in the field of PAI are explained.

Summative Assessment or also known as PAS (Semester Final Assessment) is sometimes made too difficult or too easy, making it difficult for educators to distinguish students' abilities. Therefore, it is necessary to test/analyze the test questions in the hope that the results obtained accurately reflect the students' abilities (Istiyono 2020) .

In the analysist development test items is divided into 2 forms, those are classic test theory (CTT) and item response theory (IRT). IRT is the latest analytical theory that incorporates item analysis. This modern item analysis theory was created by academics to overcome the shortcomings of the classical item analysis theory, therefore it can be said that this IRT model is a complement to the classical item analysis theory in the form of item response theory (Classical Test Theory). Item response theory (IRT) is a collection of statistical models that have been used to model responses to educational and psychological test items along with the latent traits that determine how individuals respond to those items. The theoretical and computational framework of IRT was developed from the 1950s to the 1980s and since then has been widely used by organizational researchers for various applications and research domains (Foster, Min, and Zickar 2017) . Balsis shows IRT has been used in analyzing Alzheimer's disease that affect neurological, cognitive, and behavioral processes (Balsis et al. 2018) . In other research McGrory shows IRT has been used in research on dementia (McGrory et al. 2014) . In the development of computing IRT can be used either using the Stata . application (Linden 2018) , as well as Software IRTEQ, STUIRT, and POLYEQUATE (Malatesta and Lee 2019) .

Studies on IRT are dominated by scientific studies. IRT is widely used in the study of mathematics (Anisa 2013; Halijah 2018; Kurniawan 2015; Kurniawati, Susongko, and Paridjo 2017; Sainuddin 2018) . IRT is also widely used in science studies (Arlinwibowo, Hadi, and Firdaus 2020; Ishak 2019; Sugiarto 2018) . In addition, IRT is also widely used in physics (Abdullah, Ramalis, and ... 2019; Alifa, Ramalis, and Purwana 2018; Andayani, Purwanto, and Ramalis 2019; Darmawan, Tarigan, and Wijaya nd; Fitriani, Ramalis, and Efendi 2019; Rahman 2021) .

DOI: https://doi.org/10.14421/jpai.2021.181-09

Islamic Religious Education is considered the same as other subjects. In realizing the student success is also measured through evaluation, which is a systematic process to obtain information about the effectiveness of teaching and learning activities. In addition, it can also assist teachers in achieving learning objectives and describe student achievement in accordance with predetermined criteria (Azizah and Zainudin 2020) .

Islamic Religious Education in its development seeks to develop a Muslim personality that is combined in terms of spiritual, physical, spiritual, emotional, intellectual, and social, in order to instill noble character, and can make a complete human being. In addition, it also tries to make humans as caliphs fil al-ardh, responsible, and continue to serve Allah, especially in the face of difficulties. Therefore, Islamic religious education must be taught in its entirety. However, Islamic religious education subjects in schools are still less attractive to students, and are still less successful in instilling positive behavior in students, due to monotonous learning. The low quality of learning is caused by the weak methodological aspects mastered by the teacher. In addition to teaching, teachers have prepared lesson plans and assessments in their activities, even the teacher evaluation planning has not been given much attention in its implementation. This is very influential on the success of student learning. Not to mention the questions that are made do not measure the aspects that have been planned in the RPP (Sham 2019) .

A test is said to be good as a measuring tool if it meets the test requirements, namely having: validity, reliability, objectivity, practicality and economy. Therefore, this study will examine the feasibility of tests in the field of PAI with the following competencies:

Table 1. KI KD

Odd semester	
Core Competencies	**Basic competencies**
1. Living and practice the teachings of their religion	1.1. Live the authenticity of the Qur'an as a revelation of Allah 1.2. Believing in the Koran as a guide for life 1.3. Evacuate the Qur'an appropriately and correctly in daily life. 1.4. Believing in the truth of the values contained in the main contents of the Qur'an 1.5. Charity according to the content of Surat al-Mu'minuun: 12-14; Surah al-Nahl: 78; Surah al-

DOI: https://doi.org/10.14421/jpai.2021.181-09

		Baqarah:30-32; and Surah adz-Dzaariyat: 56 (in everyday life).
2.	Living and practice honest, disciplined, responsible, caring behavior (mutual cooperation, cooperation, tolerance, peace) polite, responsive and pro-active and show attitudes as part of the solution to various problems in interacting effectively with the social and natural environment as well as in positioning themselves as a reflection of the nation in the world association.	2.1. Demonstrate a firm attitude to practice the teachings of the Koran 2.2. Demonstrate careful behavior towards syar'i arguments as the implementation of learning about the evidence of the authenticity of the Qur'an 2.3. Demonstrate behaviors that practice the teachings of the Qur'an 2.4. Demonstrate behavior that makes the Qur'an a source of law in everyday life 2.5. Having an attitude that reflects the function of humans both as servants of Allah and His caliph on earth as contained in the Surah *al-Mu'minuun* :12-14; Letter *al-Nahl* :78; Letter *al-Baqarah* :30-32; and Letters *adz-Dzaariyat* : 56.
3.	Understanding, apply, analyze factual, conceptual, procedural knowledge based on their curiosity about science, technology, art, culture, and humanities with insight into humanity, nationality, state, and civilization related to the causes of phenomena and events, as well as apply procedural knowledge in specific fields of study according to their talents and interests to solve problems	3.1. Understanding the meaning of the Qur'an according to the scholars' 3.2. Understanding the proof of the authenticity of the Qur'an 3.3. Understand the purpose and function of the Qur'an. 3.4. Understanding the main contents of the Qur'an 3.5. Understanding the verses of the Qur'an about humans and their duties as servants of Allah and caliphs on earth in QS *al-Mu'minuun* :12-14; QS *al-Nahl* :78; QS *al-Baqarah* :30-32; and QS *adz-Dzaariyat* : 56

4. Processing, reasoning, and presenting in the concrete and abstract realms related to the development of what he learned in school independently, and being able to use methods according to scientific rules	4.1. Presenting the meaning of the Qur'an conveyed by the scholars 4.2. Show examples of proofs of the authenticity of the Qur'an 4.3. Tells the story of the person who made the Qur'an according to its purpose and function 4.4. Explaining the main points of the teachings of the Qur'an along with examples in the verse 4.5. Demonstrating the memorization and meaning per word of the verses of the Qur'an about humans and their duties as servants of Allah and caliphs on earth in Surah *al-Mu'minuun* :12-14; Letter *al-Nahl* :78; Letter *al-Baqarah* :30-32; and Letters *adz-Dzaariyat* : 56.
Even semester	
Competence core	Basic competencies
1. Living and practice the teachings of their religion	1.1 Believing in authentic and hasan hadiths as the legal basis for Islamic teachings 1.2 Stick to the authentic and hasan hadith as a guide in everyday life. 1.3 Appreciate the truth values contained in the authentic hadith 1.4 Believing in the diversity of hadith and guiding the hadith selectively
2. Living and practice honest, disciplined, responsible, caring behavior (mutual cooperation, cooperation, tolerance, peace) polite, responsive and pro-active and show attitudes as part of the solution to various problems in interacting effectively with the social and natural environment as well as in positioning themselves as a	2.1. Demonstrate a critical attitude in practicing hadith as a basis in everyday life 2.2. Demonstrate behavior that uses hadith as its function has been studied 2.3. Show selective behavior towards the diversity of hadith had 2.4. Have a sincere attitude in worship as an implementation of understanding the verses of the Qur'an about sincerity in worship in Surah al-An'aam: 162-163; Surah al-Bayyinah: 5 and the hadith narrated by Bukhari from Aisha Ra. انَيَقُومُمِنَاللَّيْلِحَتَّىتَتَفَطَّرَقَدَمَاهُفَقَالَتْعَائِشَةُلِمَتَصْنَعُهَذَايَارَسُولَاللَّهِوَقَدْغَفَرَاللَّهُلَكَمَاتَقَدَّمَمِنْذَنْبِكَوَمَاتَأَخَّرَقَالَأَفَلَاأُحِبُّأَنْأَكُونَعَبْدًاشَكُورًافَلَمَّاكَثُرَلَحْمُهُصَلَّىجَالِسًافَإِذَاأَرَادَأَنْيَرْكَعَقَامَقَفَرَأَثُمَّرَكَعَ

DOI: https://doi.org/10.14421/jpai.2021.181-09

	reflection of the nation in the world association.	
3.	Understanding, apply, analyze factual, conceptual, procedural knowledge based on their curiosity about science, technology, art, culture, and humanities with insight into humanity, nationality, state, and civilization related to the causes of phenomena and events, as well as apply procedural knowledge in specific fields of study according to their talents and interests to solve problems	3.1. Understanding the meaning of hadith, sunnah, khabar, and atsar 3.2. Understanding the elements of hadith 3.3. Identify the kinds of sunnah (qauliyah, fi'liyah, taqriiriyah, and hammiyah) and their function on the Qur'an 3.4. Understand the distribution of hadith in terms of quantity and quality. 3.5. Understanding the verses of the Qur'an about sincerity in worship in QS al-An'aam: 162-163; QS al-Bayyinah: 5 and the hadith narrated by Bukhari from Aisha Ra. انَيَقُومُمِنَاللَّيْلِحَتَّىتَتَفَطَّرَقَدَمَاهُفَقَالَتْعَائِشَةُلِمَتَصْنَعُهَذَايَارَسُولَاللَّهِوَقَدْغَفَرَاللَّهُلَكَمَاتَقَدَّمَمِنْذَنْبِكَوَمَاتَأَخَّرَقَالَأَفَلَاأُحِبُّأَنْأَكُونَعَبْدًاشَكُورًافَلَمَّاكَثُرَلَحْمُهُصَلَّىجَالِسًافَإِذَاأَرَادَأَنْيَرْكَعَقَامَفَقَرَأَثُمَّرَكَعَ
4.	Processing, reasoning, and presenting in the concrete and abstract realms related to the development of what he learned in school independently, and being able to use methods according to scientific rules	4.1. Describe the substance of differences and similarities in the understanding of hadith, sunnah, khabar, and atsar 4.2. Presenting the elements of hadith 4.3. Presenting examples of various sunnah (qauliyah, fi'liyah, taqririyah, and hammiyah) 4.4. Presenting the distribution of hadith in terms of quantity and quality. 4.5. Demonstrating memorization and meaning per word of the verses of the Qur'an about sincerity in worship in Surah al-An'aam: 162-163; Surah al-Bayyinah: 5 and the hadith narrated by Bukhari from Aisha ra.

DOI: https://doi.org/10.14421/jpai.2021.181-09

انَيَقُومُمِنَاللَّيْلِحَتَّىتَتَفَطَّرَقَدَمَاهُفَقَالَتْعَائِشَةُلِمَتَصْنَعُهَذَايَارَسُولَاللَّهِوَقَدْغَفَرَاللَّهُلَكَمَاتَقَدَّ

مِنْذَنْبِكَوَمَاتَأَخَّرَقَالَأَفَلَاأُحِبُّأَنْأَكُونَعَبْدًاشَكُورًافَلَمَّاكَثُرَلَحْمُهُصَلَّىجَالِسًافَإِذَاأَرَادَ

أَنْيَرْكَعَقَامَفَقَرَأَثُمَّرَكَعَ

	Core Competencies	Basic competencies
2.	Living and practice the teachings of their religion	1.1. Living the authenticity of the Qur'an as a revelation of Allah 1.2. Believing in the Qur'an as a guide for life 1.3. Functioning the Qur'an correctly and correctly in everyday life. 1.4. Believing in the truth of the values contained in the main contents of the Qur'an 1.5. Charity according to the content of Surat al-Mu'minuun: 12-14; Surah al-Nahl: 78; Surah al-Baqarah:30-32; and Surah adz-Dzaariyat: 56 (in everyday life).
3.	Living and practice honest, disciplined, responsible, caring behavior (mutual cooperation, cooperation, tolerance, peace) polite, responsive and pro-active and show attitudes as part of the solution to various problems in interacting effectively with the social and natural environment as well as in positioning themselves as a reflection of the nation in the world association.	2.6. Demonstrate a firm attitude to practice the teachings of the Koran 2.7. Demonstrate careful behavior towards syar'i arguments as an implementation of learning about the evidence of the authenticity of the Qur'a 2.8. Demonstrate behaviors that practice the teachings of the Qur'an 2.9. Demonstrate behavior that makes the Qur'an a source of law in everyday life 2.10. Having an attitude that reflects the function of humans both as servants of Allah and His caliph on earth as contained in the Surah *al-Mu'minuun* :12-14; Letter *al-Nahl* :78; Letter *al-Baqarah* :30-32; and Letters *adz-Dzaariyat* : 56.
5.	Understanding, apply, analyze factual, conceptual, procedural knowledge based on	3.6. Understanding the meaning of the Qur'an according to the scholars' 3.7. Understanding the proof of the authenticity of the Qur'an

DOI: https://doi.org/10.14421/jpai.2021.181-09

	their curiosity about science, technology, art, culture, and humanities with insight into humanity, nationality, state, and civilization related to the causes of phenomena and events, as well as apply procedural knowledge in specific fields of study according to their talents and interests to solve problems	3.8. Understand the purpose and function of the Qur'an. 3.9. Understanding the main contents of the Qur'an 3.10. Understanding the verses of the Qur'an about humans and their duties as servants of Allah and caliphs on earth in QS *al-Mu'minuun* :12-14; QS *al-Nahl* :78; QS *al-Baqarah* :30-32; and QS *adz-Dzaariyat* : 56
6.	Processing, reasoning, and presenting in the concrete and abstract realms related to the development of what he learned in school independently, and being able to use methods according to scientific rules	4.1. Presenting the meaning of the Qur'an conveyed by the scholars 4.2. Show examples of proofs of the authenticity of the Qur'an 4.3. Tells the story of the person who made the Qur'an according to its purpose and function 4.4. Explaining the main points of the teachings of the Qur'an along with examples in the verse 4.5. Demonstrating the memorization and meaning per word of the verses of the Qur'an about humans and their duties as servants of Allah and caliphs on earth in Surah *al-Mu'minuun* :12-14; Letter *al-Nahl* :78; Letter *al-Baqarah* :30-32; and Letters *adz-Dzaariyat* : 56.

In the study of PAI there is no research that analyzes test items using IRT analysis. The research carried out is still using CTT analysis at the SMA, SMP, and SD levels (Ahmad and Sukiman 2019; Rahmaini and Taufiq 2018; Rusmayani 2020; Sarea 2018) . For this reason, it is important to conduct a study on IRT analysis on the test items used in PAI subjects.

This study will test the test items in the Year-End Assessment of PAI subjects at the MAN 2 Bantul school for the 2020/2021 academic year. The question consists of 25 multiple choice questions. The items analyzed in the

report have a total number of responses of 265 students. Item analysis using analysis with modern test theory (*Item Response Theory*).

Research Method

The research subjects were students of class MAN 2 Bantul for the academic year 2020/2021 with a total of 188 students. The object of research is a set of questions for the Even Semester Final Exam of MAN 2 Bantul for the 2020/2021 academic year in the Al-Quran Hadith Subject which consists of 25 multiple choice questions. The data used is the results of the final semester exam test results for 10th grade students of MAN 2 Bantul in the subject of Al-Quran Hadith with a total of 25 questions. The test results are then scored dichotomously where the correct answer is given a score of 1 and the wrong answer is given a score of 0.

The scoring data were analyzed with three types of parameters, namely the 1 PL, 2 PL, and 3 PL models and then concluded the fit items for the 1 PL, 2 PL, and 3 PL parameters. Then, the model fit test was carried out. The purpose of the model fit test is to find out which of the four models is the best for estimating item parameters and the ability of subjects with IRT on the test items. In this study, the model fit test was conducted by comparing the AIC (*Akaike Information Criterion*) values of all models. The analysis was carried out using the R program. The most suitable model is the one with the lowest AIC (Snipes and Taylor 2014) . After being fit with one model, the parameters and item abilities were estimated using the MLE (Maximum Likelihood) method. After that, the estimation results are interpreted. The value of the item information function is then calculated to determine the reliability of the item.

The R syntax used in the analysis is as follows:

```
library(ltm)
library(foreign)
library(erm)
dikotomus <- read.spss(file="dikotomus2.sav",to.data.frame=true)
#dichotomous <- dichotomous[,-11:-40]

#analysis 1 pl
mml1<- rasch(dichotomous)
pl1 <- data.frame(coef.rasch(mml1))
xlsx::write.xlsx(pl1,'1. 1pl.xlsx')
factor.scores(mml1,method="eap")
jpeg("1. 1pl.jpg", width = 600, height = 350)
```

DOI: https://doi.org/10.14421/jpai.2021.181-09

```
plot(mml1)
dev.off()

#analysis 2 pl
mml2 <- ltm(dichotomous~z1)
pl2 <- data.frame(coef.ltm(mml2))
xlsx::write.xlsx(pl2,'2. pl2.xlsx')
factor.scores(mml2,method="eap")
jpeg("2. 2pl.jpg", width = 600, height = 350)
plot(mml2)
dev.off()

#analysis 3 pl
mml3 <- tpm(dichotomous,start.val="random")
pl3 <- data.frame(coef.tpm(mml3))
xlsx::write.xlsx(pl3,'3. pl3.xlsx')
factor.scores(mml3,method="eap")
jpeg("3. 3pl.jpg", width = 600, height = 350)
plot(mml3)
dev.off()

#model fit test
anova(mml1,mml2)
anova(mml1,mml3)
anova(mml2,mml3)
eml1 <-rm(dichotomous)
rm(x= dichotomous)
-1*eml1$beta
person.parameter(eml1)
person.parameter(eml1)$thetapar
person.parameter(eml1)$theta.table
lrtest(rm(dichotomous),splitcr = "mean")

jpeg("4. icc.jpg", width = 600, height = 350)
plotjointicc(rm(dichotomous))
dev.off()

library(mirt)
library(xlsx)
```

DOI: https://doi.org/10.14421/jpai.2021.181-09

```
# read data 'dichotomous.txt'
data <- read.spss(file="dikotomus2.sav",to.data.frame=true)


# of existing models 'rasch', '2pl', '3pl', '3plu', and '4pl'
# for 1pl use 2pl or without itemtype with the following constraint:
# model <- 'test = 1-15
# constraint = (1-15, a1)'
test_mdl <- mirt(data = data, model = 1, itemtype = "2pl")

# abilities/theta
ability_mle <- fscores(test_mdl, method = "ml")
ability_map <- fscores(test_mdl, method = "map")
ability_eap <- fscores(test_mdl, method = "eap")
ability <- cbind(ability_mle,ability_map, ability_eap)
colnames(ability) <- c('mle', 'map', 'eap')
abilities
```

Results and Discussion

IRT 1 PL Analisis analysis

The item in the Rasch model only has one item characteristic parameter, namely the item difficulty level/index which is denoted by b . The parameter level of difficulty of this item is able to measure the level of ability of the test takers. The following are the results of the analysis of 1 PL for 25 items of Al-Quran Hadith Subjects at MAN 2 Bantul.

Table 2. Analysis of 1 PL

	Dffclt	**Dscrmn**
Q.1	1.355371	1.058506
Q.2	0.768656	1.058506
Q.3	0.877347	1.058506
Q.4	0.357835	1.058506
Q.5	0.534044	1.058506

DOI: https://doi.org/10.14421/jpai.2021.181-09

Q.6	-0.03483	1.058506
Q.7	-2.1684	1.058506
Q.8	-151,964	1.058506
Q.9	-2.01743	1.058506
Q.10	-0.69178	1.058506
Q.11	2.559723	1.058506
Q.12	-1.71062	1.058506
Q.13	0.136423	1.058506
Q.14	-1.1194	1.058506
Q.15	-0.61208	1.058506
Q.16	-2.59031	1.058506
Q.17	-0.82824	1.058506
Q.18	-2.73753	1.058506
Q.19	-0.30572	1.058506
Q.20	-0.97024	1.058506
Q.21	0.458037	1.058506
Q.22	-1.08888	1.058506
Q.23	-0.08317	1.058506
Q.24	0.989449	1.058506
Q25	-0.50817	1.058506

The results of the analysis can be explained that *Dffclt* explain the level of difficulty of the test items and *Dscrmn* explain the differentiating power of the question. Obtained value *Dscrmn* the same because the analysis used is an analysis of 1 Logistics Parameters. The estimation results are then categorized based on the range as follows.

DOI: https://doi.org/10.14421/jpai.2021.181-09

Table 3. Item Parameter Category

Parameter	Classification	Category	Quantity Item	Percentage
Difficulty Level (b)	b < -2.00	Very easy	4	16%
	-2.00 b < -1.00	Easy	4	16%
	-1.00 b < 1.00	Medium	15	60%
	1.00 b 2.00	Difficult	1	4%
	b > 2.00	Very Difficult	1	4%

Based on the category table above, it can be seen that the level distribution of difficulty in the 25 items is quite balanced. This can be understood from the proportion of very easy and easy items is 16%. While the items with a difficulty level is only 4% are the same as very difficult. For items with a moderate category of 60%. However, the distribution of difficulty levels should be increased. This can be done by increasing the types of questions that have difficult or very difficult categories.

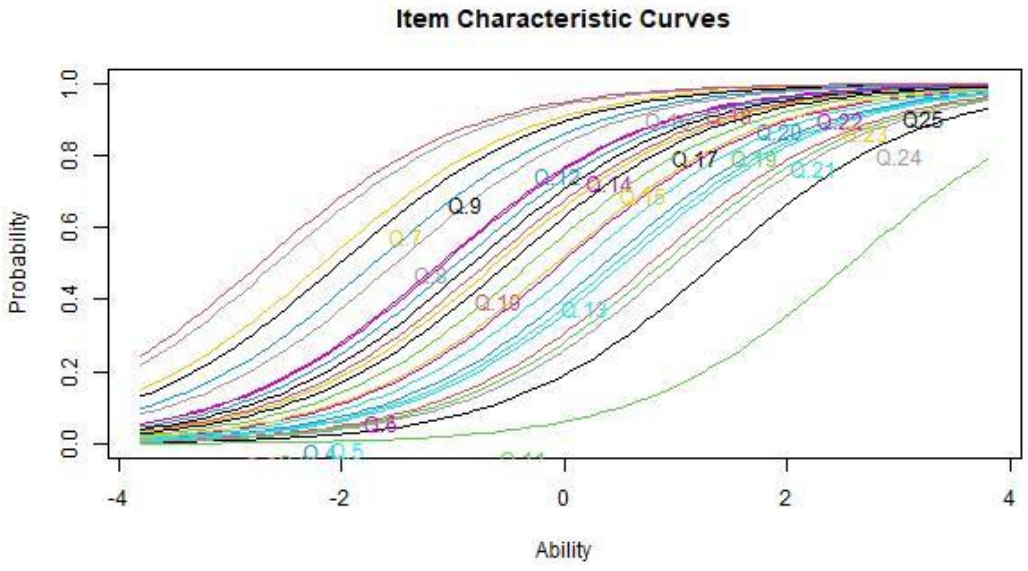

Image 1. ICC 1 PL Charts

Based on the ICC graph, it can be seen that the distribution of difficulty levels is quite good. The ability that must be possessed by

DOI: https://doi.org/10.14421/jpai.2021.181-09

students to answer 25 test items for Al-Quran Hadith subjects at MAN 2 Bantul with a probability of 50% is in the minimum range of abilities -2.5 to 2.8.

IRT 2 PL Analisis Analysis

The two-parameter model calculates item difficulty level (bi) and item discriminatory power (ai). The ability of an item to distinguish high-ability test respondents from low-ability test respondents is referred to as item distinguishing power (Istiyono 2020) . The following are the results of the 2 OT analysis for 25 items of the Al-Quran Hadith subject at MAN 2 Bantul

Table 4. PL analysis

	Dffclt	**Dscrmn**
Q.1	1.108636	1.397375
Q.2	-74.5931	-0.00887
Q.3	1.263563	0.643655
Q.4	0.29131	1.209015
Q.5	0.555723	0.922921
Q.6	-0.08773	0.706456
Q.7	-1.32942	2.771398
Q.8	-117,195	1.662072
Q.9	-133,926	2.239112
Q.10	-0.49266	2.208693
Q.11	8.750589	0.266629
Q.12	-1.2455	1.845419
Q.13	0.137757	0.751885
Q.14	-1.73133	0.624931
Q.15	-0.45859	2.002251
Q.16	-1.6886	2.29626
Q.17	-0.98994	0.874
Q.18	-1.6581	2.853659
Q.19	-0.29408	1.386694
Q.20	-0.84852	1.379353
Q.21	0.64788	0.632258
Q.22	-0.79788	1.910806
Q.23	-0.1158	1.173978
Q.24	1.427063	0.64607

DOI: https://doi.org/10.14421/jpai.2021.181-09

Q25	-0.93153	0.530164

Dffclt describes the level of difficulty of the item, while *Dscrmn* illustrates the distinguishing power, based on the above data analysis. The estimation results are then classified into the following categories based on the following ranges:.

Table 4. Item Parameter Category

Parameter	Classification	Category	Quantity Item	Percentage
Difficulty Level (b)	b < -2.00	Very easy	4	16%
	-2.00 b < -1.00	Easy	4	16%
	-1.00 b < 1.00	Medium	15	60%
	1.00 b 2.00	Difficult	1	4%
	b > 2.00	Very difficult	1	4%
Differential Power (a)	a > 0.5	Good	23	92%
	0.00 a 0.5	Pretty good	1	4%
	a < 0.00	Not good	1	4%

Based on the category table above, most of the questions are in the good category. T There are 23 items with good criteria. These items have different power values in the range of 0.00 – 0.5. The other 2 items each fall into the category of quite good and not good. The item that goes quite well is question number 11. While the items that fall into the bad category are item number 2. For item number 11 which is included in the good category, the form of the question can be improved so that it can better distinguish students' abilities. As for item number 2 which is in the bad category, it can be deleted and replaced with better items.

DOI: https://doi.org/10.14421/jpai.2021.181-09

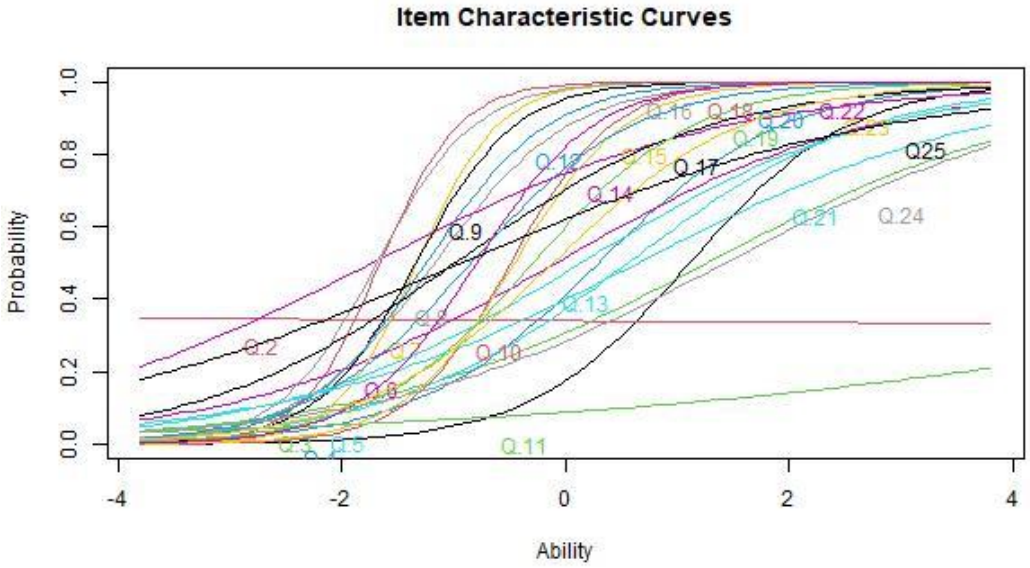

Image 2. ICC 2PL Graphics

The grain characteristic curve is shown in Figure 2. The different tilt or curvature of the 2 PL models indicates that the discriminating power for each item is different. The item curve will be better able to distinguish high-ability test takers from low-ability test takers if it is more vertical (large gradient/slope curve) (Istiyono, 2020).

3PL IRT Analysis

This three-parameter logistic model uses three parameters to indicate item characteristics, namely item difficulty level (bi), item discriminatory power (ai) and pseudo guessing (ci). The pseudo guessing parameter is the factor parameter that happens to be answered correctly which in this model is not equal to zero. Following are the results of 3 PL analysis for 25 items of Al-Quran Hadith Subject at MAN 2 Bantul.

Table 6. PL analysis

	Gussng	Dffclt	Dscrmn
Q.1	0.096198	1.07594	5.083873
Q.2	0.155359	-17.0692	-0.07445
Q.3	4.65E-06	1.361777	0.616356
Q.4	0.070711	0.487289	1.547262
Q.5	0.220381	1.017709	2.213989
Q.6	0.214478	0.612103	1.079666
Q.7	3.09E-07	-138,848	2.421695

DOI: https://doi.org/10.14421/jpai.2021.181-09

Q.8	0.238285	-0.73681	1.987298
Q.9	2.53E-07	-139,674	1.974496
Q.10	6.37E-10	-0.4316	2.031281
Q.11	3.10E-05	8.767217	0.26735
Q.12	0.256878	-0.81046	2.251701
Q.13	0.323531	0.970004	2.357427
Q.14	0.193584	-1.04118	0.702531
Q.15	5.08E-10	-0.40245	1.813948
Q.16	0.000138	-1.75455	2.161846
Q.17	1.13E-07	-0.97804	0.831136
Q.18	3.68E-07	-1.81802	2.307751
Q.19	0.257253	0.284136	3.024823
Q.20	1.02E-08	-0.81913	1.305795
Q.21	0.283958	1.116229	3.008819
Q.22	4.32E-05	-0.77556	1.731135
Q.23	0.23157	0.408504	2,916,564
Q.24	0.208857	1.350261	3.454573
Q25	0.00011	-0.87952	0.533665

Based on the data analysis above, *Dffclt* describe the level of difficulty of the test items, *Dscrmn* describe the power of difference and *Gussng* describe a pseudo guess. The estimation results are then categorized based on the range as follows.

Table 5. Item Parameter Category

Parameter	Classification	Category	Quantity Item	Percentage
Difficulty Level (b)	b < -2.00	Very easy	4	16%
	-2.00 b < -1.00	Easy	4	16%
	-1.00 b < 1.00	Medium	15	60%
	1.00 b 2.00	Difficult	1	4%
	b > 2.00	Very difficult	1	4%
Differential Power (a)	a > 0.5	Good	23	92%
	0.00 a 0.5	Pretty good	1	4%
	a < 0.00	Not good	1	4%
False Guess (c)	> 0.2	Not good	10	60%
	< 0.2	Good	15	40%

DOI: https://doi.org/10.14421/jpai.2021.181-09

On a test item, the value of ci or this pseudo guess ranges between 0 and 1. An item is said to be good if the value of ci is not more than 1/k, with k number of choices (Hullin, Drasgow, and Parsons 1983) . The number of choices for each test item on the general field mathematics test is 5 (ABCDE). Therefore, this item is said to be good if the value of c_i not more than 0.20.

Based on the results of the analysis in the category table above, it can be seen that from the 25 items, there are 10 items in the bad category and 15 in the good category. This indicates that for questions that are categorized as not good in the aspect of all guesses, even low-ability test takers have the opportunity to answer correctly even though the questions are quite difficult. (Istiyono, 2020) .

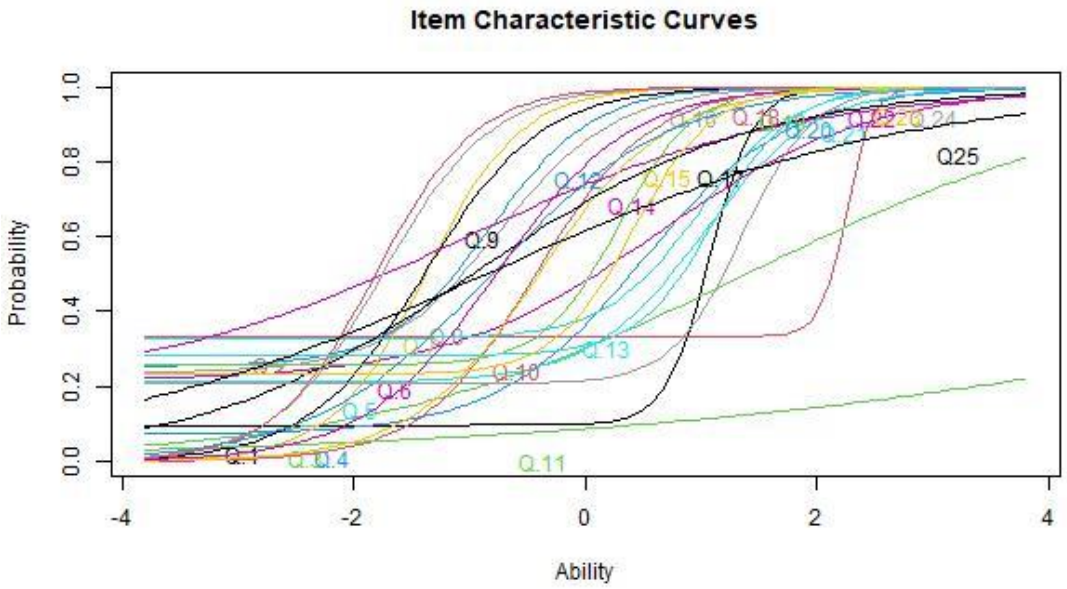

Image 3. ICC 3PL Graphics

Based on the grain characteristic curve graph for 3 PL analysis. the ci parameter gives the probability of a non-zero bottom asymptote on the characteristic curve. The picture above shows that the probability of answering correctly does not start from zero.

Model Fit Test

This model fit test aims to determine which analysis model best fits the data. If a data fits into a model, then the item will behave consistently as expected by the model. There are several ways to test the fit of the model, one of which is based on the AIC value (*Akaike*

Information Criterion). This AIC value calculates the balance between the magnitude of *likelihood* with the number of variables in the model. The most suitable model is the one with the lowest AIC (Snipes and Taylor 2014) .

The purpose of this model fit test is to find out which analysis model best fits the data. If a data fits the model, the item will behave consistently as the model predicts. There are several methods to test the fit of the model, one of which is to use the AIC value (*Akaike Information Criterion*). The balance between the likelihood and the number of variables in the model is calculated using this AIC value. The model with the lowest AIC value is the most suitable (Snipes and Taylor 2014) .

```
> anova(mml1,mml2)

 Likelihood Ratio Table
         AIC     BIC  log.Lik    LRT df p.value
mml1 4972.43 5056.58 -2460.21
mml2 4874.85 5036.67 -2387.42 145.58 24  <0.001

> anova(mml1,mml3)

 Likelihood Ratio Table
         AIC     BIC  log.Lik    LRT df p.value
mml1 4972.43 5056.58 -2460.21
mml3 4886.68 5129.41 -2368.34 183.75 49  <0.001

> anova(mml2,mml3)

 Likelihood Ratio Table
         AIC     BIC  log.Lik   LRT df p.value
mml2 4874.85 5036.67 -2387.42
mml3 4886.68 5129.41 -2368.34 38.17 25   0.045
```

Image 4. Model Fit Test

Based on the results of the above analysis, the smallest AIC value is found in the 2 PL model, which is 4874.85. So it can be concluded that the items fit the 2 PL model.

Student Ability Estimation Results

Below is a graph of the estimated ability of students to take the test. From the graph it can be seen that there are no students who have abilities below -4.00 on the test. While students who have abilities above 6.00 are 2 people from 188 students.

More Furthermore, it can be analyzed that there are 15 students with high abilities above 2.00. There are 165 students who fall into the

DOI: https://doi.org/10.14421/jpai.2021.181-09

medium category (ability from -2 to 2) and 9 students who have ability below -2 (low ability). Students with high abilities for example students number 45, 49, 55 with their respective abilities 2.498397, 2.342558 and 3.765938. Meanwhile, students with moderate abilities, for example students number 33, 64 and 84 with their respective abilities 1.267403, 0.227422, 0.764376. While students who have low abilities, for example students number 4, 23, and 56 with their respective abilities are -2.42629, -2.25918, -2.14486

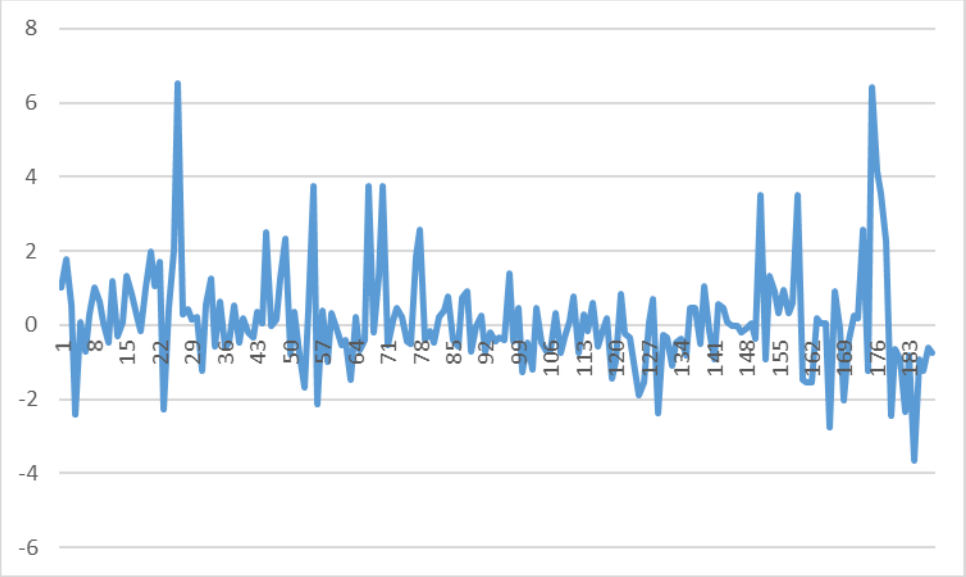

Image 5. Graph of Student Ability Distribution

If we look at the boxplot chart below, it can be seen that there are 3 outliers below and 6 outliers above. It can be understood that there are 3 students who are under the normal graph distribution. So the three students should get more attention. While the meaning of the 6 outliers above can be interpreted that there are 6 students who are smarter than the normal distribution.

DOI: https://doi.org/10.14421/jpai.2021.181-09

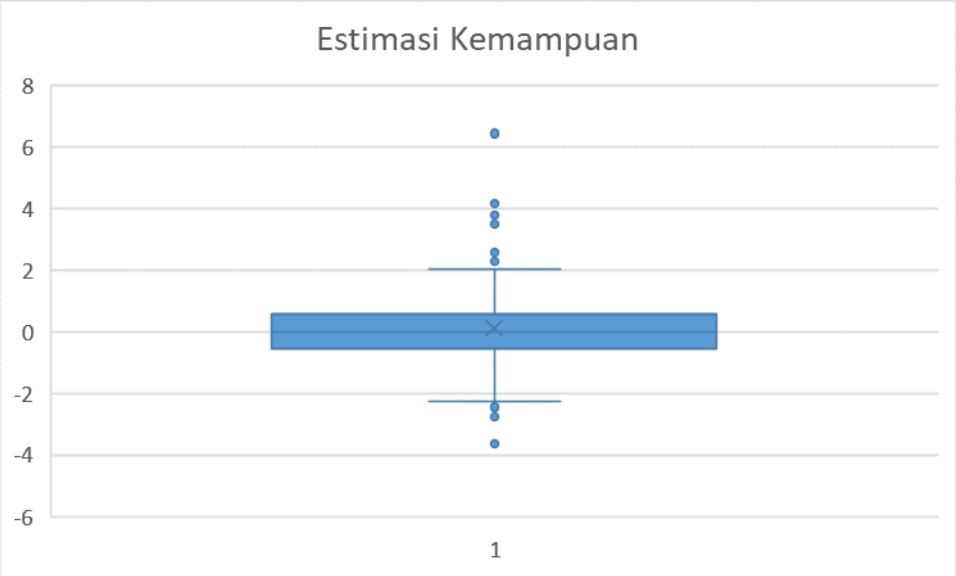

Image 6. Boxplot of Student Ability Distribution

Information Function

The following is a plot of the test information function based on these 25 items. Because the results of the model fit test show that the instrument fits the 2PL, the information function in the 2PL model is used. The plot in general can provide information that the value of the information function is a maximum of 1.6 and the test can measure well, especially on abilities below 0, approximately -1.2. That is, the test produces optimal information when given to individuals with low abilities. Then the test is overall appropriate for students with abilities around -2.5 to 1.2.

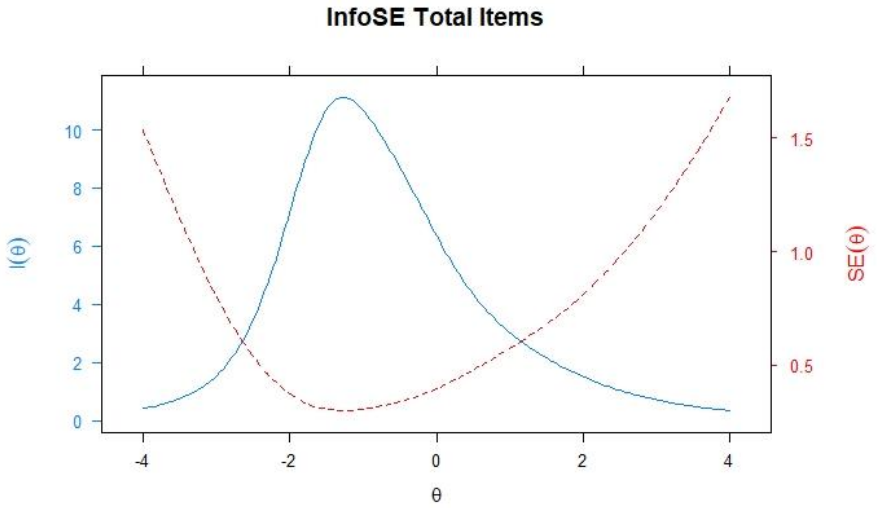

Image 7. Information Function Plot

Conclusion

From the results of the analysis that has been in conducted on 25 questions of Al-Quran Hadith subject at MAN 2 Bantul, it can be concluded: the results of the model fit test show that the instrument fit on the 2 PL model (Logistics parameter) with the lowest AIC value, namely 4874.85. The results of the parameter analysis of the level of difficulty indicate that there are 4 questions that are categorized as very easy, 4 easy questions, 15 moderate questions, 1 difficult question, and 1 very difficult question. This shows that the distribution of the difficulty level parameters is quite balanced. The results of the analysis of the different power parameters show that there are 23 good questions, 1 fairly good question, and 1 bad question. This shows that the different power level parameters are quite good. The results of the estimation of students' abilities with the MLE estimator showed that there were no students who had abilities below -4. There are 15 students with abilities above 2.00. There are 165 students who fall into the good category (ability -2 to 2 and 9 students who have abilities below -2. Based on the plot of the information function, it can be concluded that the optimal test if given to individuals with low abilities is around -1.2. Accurate questions to measure students' abilities with a range of -2.5 to 1.2.

Reference

Abdullah, M. Q., T. R. Ramalis, and ... 2019. "Karakteristik Tes Keterampilan Berpikir Kreatif Pada Mata Pelajaran Fisika SMA Materi Fluida Statis Melalui Analisis Teori Respon Butir." … *(Wahana Pendidikan Fisika ….*

Ahmad, Iqbal Faza, and Sukiman Sukiman. 2019. "Analisis Higher Order Thinking Skills (HOTS) Pada Soal Ujian Akhir Siswa Kelas 6 KMI Dalam Kelompok Mata Pelajaran Dirosah Islamiyah Di Pondok Modern Tazakka Batang." *Jurnal Pendidikan Agama Islam* 16(2):137–64. doi: 10.14421/jpai.2019.162-02.

Alifa, Tasya Fitria, Taufik Ramlan Ramalis, and Unang Purwana. 2018. "Karakteristik Tes Penalaran Ilmiah Siswa SMA Materi Mekanika Berdasarkan Analisis Tes Teori Respon Butir." *Jurnal Inovasi Dan Pembelajaran Fisika* 5(1):80–89.

Andayani, Asri, Purwanto, and Taufik Ramlan Ramalis. 2019. "Kajian Implementasi Teori Respon Butir Dalam Menganalisis Instrumen Tes Materi Fisika." *Seminar Nasional Fisika* 1(1):37–42.

Anisa, Anisa. 2013. "Perbandingan Penskoran Dikotomi Dan Politomi Dalam Teori Respon Butir Untuk Pengembangan Bank Soal Matakuliah Matematika Dasar." *Jurnal Matematika, Statistika Dan Komputasi* 9(2):95–113. doi: 10.20956/jmsk.v9i2.3402.

Arlinwibowo, Janu, Samsul Hadi, and Eric Maurteen Firdaus. 2020. "Validasi Persepsi Siswa Terhadap Pembelajaran Sains Dengan Teori Respon Butir Multidimensi." *Jurnal Ilmu Komputer Dan Matematika* 1(1):7–14.

Azizah, Nur, and Muhammad Zainudin. 2020. "Evaluasi Pembelajaran Pendidikan Agama Islam (Studi Di SMK Muhammadiyah 1 Dan SMK Muhammadiyah 2 Kota Palembang)." *Edification Journal: Pendidikan Agama Islam* 2(2):133–43. doi: 10.37092/ej.v1i2.139.

Balsis, Steve, Tabina K. Choudhury, Lisa Geraci, Jared F. Benge, and Christopher J. Patrick. 2018. "Alzheimer's Disease Assessment: A Review and Illustrations Focusing on Item Response Theory Techniques." *Assessment* 25(3):360–73. doi: 10.1177/1073191117745125.

Darmawan, M. D., D. E. Tarigan, and A. F. C. Wijaya. n.d. "Karakteristik Tes Keterampilan Proses Sains Siswa Pada Materi Fluida Statis Berdasarkan Analisis Teori Respon Butir." *WaPFi (Wahana Pendidikan …*.

Fitriani, Lani, Taufik Ramlan Ramalis, and Ridwan Efendi. 2019. "Karakterisasi Tes Keterampilan Proses Sains Materi Fluida Statis Berdasarkan Teori Respon Butir." *Omega: Jurnal Fisika Dan Pendidikan Fisika* 5(2):27–27. doi: 10.31758/omegajphysphyseduc.v5i2.27.

Foster, Garett C., Hanyi Min, and Michael J. Zickar. 2017. "Review of Item Response Theory Practices in Organizational Research: Lessons Learned and Paths Forward." *Organizational Research Methods* 20(3):465–86. doi: 10.1177/1094428116689708.

Halijah. 2018. "Kualitas Soal Ujian Sekolah Berstandar Nasional Mata Pelajaran Matematika Tingkat SMA Berdasarkan Teori Tes Klasik Dan Teori Respon Butir." masters, Universitas Negeri Makassar.

Hullin, CL, F. Drasgow, and CK Parsons. 1983. "Item Response Theory." *Homewood, Illinois: Dow Jones-Irwin*.

Ishak. 2019. "Analisis Kualitas Soal Ujian Sekolah Berbasis Nasional Mata Pelajaran IPA SMP Negeri Di Kabupaten Luwu Timur Berdasarkan Teori Respon Butir." masters, Universitas Negeri Makassar.

Istiyono, Edi. 2020. *Pengembangan Instrumen Penilaian Dan Analisis Hasil Belajar Fisika Dengan Teori Tes Klasik Dan Modern*. Yogyakarta: UNY Press.

Kurniawan, Devi Dwi. 2015. "Analisis Kualitas Soal Ujian Akhir Semester Matematika Berdasarkan Teori Respon Butir."

Kurniawati, Indah Rizqi, Purwo Susongko, and Paridjo. 2017. "Penyusunan Instrumen Tes Diagnostik Materi Ajar Operasi Hitung Aljabar dengan Pemodelan Teori Respon Butir (Studi Pengembangan pada Pembelajaran Matematika di SMP Negeri 1 Adiwerna Tahun Ajaran 2016/2017)." *JPMP (Jurnal Pendidikan MIPA Pancasakti)* 1(1). doi: 10.24905/jpmp.v1i1.784.

Linden, Ariel. 2018. "Review of Tenko Raykov and George Marcoulides's a Course in Item Response Theory and Modeling with Stata." *Stata Journal* 18(2):485–88. doi: 10.1177/1536867x1801800213.

Malatesta, Jaime, and Won-Chan Lee. 2019. "Software Review of IRTEQ, STUIRT, and POLYEQUATE for Item Response Theory Scale Linking and Equating." *Measurement: Interdisciplinary Research and Perspectives* 17(1):48–59. doi: 10.1080/15366367.2018.1539558.

Mardapi, Djemari. 2012. "Pengukuran Penilaian Dan Evaluasi Pendidikan." *Yogyakarta: Nuha Medika* 45.

Matondang, Panyahuti, Nizwardi Jalinus, and Ambiyar Ambiyar. 2020. "Analisis Pemetaan Kemampuan Numerik Siswa SMK Model Teori Respon Butir." *INVOTEK: Jurnal Inovasi Vokasional Dan Teknologi* 20(3):11–22. doi: 10.24036/invotek.v20i3.640.

McGrory, Sarah, Jason M. Doherty, Elizabeth J. Austin, John M. Starr, and Susan D. Shenkin. 2014. "Item Response Theory Analysis of Cognitive Tests in People with Dementia: A Systematic Review." *BMC Psychiatry* 14(1). doi: 10.1186/1471-244X-14-47.

Mulyani, S., R. Efendi, and T. R. Ramalis. 2021. "Karakterisasi Tes Keterampilan Pemecahan Masalah Fisika Berdasarkan Teori Respon Butir." *JURNAL Pendidikan Dan ….*

Parmaningsih, Triwik Jatu, and Dewi Retno Sari Saputro. 2021. "Rasch Analysis on Item Response Theory: Review of Model Suitability." *AIP Conference Proceedings* 2326. doi: 10.1063/5.0040305.

Rahmaini, Anida, and Aditya Nur Taufiq. 2018. "Analisis Butir Soal Pendidikan Agama Islam Di SMK N 1 Sedayu Tahun Ajaran 2017/2018." *Jurnal MUDARRISUNA: Media Kajian Pendidikan Agama Islam* 8(1):1–24. doi: 10.22373/jm.v8i1.2787.

Rahman, Aulia. 2021. "Analisis Karakteristik Tes Higher Order Thinking Skills (HOTS) Pada Materi Momentum Dan Impuls Menggunakan Teori Respon Butir." other, Universitas Pendidikan Indonesia.

Rusmayani, Rusmayani. 2020. "Analisis Butir Soal Penilaian Akhir Semester Genap Mata Pelajaran Pendidikan Agama Islam Di SMP Bintang Persada Tabanan-Bali." *Widya Balina* 5(9):42–49.

Sainuddin, Syamsir. 2018. "Analisis Karakteristik Butir Tes Matematika Pada Tes Buatan MGMP Matematika Kota Makassar berdasarkan Teori Moderen (Teori Respon Butir)." *Proximal: Jurnal Penelitian Matematika dan Pendidikan Matematika* 1(1).

Sarea, Muh Syahrul. 2018. "Karakteristik Soal Ujian Akhir Semester Pendidikan Agama Islam Dan Budi Pekerti Tingkat Sekolah Dasar." *An-Nahdhah* 11(2):303–18.

Sarea, Muh Syahrul, and Samsul Hadi. 2015. "Analisis Kualitas Soal Ujian Akhir Semester Mata Pelajaran Kimia SMA Di Kabupaten Gowa." *Jurnal Evaluasi Pendidikan* 3(1):35–43.

Snipes, Michael, and D. Christopher Taylor. 2014. "Model Selection and Akaike Information Criteria: An Example from Wine Ratings and

Prices." *Wine Economic and Policy* 1. doi: http://dx.doi.org/10.1016/j.wep.2014.03.001.

Sugiarto, Muh. 2018. "Analisis Kualitas Soal Ujian Sekolah Berstandar Nasional Mata Pelajaran Fisika Kelas XII IPA SMA Negeri Kabupaten Enrekang." masters, Universitas Negeri Makassar.

Syam, Aldo Redho. 2019. "Guru dan Pengembangan Kurikulum Pendidikan Agama Islam di Era Revolusi Industri 4.0." *TADRIS: Jurnal Pendidikan Islam* 14(1):1–18. doi: 10.19105/tjpi.v14i1.2147.