

**ANALISIS SOAL SKI KELAS X MAN 2 BANTUL YOGYAKARTA PADA
PENILAIAN AKHIR TAHUN 2020/2021**



Oleh: **Rani Putri Prihatin**

NIM: 20204011011

TESIS

Diajukan kepada Program Magister (S2)

Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga

untuk

Memenuhi Salah Satu Syarat guna Memperoleh

Gelar Magister Pendidikan (M. Pd)

Program Studi Pendidikan Agama Islam

YOGYAKARTA

2022

PERNYATAAN KEASLIAN

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Rani Putri Prihatin S.Pd
NIM : 20204011011
Jenjang : Magister
Program Studi : Pendidikan Agama Islam

Menyatakan bahwa tesis ini secara keseluruhan adalah hasil penelitian/karya sendiri, kecuali pada bagian-bagian yang dirujuk sumbernya.

Yogyakarta, 4 Juli 2022

Saya yang menyatakan



Rani Putri Prihatin S. Pd

NIM. 20204011011

PERNYATAAN BEBAS PLAGIASI

PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama : Rani Putri Prihatin S.Pd
NIM : 20204011011
Jenjang : Magister
Program Studi : Pendidikan Agama Islam

Menyatakan bahwa naskah tesis ini secara keseluruhan benar-benar bebas dari plagiasi.

Jika di kemudian hari terbukti melakukan plagiasi, maka saya siap ditindak sesuai ketentuan hukum yang berlaku.

Yogyakarta, 4 Juli 2022

Saya yang menyatakan



Rani Putri Prihatin S. Pd

NIM. 20204011011

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

SURAT KETERANGAN BERJILBAB

SURAT KETERANGAN BERJILBAB

Assalamu'alaikum Warahmatullahi Wabarokatuh

Saya yang bertanda tangan di bawah ini :

Nama : Rani Putri Prihatin S.Pd

NIM : 20204011011

Jenjang : Magister

Program Studi : Pendidikan Agama Islam

menyatakan dengan sesungguhnya bahwa pas foto yang diserahkan dalam pendaftaran munaqosyah ini adalah foto saya yang menggunakan jilbab. Demikian surat pernyataan ini saya buat dengan sesungguhnya dan penuh kesadaran, jika pernyataan ini tidak benar, maka saya sanggup menerima resiko yang telah ditetapkan.

Wassalamualikum Warrahmatullahi Wabarokatuh

Yogyakarta, 4 Juli 2022

Saya yang menyatakan



Rani Putri Prihatin S. Pd
NIM.20204011011

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

NOTA DINAS PEMBIMBING

NOTA DINAS PEMBIMBING

Kepada Yang Terhormat,
Dekan Fakultas Ilmu Tarbiyah dan Keguruan
UIN Sunan Kalijaga Yogyakarta

Assalamu'alaikum Warahmatullahi Wabarokatuh

Setelah melakukan bimbingan, arahan serta koreksi terhadap penulisan tesis yang berjudul

ANALISIS SOAL SKI KELAS X MAN 2 BANTUL YOGYAKARTA PADA

PENILAIAN AKHIR TAHUN 2020/2021

yang ditulis oleh:

Nama : Rani Putri Prihatin

NIM : 20204011011

Jenjang : Magister (S2)

Program Studi : Pendidikan Agama Islam

Saya berpendapat bahwa tesis tersebut sudah dapat diajukan kepada Program Magister (S2) Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta untuk diajukan dalam rangka memperoleh gelar Magister Pendidikan (M.Pd).

Wassalamu'alaikum Warrahmatullahi Wabarokatuh

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Yogyakarta, 22 Juli 2022

Pembimbing

Dr. Hj. Siti Fatonah, M.Pd

NIP: 19710205 199903 2 00

PENGESAHAN TUGAS AKHIR



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS ILMU TARBIYAH DAN KEGURUAN
Jl. Marsda Adisucipto Telp. (0274) 513056 Fax. (0274) 586117 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-2144/Un.02/DT/PP.00.9/08/2022

Tugas Akhir dengan judul : ANALISIS SOAL SKI KELAS X MAN 2 BANTUL YOGYAKARTA PADA PENILAIAN AKHIR TAHUN 2020/2021

yang dipersiapkan dan disusun oleh:

Nama : RANI PUTRI PRIHATIN, S.Pd
Nomor Induk Mahasiswa : 20204011011
Telah diujikan pada : Senin, 01 Agustus 2022
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang
Dr. Siti Fatonah, S.Pd., M.Pd
SIGNED

Valid ID: 62f6e192c545d



Penguji I
Prof. Dr. Abdul Munip, S.Ag., M.Ag.
SIGNED

Valid ID: 62fe0442cd53f



Penguji II
Dr. Dwi Ratnasari, S.Ag., M.Ag
SIGNED

Valid ID: 62ff30f0880cd



Yogyakarta, 01 Agustus 2022
UIN Sunan Kalijaga
Dekan Fakultas Ilmu Tarbiyah dan Keguruan

Prof. Dr. Hj. Sri Sumarni, M.Pd.
SIGNED

Valid ID: 62ffa095abf3e

PESETUJUAN TIM PENGUJI UJIAN TESIS

PERSETUJUAN TIM PENGUJI




UJIAN TESIS

Tesis Berjudul :

ANALISIS SOAL SKI KELAS X MAN 2 BANTUL YOGYAKARTA PADA PENILAIAN AKHIR TAHUN
2020/2021

Nama : Rani Putri Prihatin
NIM : 20204011011
Program Studi : Pendidikan Agama Islam
Fakultas : Ilmu Tarbiyah dan Keguruan

Telah disetujui tim penguji munaqosyah

Ketua/Pembimbing : Dr. Hj. Siti Fatonah, M. Pd. ()
Sekretaris/Penguji I : Prof. Dr. H. Abdul Munip, M. Ag. ()
Penguji II : Dr. Dwi Ratnasari, M. Ag. ()

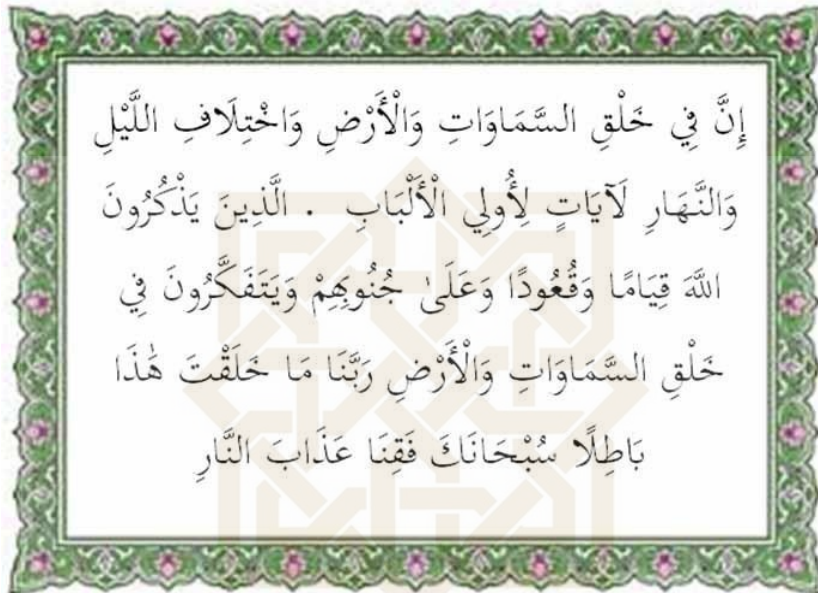
Diuji di Yogyakarta pada :

Tanggal : 1 Agustus 2022
Waktu : 09.00 - 10.00 WIB.
Hasil : A- (91,66)
IPK : 3,74
Predikat : Sangat Memuaskan

*coret yang tidak perlu

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

MOTTO



Artinya:

Sesungguhnya dalam penciptaan langit dan bumi, dan silih bergantinya malam dan siang terdapat tanda-tanda bagi orang-orang yang berakal, (yaitu) orang-orang yang mengingat Allah sambil berdiri atau duduk atau dalam keadan berbaring dan mereka memikirkan tentang penciptaan langit dan bumi (seraya berkata): "Ya Tuhan kami, tiadalah Engkau menciptakan ini dengan sia-sia, Maha Suci Engkau, maka peliharalah kami dari siksa neraka.

(QS. Ali Imron: 190-191)

STATE UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PERSEMBAHAN

Tesis ini penulis persembahkan kepada:

Almamater tercinta, Program Studi Pendidikan Agama Islam

Fakultas Ilmu Tarbiyah dan Keguruan

Universitas Islam Negeri Sunan Kalijaga Yogyakarta



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

ABSTAK

Rani Putri Prihatin, *Analisis Soal SKI Kelas X MAN 2 Bantul Yogyakarta Pada Penilaian Akhir Tahun 2020/2021*. Tesis, Yogyakarta: Program Studi Pendidikan Agama Islam Fakultas Ilmu Tarbiyah dan Keguruan Universitas Islam Negeri Sunan Kalijaga Yogyakarta, 2022.

Ujian akhir adalah kegiatan yang dilakukan oleh satuan pendidikan untuk mengukur pencapaian kompetensi peserta didik pada akhir semester. Instrument tes menjadi salah satu alat ukur untuk mengetahui kompetensi yang dimiliki oleh siswa. Dalam hal ini peneliti menggunakan sampel di MAN 2 Bantul Yogyakarta, dalam soal yang diujikan cenderung lebih banyak menguji aspek ingatan. Terdapat perbedaan format dalam perangkat administrasi/manajemen pelaksanaan KBK pada dimensi perencanaan, pelaksanaan dan evaluasi yang digunakan guru pada lima kabupaten/Kota di lingkungan Dinas Pendidikan Propinsi Daerah Istimewa Yogyakarta. Oleh karena itu, perlu ada upaya untuk membantu guru dalam melakukan analisis butir soal pada hasil belajar peserta didik. Pembuatan soal SKI oleh guru matapelajaran berdasarkan kisi-kisi yang disepakati MGMP (Musyawarah Guru Mata Pelajaran) MA Daerah Istimewa Yogyakarta

Penelitian ini merupakan penelitian deskriptif eksploratif dengan pendekatan kuantitatif. Pengumpulan data dilakukan melalui dokumentasi jawaban siswa, soal dan kisi-kisi soal ujian akhir. Data respon siswa dianalisis menggunakan pendekatan teori respon butir, SPSS dan R.

Hasil analisis yang telah dilakukan terhadap 25 soal mata pelajaran Sejarah Kebudayaan Islam MAN 2 Bantul dapat ditarik kesimpulan bahwa hasil uji validitas isi dinyatakan sangat tinggi dengan hasil 0,60, validitas empiris dinyatakan 25 butir soal valid dan 1 soal tidak valid di no soal 4, kategori level kognitif soal ujian akhir tahun dalam taksonomi bloom terdapat C1 21 dan C2 4 soal. Uji reliabilitas analisis soal terdapat nilai 0,82748 dengan kategori sangat tinggi. Hasil analisis parameter 1PL, diperoleh tingkat kesukaran dari 25 butir soal sudah cukup berimbang. Hal ini dapat diproporsi 7 butir soal dikategorikan sangat mudah, 5 butir soal mudah, 4 butir soal sukar dan 9 butir sedang. Hasil analisis parameter butir dengan model 2PL, parameter tingkat kesukaran 2 butir sangat mudah, 11 butir mudah, 11 butir sedang, dan 1 butir sangat sukar. Model 2PL ini tidak ada butir yang masuk dalam kategori sukar. Hasil analisis parameter daya beda menunjukkan terdapat 23 soal baik, 1 soal cukup baik, dan 1 soal tidak baik. Hal tersebut menunjukkan parameter tingkat daya beda sudah cukup baik. Hasil analisis model 3PL menunjukkan tingkat kesukaran 1 butir sangat mudah, 9 butir mudah, 13 butir sedang, 2 butir sukar. Daya beda butir soal 24 butir baik, 1 cukup baik, tebak semu dari butir soal 8 butir tidak baik dan 17 butir baik. Hasil Analisis kecocokan model 1PL, 2PL, 3PL model yang menghasilkan butir yang cocok terdapat pada model 1PL.

Kata kunci: *Analisis, Penilaian, Soal Akhir Tahun*

ABSTRACT

Rani Putri Concerned, Analysis of SKI Questions for Class X MAN 2 Bantul Yogyakarta in the 2020/2021 Final Assessment. Thesis, Yogyakarta: Islamic Religious Education Study Program, Faculty of Tarbiyah and Teacher Training, Sunan Kalijaga State Islamic University, Yogyakarta, 2022.

Final exam is an activity carried out by the education unit to measure the achievement of student competencies at the end of the semester. The test instrument is one of the measuring tools to determine the competencies possessed by students. In this case, the researcher used a sample at MAN 2 Bantul Yogyakarta, in which the questions tested tended to test more aspects of memory. There are different formats in the administrative or management tools for the implementation of KBK on the dimensions of planning, implementation, and evaluation used by teachers in five districts/cities within the Education Office of the Special Region of Yogyakarta. Therefore, there needs to be an effort to assist teachers in conducting item analysis on student learning outcomes. The making of SKI questions by subject teachers is based on the grid agreed upon by the MGMP (Subject Teacher Consultation) of the Special Region of Yogyakarta.

This research is an exploratory descriptive study with a quantitative approach. Data was collected through documentation of student answers, questions, and final exam questions. Student response data were analyzed using the item response theory approach, SPSS, and R.

The results of the analysis that has been carried out on 25 questions about the Islamic Cultural History subject at MAN 2 Bantul can be concluded that the results of the content validity test are stated to be very high with a result of 0.60, the empirical validity is stated to be 25 valid items and 1 item is invalid in question no. 4, the cognitive level category of year-end exam questions in Bloom's taxonomy contains C1 21 and C2 4 questions. The reliability test of the question analysis has a value of 0.82748 with a very high category. The results of the IPL parameter analysis showed that the difficulty level of the 25 items was quite balanced. It can be proportioned that 7 items are categorized as very easy, 5 items are easy, 4 items are difficult and 9 items are moderate. The results of the item parameter analysis with the 2PL model, the level of difficulty parameters 2 items are very easy, 11 items are easy, 11 items are moderate, and 1 item is very difficult. In this 2PL model, there are no items that fall into the difficult category. The results of the analysis of the different power parameters showed that there were 23 good questions, 1 good question, and 1 bad question. This shows that the different power level parameters are quite good. The results of the 3PL model analysis show that the difficulty level is 1 very easy item, 9 easy items, 13 moderate items, and 2 difficult items. The differentiating power of 24 items is good, 1 is quite good, false guesses from the questions, 8 items are not good and 17 items are good. The results of the analysis of the suitability of the IPL, 2PL, and 3PL models that produce suitable items are found in the IPL model.

Keywords: Analysis, Assessment, Year-End Question

KATA PENGANTAR

الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ، أَشْهَدُ أَنْ لَا إِلَهَ إِلَّا اللَّهُ وَ أَشْهَدُ أَنَّ مُحَمَّدًا رَسُولُ اللَّهِ، وَالصَّلَاةُ وَالسَّلَامُ عَلَى أَشْرَفِ الْأَنْبِيَاءِ
وَالْمُرْسَلِينَ وَعَلَى آلِهِ وَصَحْبِهِ أَجْمَعِينَ، أَمَّا بَعْدُ

Dengan menyebut nama Allah Yang Maha Pengasih lagi Maha Penyayang. Segala puji bagi Allah Tuhan semesta alam yang telah melimpahkan taufiq, rahmat, dan hidayah-Nya, sehingga penulis dapat menyelesaikan penyusunan tesis ini. Peneliti menyadari bahwa penyelesaian tesis ini atas pertolongan Allah Swt. Shalawat dan salam semoga terlimpahkan kepada Nabi Muhammad saw. sebagai sosok teladan dalam dunia pendidikan, keluarga, dan para sahabat. Tesis ini merupakan kajian singkat tentang Analisis Soal SKI Kelas X MAN 2 Bantul Yogyakarta Pada Penilaian Akhir tahun 2020/2021. Peneliti sepenuhnya menyadari bahwa tesis ini tidak akan terwujud tanpa adanya bantuan, bimbingan, dan dukungan dari berbagai pihak. Dengan segala kerendahan hati peneliti mengucapkan banyak terima kasih kepada:

1. Bapak Prof. Dr. Phil. Al Makin, S.Ag., M.A, selaku Rektor UIN Sunan Kalijaga, yang telah memberikan pengarahan dan motivasi selama menjadi mahasiswa UIN Sunan Kalijaga.
2. Ibu Prof. Dr. Hj. Sri Sumarni, M. Pd, selaku Dekan Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta, yang telah membantu dalam penyusunan tesis dan memberi pengarahan selama menjadi mahasiswa Fakultas Ilmu Tarbiyah dan Keguruan.
3. Bapak Prof. Dr. Mahmud Arif, M. Ag selaku Ketua Prodi Magister Pendidikan Agama Islam yang telah memberikan banyak motivasi untuk menjadi mahasiswa yang kreatif, inovatif, dan produktif.
4. Ibu Dr. Dwi Ratnasari, S.Ag., M.Ag., selaku Sekretaris Prodi Magister Pendidikan Agama Islam dan selaku Dosen Penasehat Akademik yang telah memberikan arahan dan motivasi dalam menempuh jenjang perkuliahan di program studi PAI.

5. Ibu Dr. Hj. Siti Fatonah, M. Pd, selaku Pembimbing Tesis yang telah mencurahkan ketekunan, kesabaran, dukungan, motivasi, meluangkan waktu, tenaga, pikiran, dan bimbingan dalam penyusunan tesis ini.
6. Segenap dosen dan karyawan Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta yang telah sabar membimbing penulis selama ini.
7. Bapak Drs. H. Ulul Ajib, M. Pd, selaku Pimpinan MAN 2 Bantul Yogyakarta dan seluruh Bapak Ibu guru yang telah memberikan ijin penelitian dan segala bentuk dukungan lainnya dalam penyelesaian tesis ini.
8. Orang tua tercinta Ibu Sri Miyatun yang telah memberikan motivasi, nasihat, kasih sayang, dan membimbing penulis dengan penuh ketulusan dan kesabaran.
9. Teman-teman seperjuangan Magister PAI 2020, Korp Granit, keluarga PMII Rayon Wisma Tradisi terima kasih untuk kebersamaan, persahabatan selama menempuh pendidikan di UIN Sunan Kalijaga Yogyakarta.
10. Semua pihak yang telah berjasa dalam penulisan tesis ini yang tidak bisa disebutkan satu persatu

Peneliti berdoa semoga semua bantuan, bimbingan, dukungan, tersebut diterima sebagai amal baik oleh Allah SWT, amin.

Yogyakarta, 19 Juli 2022

Penulis,



Rani Putri Prihatin
NIM. 20204011011

DAFTAR ISI

| | |
|---------------------------------------------------------------------|-------|
| PERNYATAAN KEASLIAN..... | ii |
| PERNYATAAN BEBAS PLAGIASI | iii |
| SURAT KETERANGAN BERJILBAB | iv |
| NOTA DINAS PEMBIMBING | v |
| PENGESAHAN TUGAS AKHIR | vi |
| PESETUJUAN TIM PENGUJI UJIAN TESIS | vii |
| MOTTO | viii |
| PERSEMBAHAN..... | ix |
| ABSTAK..... | x |
| <i>ABSTRACT</i> | xi |
| KATA PENGANTAR | xii |
| DAFTAR ISI..... | xiv |
| DAFTAR TABEL..... | xvi |
| DAFTAR GRAFIK..... | xviii |
| DAFTAR GAMBAR | xix |
| BAB I PENDAHULUAN | 1 |
| A. Latar Belakang | 1 |
| B. Rumusan Masalah..... | 15 |
| C. Tujuan dan Manfaat Penelitian | 15 |
| D. Kajian Pustaka | 17 |
| E. Kerangka Teori Analisis Butir Soal | 23 |
| 1. Analisis Butir Soal | 23 |
| 2. Validitas Isi/Konten | 25 |
| 3. Taksonomi Bloom..... | 28 |
| 4. Validitas Empiris | 32 |
| 5. Reliabilitas | 33 |
| 6. Teori Respon Butir (<i>Item Response Theory</i>)..... | 36 |
| 7. Uji Asumsi Item Response Theory | 45 |
| 8. Estimasi Kemampuan dengan Metode <i>Maximum Likelihood</i> | 48 |
| 9. Kecocokan Model | 52 |

| | |
|----------------------------------------------------------------------------------------------------------------|------------|
| F. Metode Penelitian | 54 |
| 1. Desain Penelitian | 54 |
| 2. Variabel Penelitian..... | 55 |
| 3. Objek Penelitian..... | 55 |
| 4. Instrumen Penelitian | 56 |
| 5. Teknik Pengumpulan Data..... | 57 |
| 6. Teknik Analisis Data..... | 60 |
| G. Sistemastika Pembahasan | 69 |
| BAB II GAMBARAN UMUM MADRASAH ALIYAH NEGERI 2 BANTUL | 71 |
| A. Sejarah Singkat | 71 |
| B. Lokasi Madrasah..... | 72 |
| C. Visi, Misi dan Tujuan Madrasah..... | 73 |
| D. Struktur Organisasi MAN 2 Bantul | 78 |
| E. Keadaan Guru Pengajar | 79 |
| F. Keadaan Siswa MAN 2 Bantul | 80 |
| BAB III HASIL ANALISIS BUTIR SOAL AKHIR TAHUN PELAJARAN SKI KELAS X DI MAN 2 BANTUL TAHUN 2020/2021 | 81 |
| A. Validitas Isi | 81 |
| B. Taksonomi Bloom..... | 86 |
| C. Validitas Empiris | 91 |
| D. Estimasi Reliabilitas..... | 92 |
| E. Analisis Teori Respon Butir Model 1PL..... | 94 |
| F. Analisis Teori Respon Butir Model 2PL..... | 97 |
| G. Analisis Teori Respon Butir Model 3PL..... | 100 |
| BAB IV PENUTUP | 113 |
| A. Kesimpulan | 113 |
| B. Rekomendasi..... | 115 |
| DAFTAR PUSTAKA | 116 |
| LAMPIRAN-LAMPIRAN | 124 |

DAFTAR TABEL

- Tabel 1. Kompetensi Dasar Mata Pelajaran SKI Kelas X, 13
- Tabel 2. Tabulasi Silang Hasil Penilaian Kedua Validator, 26
- Tabel 3. Kriteria Validitas, 27
- Tabel 4. Revisi Taksonomi Bloom, 28
- Tabel 5. Kategori dalam Proses Kognitif, 31
- Tabel 6. Kriteria Reliabilitas, 35
- Tabel 7. Tabulasi Silang Hasil Penilaian Kedua Validator, 61
- Tabel 8. Kriteria Validitas, 61
- Tabel 9. Kriteria Reliabilitas, 66
- Tabel 10. Kriteria kecocokan butir dengan model rasch, 66
- Tabel 11. Kriteria Lolos tidanya suatu butir soal, 67
- Tabel 12. Kriteria Indeks Kesukaran Butir, 67
- Tabel 13. Kriteria kualitas butir menurut pendekatan teori respon butir, 67
- Tabel 14. Kriteria kemampuan responden, 68
- Tabel 15. Data Guru Pengajar di MAN 2 Bantul, 80
- Tabel 16. Penilaian Validitas Isi Para Ahli, 81
- Tabel 17. Tabulasi Silang Hasil Penilaian Kedua Validator, 84
- Tabel 18. Hasil Penilaian Validator, 85
- Tabel 19. Kriteria Validitas, 87
- Tabel 20. Level Kognitif Soal, 89
- Tabel 21. Level Kognitif Kometensi Dasar, 91
- Tabel 22. Validitas Empiris, 94
- Tabel 23. Kriteria Reliabilitas, 95
- Tabel 24. Estimasi Parameter Butir Model 1PL, 97
- Tabel 25. Kategori Parameter Butir 1PL ,97
- Tabel 26. Estimasi Pameter Butir Model 2PL, 100
- Tabel 27. Kategori Parameter Butir 2PL, 100
- Tabel 28. Estmasi Pameter Butir Model 3PL, 103

Tabel 29. Kategori Parameter Butir, 104

Tabel 30. Item Fit Model 1PL, 107

Tabel 31. Item Fit Model 2PL, 108

Tabel 32. Item Fit Model 3PL, 110

Tabel 33. Kecocokan Model, 113



DAFTAR GRAFIK

- Grafik 1. Hasil Tabulasi S ilang Para Ahli, 85
Grafik 2. Kategori Level Kognitif Indikator Soal, 88
Grafik 3. Kategori Parameter Butir 1PL, 96
Grafik 4. ICC per Item Model 1PL, 97
Grafik 5. Kategori Parameter Butir 2PL, 99
Grafik 6. ICC per item Model 2PL, 100
Grafik 7. Kategori Parameter Butir 3PL, 102
Grafik 8. ICC per item Model 3PL, 103
Grafik 9. Uji Kecocokan Model, 111



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR GAMBAR

Gambar 1 Kurva Karakteristik Butir Model 1PL 41

Gambar 2 Kurva karakteristik butir model 2P 42

Gambar 3. Kurva karakteristik butir model 3P, 42

Gambar 4. Model Nilai AIC, 108



BAB I

PENDAHULUAN

A. Latar Belakang

Penilaian Akhir Tahun adalah kegiatan yang dilakukan oleh satuan pendidikan pada akhir semester genap untuk mengukur pencapaian kompetensi peserta didik pada akhir semester genap pada satuan pendidikan yang menggunakan sistem paket.¹ Dalam sebuah penilaian idealnya, instrumen tes hasil belajar yang disusun oleh guru dapat memberikan banyak informasi mengenai kemampuan siswa. Akan tetapi, persoalan yang banyak muncul saat guru menyusun instrumen tes ini ternyata masih banyak ditemukan kelemahan dalam proses penyusunannya sehingga menghasilkan tes yang tidak memiliki daya ukur yang valid. Instrumen tes yang tidak memiliki daya ukur yang valid tidak akan memberikan informasi apapun mengenai kemampuan peserta tes. Untuk mengevaluasi pencapaian proses pembelajaran yang umum dilakukan oleh guru adalah dengan menyusun instrumen tes hasil belajar.

Pada dasar analisis instrumen butir soal sangat penting untuk dilakukan oleh guru. Hal tersebut bertujuan agar guru dapat mengetahui kualitas dan validitas dari butir soal yang telah di susun. Selain itu, melalui analisis butir soal guru akan mendapatkan umpan balik yang nantinya dapat dijadikan acuan dalam pengambilan keputusan. Soal yang masuk kategori layak dapat langsung

¹ Kemendikbud, *Panduan Penilaian Oleh Pendidik Dan Satuan Pendidik Untuk Sekolah Menengah Atas*, (Jakarta: Direktorat Jenderal Pendidikan Dasar dan Menengah, 2017), hlm. 217.

digunakan, soal dengan kategori kurang layak dapat direvisi, sedangkan soal dengan kategori tidak layak dapat digugurkan sehingga harus dihapus atau diganti dengan soal lainnya. Namun sayangnya, pengetahuan dan keterampilan guru dalam melakukan analisis butir soal masih tergolong rendah.² Sebagian besar guru cenderung membuat soal sesuai dengan kompetensi inti dan kompetensi dasar yang telah ditentukan tanpa mengukur terlebih dahulu apakah siswa telah memahami materi-materi yang akan diujikan. Dengan demikian sudah dapat dipastikan bahwa kompetensi siswa tentunya tidak dapat diukur secara tepat.³

Sekolah dan guru menjadi satu bagian yang penting dalam memberikan fasilitas dalam proses pembelajaran siswa, tanpa adanya guru, siswa tidak akan berkembang dan meningkat hasil belajarnya. Upaya yang guru lakukan untuk mengetahui peningkatan pengetahuan siswa salah satunya dengan melakukan penilaian atau biasa disebut juga sistem penilaian hasil belajar siswa selama di Madrasah. Hasil belajar dilakukan untuk mengukur pemahaman dan penguasaan pada materi yang diberikan oleh guru pada siswa, yang artinya penilaian ini dapat memberikan gambaran dari masing-masing siswa dalam mencapai kompetensinya.⁴ Istiyono menjelaskan bahwa Penilaian adalah proses mengumpulkan informasi tentang siswa dan kelas untuk tujuan pengambilan

² Ati Sumiati, Umi Widiastuti, and Usep Suhud, "Workshop Teknik Menganalisis Butir Soal Dalam Meningkatkan Kompetensi Guru Di SMK Cileungsi Bogor," dalam *Jurnal Pemberdayaan Masyarakat Madani (JPMM)* 2, no. 1, 2018, hlm. 53.

³ Riza Yonisa Kurniawan, "Pemberian Pelatihan Analisis Butir Soal Bagi Guru Di Kabupaten Jombang: Efektif?," dalam *Jurnal Pemberdayaan Masyarakat Madani (JPMM)* 1, no. 2, 2017, hlm. 93.

⁴ Sukma Adi Perdana, "Analisis Kualitas Instrumen Pengukuran Pemahaman Konsep Persamaan Kuadrat Melalui Teori Tes Klasik Dan Rasch Model," dalam *Jurnal Kiprah* 6, no. 1, 2018, hlm. 41–48.

keputusan instruksional.⁵ Untuk melakukan penilaian diperlukan instrumen yang dapat mengukur kemampuan siswa secara akurat. Dalam hal ujian sumatif atau akhir semester, biasanya dilakukan dengan pilihan ganda. Proses pemberian angka pada sesuatu atau seseorang berdasarkan aturan disebut pengukuran. Kegiatan menentukan angka secara sistematis untuk suatu objek didefinisikan sebagai pengukuran oleh Mardapi.⁶ Oleh karena itu penilaian sangat penting untuk dilakukan oleh guru, dengan melakukan penilaian guru dapat mengetahui kemampuan dari siswa, dalam melaksanakan kegiatan ini salah satu cara adalah menggunakan soal tes dan non tes. Soal tes ini sering digunakan pada sekolah atau madrasah untuk mengukur kemampuan siswa dalam melaksanakan penilaian. Soal tes dinilai mampu mengkomodir setiap materi yang telah disampaikan oleh guru kepada siswa pada setiap pembelajaran dalam soal-soal uraian maupun pilihan ganda.⁷

Yahya Qohar sebagaimana dikutip oleh badrun badrun mengatakan bahwa sebuah evaluasi harus memenuhi 6 (enam) syarat, yaitu: (1) Reliabel; (2) Valid; (3) Obyektif; (4) Harus Diskriminatif; (4) Imperhensif; dan (5) Mudah Digunakan. Diskriminatif maksudnya ia mampu membedakan nilai dan ukuran siswa dalam capaiannya terhadap materi pelajaran yang telah diberikan. Dan

⁵ Istiyono, *Pengembangan Instrumen Penilaian Dan Analisis Hasil Belajar Fisika Dengan Teori Tes Klasik Dan Modern*, (Yogyakarta: UNY Press, 2020), hlm. 89.

⁶ Mardapi, *Pengukuran Penilaian Dan Evaluasi Pendidikan*, (Yogyakarta: Nuha Medika, 2012), hlm. 45.

⁷ Sukma Adi Perdana, "Analisis Kualitas Instrumen Pengukuran Pemahaman Konsep Persamaan Kuadrat Melalui Teori Tes Klasik dan Rasch Model, dalam *Jurnal Kiprah*, Vol. VI (1), 2018, hlm. 42.

sejalan dengan pendapat ini, yakni bahwa sebuah evaluasi harus memenuhi syarat diskriminatif, maka sebuah tes haruslah dianalisis dari segi tingkat kesulitan dan daya pembedanya.⁸

Syamsudin mengatakan bahwa ada 4 (empat) cara yang dapat digunakan untuk menilai sebuah test yang baik, yaitu: (1) meneliti secara jujur soal-soal yang telah disusun; (2) mengadakan analisis soal (item analysis); (3) memeriksa validitas (checking validity); dan (4) memeriksa reliabilitas (checking reliability). Cara yang pertama adalah cara yang bersifat subyektif, sedang 3 (tiga) yang lain dapat dilaksanakan dengan secara kuantitatif dan terukur. Analisis soal dapat dilakukan dengan cara: (1) Tingkat kesukaran (difficulty level); (2) Daya Pembeda (Discriminative test); dan (3) Pola Jawaban soal.⁹

Menurut Azis tingkat kesukaran adalah seberapa mudah atau sulitnya suatu butir soal bagi sekelompok siswa. Secara umum dapat dikatakan bahwa tingkat kesukaran merupakan tingkat mudah atau tidaknya suatu soal yang diberikan pada sekelompok siswa. Purwanto menjelaskan bahwa daya pembeda (DB) adalah kemampuan butir soal THB (Tes Hasil Belajar) membedakan siswa yang mempunyai kemampuan tinggi dan rendah.¹⁰ Analisis daya pembeda ini bertujuan untuk mengetahui kesanggupan soal dalam membedakan siswa yang

⁸ Badrun Kartowagiran, *Pengantar Teori Tes Klasik*, (Yogyakarta: Pascasarjana UNY & Dinas Pendidikan Propinsi Yogyakarta, 2009), hlm. 45.

⁹ Sri Wening, *Materi Evaluasi Pembelajaran*, (Yogyakarta: Fakultas Teknik Universitas Negeri Yogyakarta, 2012), hlm. 415.

¹⁰ Purwanto, *Evaluasi Hasil Belajar*, (Yogyakarta: Pustaka Pelajar, 2010), hlm. 65.

tergolong mampu (tinggi prestasinya) dengan siswa yang tergolong lemah prestasinya.¹¹

Sementara pola jawaban soal adalah distribusi testee dalam hal menentukan pilihan jawaban pada soal bentuk pilihan ganda. Pola jawaban soal diperoleh dengan menghitung banyaknya testee yang memilih pilihan jawaban a, b, c, atau d atau yang tidak memilih pilihan manapun (blangko). Dalam istilah evaluasi disebut omit, disingkat O. Dari pola jawaban soal dapat ditentukan apakah pengecoh (distractor) berfungsi sebagai pengecoh dengan baik atau tidak. Pengecoh yang tidak dipilih sama sekali oleh testee berarti bahwa pengecoh itu jelek, terlalu menyolok menyesatkan. Sebaliknya sebuah *distractor* (pengecoh) dapat dikatakan berfungsi dengan baik apabila *distractor* tersebut mempunyai daya tarik yang besar bagi pengikut-pengikut tes yang kurang memahami konsep atau kurang menguasai bahan.

Penilaian hasil belajar oleh pendidik dilakukan secara berkesinambungan untuk memantau proses, kemajuan dan perbaikan hasil dalam bentuk ulangan harian, ujian tengah semester, ujian akhir semester dan ujian kenaikan kelas. Penilaian hasil belajar oleh peserta didik digunakan untuk menilai pencapaian standar kompetensi lulusan untuk semua mata pelajaran. Penilaian hasil belajar oleh pemerintah dalam bentuk ujian nasional bertujuan untuk menilai pencapaian kompetensi lulusan untuk semua mata pelajaran.¹² Penilaian didasarkan pada data

¹¹ N.Sudjana, *Penilaian Hasil Proses Belajar Mengajar*, (Bandung: Remaja, 2012), hlm. 78.

¹² Sudaryono Sudaryono, "Implementasi Teori Responsi Butir (*Item Response Theory*) Pada Penilaian Hasil Belajar Akhir Di Sekolah," dalam *Jurnal Pendidikan Dan Kebudayaan* 17, no. 6, 2011, hlm. 720.

yang mencerminkan kemampuan siswa.¹³ Penilaian bertujuan untuk mengetahui karakteristik suatu objek yang akan diukur. Khususnya pengukuran pendidikan meliputi pengukuran hasil belajar yang mencakup bermacam-macam bidang, tergantung objek hasil belajar apa yang ingin diukur. Dapat dikatakan bahwa pengukuran bersifat kuantitatif, untuk menentukan nilai secara kuantitatif diperlukan alat ukur, alat ukur yang dimaksud salah satunya berupa tes. Alat ukur atau instrumen tes yang umum digunakan dalam mengevaluasi hasil belajar peserta didik adalah seperangkat soal.

Secara umum, alat penilaian (instrumen) dapat dikategorikan ke dalam dua bentuk yakni: 1) Tes; dan 2) Bukan Tes (non-test). Alat pengukuran yang termasuk ke dalam kategori non tes adalah: a) Kuesioner; b) Wawancara; c) Daftar Cocok (check list); d) Pengamatan atau Observasi; e) Penugasan; f) Portofolio; g) Jurnal; h) Inventori; i) Penilaian diri (Self Assesment); j) Penilaian oleh teman (peer assessment). Sedangkan tes adalah sejumlah pertanyaan yang harus dijawab, atau pertanyaan-pertanyaan yang harus dipilih atau ditanggapi, tugas-tugas yang harus dikerjakan oleh orang yang diuji pada waktu tertentu. Tes merupakan sejumlah pertanyaan yang memiliki jawaban benar atau salah, pertanyaan yang membutuhkan jawaban atau diberi tanggapan untuk mengukur tingkat kemampuan seseorang dalam aspek tertentu.¹⁴

¹³ Siti Fatonah, "Evaluasi Pelaksanaan Asesmen Otentik Kurikulum 2013 Di MI Yogyakarta," dalam *Al-Bidayah: Jurnal Pendidikan Dasar Islam* 8, no. 2, 2016, hlm. 115.

¹⁴ Sri Wening, *Materi Evaluasi Pembelajaran...*, hlm. 415.

Dalam kaitannya dengan penyusunan sebuah tes, Badrun Kartowagiran memaparkan tentang langkah-langkah penyusunan tes secara umum, yaitu: (1) Penentuan tujuan tes; (2) penyusunan kisi-kisi; (3) Penulisan soal; (4) Penelaahan soal (Review dan Revisi Soal); (5) Uji coba soal; (6) perakitan soal menjadi perangkat tes. Penentuan tujuan dapat berupa tujuan khusus, yaitu melihat tingkat capaian suatu program. Kisi-kisi soal adalah deskripsi mengenai ruang lingkup soal yang akan diujikan serta memberikan perincian mengenai soal-soal yang diperlukan dalam tes tersebut. Penulisan soal adalah penjabaran jenis dan tingkat perilaku yang hendak diukur menjadi pertanyaan-pertanyaan yang karakteristiknya sesuai dengan perinciannya dalam kisi-kisi. Telaah soal adalah kajian soal secara teoritis. Uji coba soal merupakan langkah untuk mengetahui kualitas soal yang akan diujikan dengan melihat secara empiris respons peserta tes. Perakitan soal adalah penyajian soal menjadi perangkat tes yang terpadu.¹⁵

Alat ukur yang bisa digunakan dalam proses evaluasi proses pembelajaran bisa berupa tugas-tugas rumah, kuis, ujian tengah semester (UTS), dan ujian akhir semester (UAS). Tes merupakan salah satu bentuk instrumen yang digunakan untuk melakukan pengukuran atas sejumlah pertanyaan yang memiliki jawaban benar atau salah, atau semua benar atau sebagian benar dengan tujuan untuk mengetahui pencapaian belajar atau kompetensi yang telah dicapai peserta didik dalam bidang tertentu.¹⁶ Penilaian Sumatif atau disebut juga dengan PAS

¹⁵ Badrun Kartowagiran, *Pengantar Teori Tes Klasik...*, hlm. 2–4.

¹⁶ Mardapi, *Pengukuran Penilaian Dan Evaluasi Pendidikan...*, hlm. 45.

(Penilaian Akhir Semester) terkadang dibuat terlalu sulit atau terlalu mudah sehingga menyulitkan pendidik untuk membedakan kemampuan siswa. Oleh karena itu, perlu dilakukan pengujian/analisis terhadap soal-soal tes dengan harapan hasil yang diperoleh secara akurat mencerminkan kemampuan siswa.¹⁷

Soal yang dikembangkan dalam UAS umumnya berbentuk tes objek (pilihan ganda) dalam soal pilihan ganda dibagi menjadi dua sistem yaitu pernyataan (*stem*) dan alternatif jawaban (*option*). Soal *stem* biasanya berupa pernyataan atau pertanyaan, sedangkan soal *option* merupakan beberapa dari pilihan jawaban. Salah satu alternatif pilihan jawaban merupakan kunci jawaban, sedangkan yang lainnya disebut sebagai pengecoh (*distractors*). Soal yang baik harus memiliki pengecoh yang relatif homogen, sehingga tidak mudah ditebak oleh siswa.¹⁸ Teori ini dinamakan teori tes klasik atau disebut sebagai *Classical test Theory* (CTT) dalam proses melakukan analisis item, kelebihan dalam teori ini adalah taraf kesukaran dan daya diskriminasi item dapat diukur secara manual.¹⁹ Jika tes sulit artinya tingkat kemampuan peserta didik rendah, dengan kata lain jika tes diujikan pada peserta dengan kemampuan rendah, maka tingkat kesukaran tes akan menjadi tinggi. Sebaliknya jika tes mudah artinya tingkat kemampuan peserta didik tinggi, dengan kata lain tingkat kesukaran soal akan menjadi rendah ketika tes diujikan pada kelompok berkemampuan tinggi, kedua

¹⁷ Istiyono, *Pengembangan Instrumen...*, hlm. 88.

¹⁸ Dewi Juliah Ratnaningsih and Isfarudi Isfarudi, "Analisis Butir Tes Objektif Ujian Akhir Semester Mahasiswa Universitas Terbuka Berdasarkan Teori Tes Modern," dalam *Jurnal Pendidikan Terbuka Dan Jarak Jauh* 14, no. 2, 2013, hlm. 98–109.

¹⁹ Firmanto Adi Nurcahyo, "Aplikasi IRT Dalam Analisis Aitem Tes Kognitif," dalam *Jurnal Psikologi*, no. 2, 2017, hlm. 75.

tingkat kesukaran soal didefinisikan sebagai proporsi peserta didik dalam grup yang menjawab soal benar.

Hambleton, Swaminathan, & Rogers mengemukakan Teori tes klasik memang telah mendominasi dan banyak digunakan dalam dunia pengukuran beberapa tahun ini, hampir semua konsep validitas dan realibilitas yang dikenal saat ini merupakan pengembangan dari teori tes klasik, namun dalam hal ini teori tes klasik memiliki beberapa keterbatasan. Oleh karena itu mulai berkembang teori respon butir atau *Item Response Theory* (IRT) yang merupakan teori yang dikembangkan untuk memperbaiki keterbatasan yang ada di teori tes klasik. Teori respon butir merupakan salah satu cara untuk menilai kelayakan butir dengan membandingkan rerata penampilan butir terhadap tampilan bukti kemampuan yang diramalkan oleh model, tujuan utama dari teori respon butir dikembangkan adalah untuk mengatasi kelemahan dari teori tes klasik yang tidak independen terhadap sampel.²⁰

Terdapat 3 model teori respon butir, model pengukuran tersebut berdasarkan jumlah parameter butir yang dimasukkan ke dalam model, yaitu model satu parameter (1PL) adalah (b) tingkat kesukaran butir soal, model dua parameter (2PL) adalah (b) tingkat kesukaran butir soal dan (a) daya pembeda,

²⁰ Bambang Subali and Pujiyati Suyata, "Panduan Analisis Data Pengukuran Pendidikan Untuk Memperoleh Bukti Empirik Kesahihan Menggunakan Program *Quest*," (Yogyakarta: Lembaga Penelitian Dan Pengabdian Pada Masyarakat UNY, 2011), hlm. 1.

model tiga parameter adalah (b) tingkat kesukaran butir soal dan (a) daya pembeda (g) tebakan semu.²¹

Arikunto menyatakan analisis butir soal memiliki beberapa manfaat diantaranya: 1) mengidentifikasi butir-butir soal yang baik dan kurang baik, 2) dapat memperoleh informasi untuk memperbaiki tes yang digunakan baik dari segi isi dan konstruksinya, 3) memperoleh gambaran tentang keadaan tes yang disusun.²²

Konsep dasar teori respon butir adalah performansi subjek pada suatu tes yang dapat diprediksikan atau dijelaskan oleh seperangkat faktor yang disebut traits, latent traits atau abilitas serta hubungan antara performansi subjek pada suatu butir dan seperangkat kemampuan laten yang mendasarinya dapat digambarkan oleh suatu fungsi yang menaik secara monoton yang disebut sebagai kurva karakteristik butir (*Item characteristic curve-ICC*).²³

Pemahaman *Item Response Theory* (IRT) menilai perilaku seseorang dapat dijelaskan oleh karakteristik orang yang bersangkutan sampai pada batasan-batasan tertentu.²⁴ Hambleton dan Swaminathan menyatakan bahwa teori respon butir merupakan salah satu cara untuk menilai kelayakan butir dengan membandingkan rerata penampilan butir terhadap tampilan bukti kemampuan kelompok.²⁵ *Item Response Theory* (IRT) memiliki keunggulan yaitu *not group*

²¹ Heri retnawati, *Teori Respon Butir Dan Penerapannya Untuk Peneliti, Praktis Pengukuran Dan Penguji*, (Yogyakarta: Nuha Medika , 2014), hlm. 1.

²² Suharsimi Arikunto, *Prosedur Penelitian...*, hlm. 42.

²³ Ibid., hlm. 65.

²⁴ Djemari Mardapi, *Teknik Penyusunan...*, hlm. 88.

²⁵ Ratnaningsih and Isfarudi, *Analisis Butir...*, hlm. 99.

dependent and not item dependent. Sehingga kesulitan yang di alam ketika menggunakan teori res klasik bisa teratasi dengan metode statistik IRT yang mengestikasi peserta didik dengan tingkat kemampuan yang berbeda.²⁶

Teori respon butir menekankan pada probabilitas jawaban benar peserta tes, parameter butir dan parameter peserta tes dihubungkan melalui suatu fungsi matematis atau model formula matematis. Dalam formula ini kemungkinan peserta tes menjawab soal dipahami sebagai fungsi logistik perbedaan parameter yang dimasukkan ke dalam model.²⁷

Mata Pelajaran Sejarah Kebudayaan Islam di Madrasah Aliyah merupakan salah satu mata pelajaran yang menelaah tentang asal-usul, perkembangan, peranan kebudayaan/ peradaban Islam di masa lampau, mulai dari dakwah Nabi Muhammad pada periode Makkah dan periode Madinah, kepemimpinan umat setelah Rasulullah SAW wafat, sampai perkembangan Islam periode klasik (zaman keemasan) pada tahun 650 M– 1250 M, abad pertengahan/zaman kemunduran (1250 M–1800 M), dan masa modern/zaman kebangkitan (1800-sekarang), serta perkembangan Islam di Indonesia dan di dunia. Berdasarkan penjelasan di atas, pembelajaran Sejarah Kebudayaan Islam (SKI) pada hakikatnya adalah aktivitas transfer ilmu pengetahuan yang dilakukan oleh guru

²⁶ Mahanani, "Analisis Soal Internasional *Competitions and Assessments for Schools (ICAS)* Dengan Menggunakan Metode *Item Response Theory (IRT)* Dan *Classical Test (CTT)*", *Tesis*, (Semarang: Universitas Negeri Semarang, 2015), hlm. 3.

²⁷ M. Arif Hidayatulloh, "Analisis Kualitas Butir Soal Ulangan Semester Genap Bahasa Arab Kelas VIII Tahun Pelajaran 2012/2013 Dengan Program Quest Di SMP Muhammadiyah 3 Depok Sleman Daerah Istimewa Yogyakarta", *Tesis*, (Yogyakarta: Universitas Islam Negeri Sunan Kalijaga Yogyakarta, 2013), hlm. 11.

kepada siswa yang berhubungan erat dengan peristiwa masa silam, baik itu peristiwa politik, sosial, maupun ekonomi yang memang benar-benar terjadi dalam suatu negara Islam dan dialami oleh masyarakat Islam.

Penyusunan butir soal Penilaian Akhir Tahun Mata Pelajaran SKI kelas X MAN 2 Bantul dilakukan guru pengampu sendiri. Butir soal tersebut belum pernah dianalisis oleh pihak sekolah yakni guru mata pelajaran SKI, sehingga belum diketahui kualitas butir soal tersebut. Sedangkan sesuai dengan uraian penjelasan sebelumnya dapat diketahui bahwa evaluasi pembelajaran yakni tes berperan sangat penting untuk keputusan tindak lanjut kepada peserta didik mengenai ketercapaian tujuan pembelajaran. Persoalan yang akan dikaji oleh peneliti adalah berangkat dari aspek penilaian menggunakan soal yang dirancang sedemikian baiknya sehingga menimbulkan pertanyaan, apakah penilaian tersebut sudah sesuai dengan kemampuan siswa yang menjawab, hal ini juga berkaitan dengan tingkat kevalidan dari soal tes yakni sejauh mana tes tersebut benar mengukur aspek yang perlu diukur.²⁸

Penilaian adalah sebuah proses tafsiran segala informasi cara sistematis, berkala, komprehensif terkait dengan proses dan hasil tumbuh kembang yang dicapai oleh siswa melalui kegiatan belajar mengajar, dan menafsirkan hal tersebut sebagai pemutusan penilaian.²⁹ Dalam hal ini peneliti menggunakan sampel di MAN 2 Bantul Yogyakarta, dalam soal yang diujikan cenderung lebih

²⁸ Sudaryono, "Implementasi Teori...", hlm. 720.

²⁹ Ani Oktarina and Siti Fatonah, "Pengamatan Tentang Pembelajaran Dan Penilaian Pada Anak Usia Dini Di Era Pandemi Covid-19," dalam *Cakrawala Dini: Jurnal Pendidikan Anak Usia Dini* 12, no. 1, 2021, hlm. 32.

banyak menguji aspek ingatan. Peserta didik harus mulai dilatih berpikir tingkat tinggi, melatih peserta didik untuk terampil ini dapat dilakukan guru dengan cara melatih soal-soal yang sifatnya mengajak peserta didik berpikir dalam level analisis, sintesis dan evaluasi padahal guru memerlukan instrumen penilaian dalam bentuk soal-soal untuk melaksanakan penilaian, baik untuk menguji aspek kognitif, afektif, maupun psikomotor.

Terdapat perbedaan format dalam perangkat administrasi/manajemen pelaksanaan KBK pada dimensi perencanaan, pelaksanaan dan evaluasi yang digunakan guru pada lima kabupaten/Kota di lingkungan Dinas Pendidikan Propinsi Daerah Istimewa Yogyakarta.³⁰ Hal ini mengindikasikan bahwa kemampuan guru dalam membuat laporan hasil evaluasi belajar kurang memadai. Selama ini guru juga merasa kebingungan dalam membuat laporan tersebut, terutama dalam hal melakukan analisis butir soal. Oleh karena itu, perlu ada upaya untuk membantu guru dalam melakukan analisis soal pada hasil belajar peserta didik. Analisis butir soal meliputi analisis butir soal kuantitatif dan analisis butir soal kualitatif. Dalam penelitian ini akan dikaji analisis butir soal. Hasil analisis meliputi: tingkat kesukaran butir soal, daya pembeda soal, statistik sebaran jawaban, kehandalan/reliabilitas tes, kesalahan pengukuran (standar error), dan distribusi skor serta skor setiap peserta tes. Penelitian ini bertujuan untuk

³⁰ Maria Dominika Niron, *“Pendampingan Manajemen Pelaksanaan Kurikulum Berbasis Kompetensi di Sekolah se Propinsi DIY”*, (Yogyakarta: Departemen Pendidikan Nasional Universitas Negeri Yogyakarta, 2005), hlm.68.

mengetahui kualitas soal tes soal akhir tahun SKI Kelas X di MAN 2 Bantul Yogyakarta.

Soal Penilaian Akhir Tahun Mata Pelajaran SKI di MAN 2 Bantul terdiri dari 25 butir soal dalam bentuk pilihan ganda. Butir-butir soal tersebut disusun berdasarkan kompetensi dasar sebagai berikut:

| Kompetensi Dasar | Materi Pokok | Kelas/ Semester | Indikator Soal |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|---------------------------------------------------------------------------|
| 3.1. Mengevaluasi proses lahirnya Daulah Abbasiyah | Peradaban Islam Pada Masa Daulah Abbasiyah, Islam Pada Masa Daulah Usmani, Peradaban Islam Pada Masa Daulah Mughal Di India, Peradaban Islam Pada Masa Daulah Syafawi | XI/1&2 | Siswa mengetahui Tokoh pembaruan di Mesir |
| 3.2. Mengevaluasi perkembangan peradaban dan ilmu pengetahuan pada masa Daulah Abbasiyah | | | Siswa mengetahui latarbelakang Berdirinya Abbasiyyah |
| 3.3. Mengevaluasi proses lahirnya Daulah Usmani | | | Siswa mengetahui Tafsir Alquran modern Muhammad Rasyid Rida |
| 3.4. Mengevaluasi perkembangan peradaban dan ilmu pengetahuan pada masa Daulah Usmani | | | Siswa mengetahui pendapat Jamaludin Al-Afghani tentang dunia Nasrani |
| 3.5. Mengevaluasi sejarah lahirnya Daulah Mughal | | | Siswa mengetahui peninggalan sejarah Khalifah Harun Ar Rasyid |
| 3.6. Mengevaluasi perkembangan ilmu pengetahuan dan peradaban pada masa Daulah Mughal | | | Siswa mengetahui Pendapat Muhammad Iqbal tentang peradaban Barat |
| 3.7. Mengevaluasi proses lahirnya Daulah Syafawi | | | Siswa mengetahui khalifah yang paling berjaya pada masa Dinasti Usmani II |
| 3.8. Mengevaluasi perkembangan peradaban dan ilmu pengetahuan pada masa Daulah Syafawi | | | Siswa mengetahui khalifah dinasti Abbasiyah |
| 3.9. Menganalisis kemunduran Islam dan latar belakang munculnya gerakan tajdid | | | Siswa mengetahui pusat pemerintahan Abbasiyah |
| 3.10. Menganalisis tokoh-tokoh pembaruan dalam Islam dan ide-ide pembaruannya (Ali Pasha, Jamaluddin Al-Afghani, Muhammad Abduh, Rasyid Ridha, Muhammad Iqbal) | | | Siswa mengetahui ide pembaruan Muhammad Rasyid Rida |
| 3.11. Menganalisis nilai positif dari gerakan pembaruan dunia Islam | | | Siswa mengetahui majalah yang diterbitkan Muhammad Abduh |
| 3.12. Menganalisis pengaruh gerakan pembaruan terhadap perkembangan Islam di Indonesia | | | Siswa mengetahui ilmuan 4 madzhab |
| 3.13. Menganalisis munculnya organisasi Islam sebagai dampak dari adanya gerakan pembaruan | | | Siswa mengetahui Pembaruan Al-Afghani dengan gerakan Pan-Islamisme |
| 4.1 Menilai proses lahirnya Daulah Abbasiyah | Kemunduran Umat Islam, Gerakan Pembaruan Dalam Islam, Dan Pengaruh Pembaruan Islam Di Indonesia | | Siswa mengetahui kapan terjadinya perang salib |
| 4.2 Mengapresiasi perkembangan peradaban dan ilmu pengetahuan pada masa Daulah Abbasiyah | | | Siswa mengetahui ibrah imperialisme penjajahan Barat terhadap dunia Islam |
| 4.3 Menilai proses berdirinya Daulah Usmani | | | Siswa mengetahui pendiri daulah Usmani |

Tabel 1. Kompetensi Dasar Mata Pelajaran SKI Kelas X

Peneliti mengambil sampel ini karena pembuatan soal SKI kelas X di MAN 2 Bantul Yogyakarta dibuat oleh guru matapelajaran berdasarkan kisi-kisi yang disepakati oleh MGMP (Musyawarah Guru Mata Pelajaran) MA Daerah Istimewa Yogyakarta. Maka dalam hal ini untuk menjaga kualitas soal diperlukan analisis kualitas butir soal matapelajaran SKI di MAN 2 Bantul Yogyakarta, dan

penelitian ini diharapkan dapat digunakan sebagai acuan perbaikan soal dimasa mendatang.

B. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, peneliti merumuskan permasalahan kualitas soal akhir tahun SKI kelas X di MAN 2 Bantul Yogyakarta tahun 2020/2021 sebagai berikut :

1. Bagaimana validitas isi soal tersebut?
2. Bagaimana validitas empirik soal tersebut?
3. Berapa reabilitas soal tersebut?
4. Bagaimana analisis soal tersebut berdasarkan teori respon butir model 1PL?
5. Bagaimana analisis soal tersebut berdasarkan teori respon butir model 2PL?
6. Bagaimana analisis soal tersebut berdasarkan teori respon butir model 3PL?

C. Tujuan dan Manfaat Penelitian

Manfaat dan Tujuan penelitian kualitas soal akhir tahun SKI kelas X di MAN 2 Bantul Yogyakarta tahun 2020/2021 sebagai berikut antara lain:

1. Tujuan Penelitian

- a. Untuk menganalisis Validitas Isi Soal Akhir Tahun SKI Kelas X di MAN 2 Bantul Yogyakarta Tahun 2020/2021.

- b. Untuk menganalisis Validitas empirik soal Akhir Tahun SKI Kelas X di MAN 2 Bantul Yogyakarta Tahun 2020/2021.
- c. Untuk mencari nilai reabilitas soal Akhir Tahun SKI Kelas X di MAN 2 Bantul Yogyakarta Tahun 2020/2021.
- d. Untuk menganalisis analisis soal tersebut berdasarkan teori respon butir model 1PL.
- e. Untuk menganalisis analisis soal tersebut berdasarkan teori respon butir model 2PL.
- f. Untuk menganalisis analisis soal tersebut berdasarkan teori respon butir model 3PL.

2. Manfaat Penelitian

a. Manfaat Teoritis

- 1) Penelitian ini diharapkan menjadi sebuah karya ilmiah yang menambah hazanah pustaka dan literatur terutama pendidikan Islam.
- 2) Penelitian ini diharapkan dapat memberikan sumbangan pemikiran dalam pengembangan pendidikan terutama pendidikan Islam.

b. Manfaat Praktis

- 1) Bagi akademisi diharapkan penelitian ini dapat menjadi sumbangan dalam melakukan peningkatan kualitas penilaian pada soal tes.
- 2) Bagi pembaca diharapkan dapat menambah wawasan dan pengetahuan mengenai proses peningkatakn kualitas soal yang diujikan pada peserta tes.

- 3) Bagi guru, dapat digunakan sebagai bahan masukan mengenai pembuatan soal ujian akhir SKI bagi kelas X pada tahun-tahun berikutnya.
- 4) Sebagai bahan pertimbangan terhadap penelitian serupa pada tempat lain dan menjadi bahan rujukan bagi mereka yang berminat melakukan penelitian serupa dalam lingkup yang lebih luas dan mendalam dimasa yang akan datang.

D. Kajian Pustaka

Peneliti mengambil dari beberapa penelitian yang terkait dengan judul yang peneliti teliti diantaranya adalah:

Pertama, tesis dengan judul “Analisis Soal *Internasional Competitions And Assessments For Schools* (ICAS) Dengan Menggunakan Metode *Item Response Theory* (IRT) dan *Classical Test Theory* (CTT)”. Karya Mahanani mahasiswa Universitas Negeri Semarang tahun 2015. Hasil penelitiannya, peneliti masih menggunakan teori tes klasik dalam melakukan penilaian tes.³¹

Kedua, Jurnal penelitian dengan judul “Analisis Butir tes Objektif Ujian Akhir Semester Mahasiswa Universitas Terbuka Berdasarkan Teori Tes Modern”. Karya Dewi Juliah Ratnaningsih mahasiswa Ilmu matematika dan Pendidikan Alam Universitas Terbuka. Penelitian tersebut berfokus pada objek yang berdasarkan parameter rasch.³² Sedangkan dalam penelitian ini peneliti

³¹ Mahanani, Analisis Soal Internasional..., hlm.75.

³² Ratnaningsih and Isfarudi, “Analisis Butir...”, hlm. 59.

menggunakan data dari hasil peserta ujian SKI kelas X di MAN 2 Bantul Yogyakarta.

Ketiga, jurnal penelitian yang berjudul “Analisis Butir Soal Ujian Akhir Semester Matematika Dengan Teori Respon Butir” analisis kualitas dengan teori respon butir menggunakan model dua parameter logistic (2PL) yakni parameter daya beda (a) dan parameter tingkat kesukaran (b). Sedangkan dalam penilitian ini peneliti menggunakan parameter (1PL, 2PL dan 3 PL) yakni parameter daya beda (a) dan parameter tingkat kesukaran (b) dan tebak semu (g).

Keempat, Penelitian yang dilakukan oleh Abdul Mutholib dengan judul “Analisis Butir Soal Bahasa Arab Ujian Akhir Madrasah Berstandar Nasional (UAMBN) Bagi Siswa Madrasah Ibtidaiyah (MI) Tahun Pelajaran 2012-2013”. Dalam penelitian ini digunakan Analisis kuantitatif untuk mengetahui dan menganalisis validitas, realibilitas, tingkat kesulitan, daya beda, dan penyebaran pilihan jawaban atau distraktor butir soal mata pelajaran Bahasa Arab UAMBN MI Tahun 2013. Dengan bantuan program komputer, yakni program ANATES Versi 4.0.2.

Hasilnya dari segi validitas dan reliabilitas menunjukkan kriteria yang sangat tinggi. Dalam uji validitas tingkat kesesuaian 100% seluruh butir soal tersebut dengan kisi-kisi yang telah ditetapkan. Sedangkan dalam reliabilitas memiliki tingkat keajegan/konsistensi yang tinggi, dengan angka reliabilitas tes 0,73 (berada pada interval 0,60-0,80), kemudian dari segi tingkat kesulitan diketahui bahwa terdapat 8% butir soal dikategorikan sangat mudah; 26% butir

soal dikategorikan mudah; 52% butir soal sedang; 12% butir soal yang dikategorikan sukar; dan 2% butir soal sangat sukar. Kemudian dari segi daya beda ditemukan 46% butir soal dapat diterima baik; 18% butir soal dapat diterima tapi perlu diperbaiki; 12% butir soal masih perlu perbaikan; dan 24% butir soal yang tidak dipakai/dibuang.

Kelima, Penelitian yang dilakukan oleh Ahmad Nur Kholis dengan judul “Analisis Tingkat Kesulitan (Difficulty Level) Soal Pada Buku Sejarah Kebudayaan Islam Kelas 8 Kurikulum 2013 R.” Analisis butir soal juga digunakan dalam penelitian ini. Hasilnya menunjukkan bahwa soal mata pelajaran Sejarah Kebudayaan Islam (SKI) untuk kelas 8 MTs pada pembahasan: (1) Jejak Peradaban Dinasti Bani Abbasiyah; dan (2) Kecemerlangan Ilmuwan Muslim Dinasti Bani Abbasiyah, pada Buku Guru yang diterbitkan oleh Kementerian Agama tahun 2015 disajikan dengan proporsi: 100% soal dalam kategori sulit (sukar).

Keenam, Penelitian yang dilakukan oleh maolina Nurriska Setiawati dengan judul “Analisis Parameter Tes Penilaian Akhir Semester Fisika Kelas X Dengan Teori Respon Butir” Metode yang digunakan pada penelitian ini adalah metode deskriptif kuantitatif. Partisipan dalam penelitian ini adalah 215 peserta didik kelas X MIPA di salah satu SMA Kota Bandung. Seluruh data yang diperoleh dianalisis menggunakan teori respon butir model 2 PL. asil analisis menunjukkan bahwa soal penilaian akhir semester ini memiliki 2 butir soal yang tidak valid dari 25 butir soal pilihan ganda (butir ke-5 dan ke-12), nilai parameter

daya pembeda (a) sebesar 0,354 yang dikategorikan baik, dan nilai parameter tingkat kesukaran (b) sebesar -1,4 yang berada dalam kategori mudah. Soal penilaian akhir semester ini memiliki nilai puncak fungsi informasi sebesar 25,132 dan SEM (Standard Error of Measurement) sebesar 0,199 dengan perpotongan antara kedua kurva tersebut berada pada rentang -3,619 sampai 1,206 sehingga soal tersebut reliabel untuk mengukur kemampuan sangat rendah hingga tinggi.³³

Ketujuh Penelitian ini dilakukan Oleh Iqbal Faza Ahmad dan Sukiman yang berjudul “Analisis *Higher Order Thinking Skills* (HOTS) pada Soal Ujian Akhir Siswa Kelas 6 KMI Dalam Kelompok Mata Pelajaran Dirasah Islamiyah Di Pondok Modern Tazakka Batang” Metode yang digunakan dalam penelian ini kualitatif deskriptif analitik. Penelitian ini bertujuan mengetahui komposisi soal ujian akhir siswa kelas 6 KMI menggunakan analisis *higher order thinking skills* (HOTS). Perbedaan dengan penelitian ini, peneliti menggunakan analisis teori respon butir pada soal SKI akhir tahun 2020/2021.³⁴

Kedelapan Penelitian ini dilakukan oleh Iqbal Faza Ahmad dkk dengan judul “*Trends in Implementation of Higher-Order Thinking Skills in Islamic Religious Education in Madrasahs and Schools: A Systematic Literature Review*” metode yang digunakan studi literatur sistematis. Peneliti menemukan upaya

³³ Maulina Nurriska Setiawati, Purwanto Purwanto, and Taufik Ramlan Ramalis, “Analisis Parameter Tes Penilaian Akhir Semester Fisika Kelas X Dengan Teori Respon Butir”, dalam *Jurnal Pendidikan Fisika*, vol 7, no. 1, 2022, hlm. 9.

³⁴ Iqbal Faza Ahmad dan Sukiman Sukiman, “Analisis *Higher Order Thinking Skills* (HOTS) Pada Soal Ujian Akhir Siswa Kelas 6 KMI Dalam Kelompok Mata Pelajaran Dirasah Islamiyah Di Pondok Modern Tazakka Batang”, dalam *Jurnal Pendidikan Agama Islam*, vol. 16, no. 2, 2019, hlm. 64.

guru, pendidik dan ahli dalam menerapkan HOTS dalam pembelajaran PAI dapat dikelompokkan menjadi dua kategori. Pertama mereka melakukan inovasi proses pembelajaran berbasis HOTS, dan kedua mereka yang mengembangkan instrumen penilaian berbasis HOTS. Beberapa inovasi dalam proses pembelajaran dilakukan dengan menerapkan berbagai metode, model, dan strategi pembelajaran. Penerapan pembelajaran berbasis HOTS dalam pembelajaran Pendidikan Agama Islam dapat meningkatkan kualitas pembelajaran menjadi lebih efektif, efisien, menyenangkan, dan bermakna. Peningkatan ini kemudian berdampak pada peningkatan kualitas prestasi belajar. Penelitian ini memberikan sorotan dan membantu untuk memahami hasil penelitian sebelumnya yang membahas HOTS menggunakan tinjauan sistematis dalam pelaksanaan Pendidikan Agama Islam di Madrasah dan Sekolah.³⁵

Kesembilan penelitian ini dilakukan Hasnah yang berjudul “Analisis Kualitas Soal Matematika Ujian Sekolah Kelas XII IPA SMA Negeri di Watansoppeng Berdasarkan Teori Respon Butir” metode penelitian yang digunakan adalah kuantitatif deskriptif. Hasil dari penelitian ini adalah Pada Model Logistik 1 Parameter, untuk aspek tingkat kesukaran sebanyak 22.5% butir berada pada kategori baik, dan 77.5% berada pada kategori buruk. Soal Matematika Ujian Sekolah Kelas XII IPA SMA Negeri di Watansoppeng memberi informasi yang maksimal jika dikenakan pada peserta tes yang berkemampuan sekitar -4 hingga

³⁵ Iqbal Faza Ahmad, “Trends in the Implementation of Higher-Order Thinking Skills in Islamic Religious Education in Madrasahs and Schools: A Systematic Literature Review,” dalam *Jurnal Pendidikan Islam*, vol. 9, no. 2, 2020, hlm. 195–216.

-2. Untuk aspek daya beda, 50% butir berada pada kategori baik dan 50% berada pada kategori buruk.³⁶

Kesepuluh penelitian ini dilakukan oleh Nenden Annisa Rosidah, dkk yang berjudul “Karakteristik Tes Keterampilan Berpikir Kritis (KBK) Berdasarkan Pendekatan Teori Respon Butir” metode penelitian ini menggunakan metode campuran (*mixed methods*) dengan desain *concurrent embedded* dengan metode kuantitatif sebagai metode primer. Analisis data dalam penelitian ini didasarkan pada respons 108 siswa yang telah mendapatkan pembelajaran tentang suhu dan kalor kelas X SMA di Tasikmalaya dan Bandung. Hasil analisis dari kurva karakteristik tes menunjukkan bahwa indeks kesukaran (*bjk*) terentang antara -1,49 sampai 1,34 dengan indeks kesukaran yang berbeda pada tiap kategorinya. Perpotongan antara total fungsi informasi dan SEM (*Standard error measurement*) berada pada rentang -1,70 sampai 1,20. Hal ini berarti, keseluruhan tes ini dapat digunakan untuk mengukur kemampuan berpikir kritis siswa dari tingkat kemampuan rendah (di bawah rata-rata) sampai kemampuan tinggi (di atas rata-rata).³⁷

³⁶ Hasnah Hasnah, “Analisis Kualitas Soal Matematika Ujian Sekolah Kelas XII IPA SMA Negeri Di Watansoppeng Berdasarkan Teori Respon Butir,” dalam *Jurnal PEP Educational Assessment*, vol. 1, no. 1, 2017, hlm. 27–34.

³⁷ Nenden Annisa Rosidah, Taufik Ramlan Ramalis, dan Iyon Suyana, “Karakteristik Tes Keterampilan Berpikir Kritis (KBK) Berdasarkan Pendekatan Teori Respon Butir,” dalam *Jurnal Inovasi Dan Pembelajaran Fisika*, vol. 5, no. 1, 2018, hlm. 77.

E. Kerangka Teori Analisis Butir Soal

1. Analisis Butir Soal

Anas Sudijono berpendapat bahwa soal dianggap berkualitas jika butir soalnya mampu menjalankan fungsinya sebagai alat ukur. Untuk mengetahui apakah tes tersebut menjalankan fungsinya menurut Wahidamurni harus memenuhi 2 indikator yaitu validitas dan realibilitas skor yang dihasilkan.³⁸ Untuk menentukan validitas dan realibilitas skor dilakukan analisis butir soal. Bermutu atau tidaknya butir-butir soal tes hasil belajar juga dapat diketahui dari derajat kesukaran atau tingkat kesulitan yang dimiliki masing-masing butir soal tersebut. Butir soal tes hasil belajar dapat dikatakan sebagai butir soal yang baik apabila butir tersebut tidak terlalu sukar dan tidak pula terlalu mudah dengan kata lain, derajat kesukaran butir soal adalag sedang dan cukup.³⁹

Butir soal tes yang baik juga harus dapat menunjukkan daya pembeda, daya pembeda adalah kemampuan soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang kurang (berkemampuan rendah), menurut Anastasi dan Urbani daya beda berhubungan dengan derajat kemampuan butir membedakan dengan baik perilaku pengambil tes dalam tes yang dikembangkan. Soal dapat dikatakan mempunyai daya pembeda jika soal tersebut dapat dijawab oleh berkemampuan tinggi dan tidak dapat dijawab oleh siswa berkemampuan rendah. Jika suatu soal data dikerjakan oleh siswa yang

³⁸ Wahidmurni, *Evaluasi Pembelajaran, Kompetensi Dan Praktik*, (Yogyakarta: Nuha Litera, 2010), hlm. 84.

³⁹ Anas Sudijono, *Pengantar Evaluasi...*, hlm. 370.

berkemampuan tinggi maupun siswa yang berkemampuan rendah, berarti soal tersebut tidak mempunyai daya beda demikian pula jika soal tersebut tidak bisa dijawab oleh siswa berkemampuan tinggi dan siswa yang berkemampuan rendah, berarti soal tersebut tidak baik, sebab tidak mempunyai daya pembeda.⁴⁰

Kamus Besar Bahasa Indonesia (KBBI) Analisis adalah penyelidikan terhadap suatu peristiwa (seperti karangan, perbuatan, kejadian atau peristiwa) untuk mengetahui keadaan sebenarnya. Analisis soal merupakan suatu kegiatan yang harus dilakukan pendidik untuk meningkatkan mutu soal yang telah tertulis. Kegiatan ini merupakan proses pengumpulan, peringkasan, dan penggunaan informasi dari jawaban siswa untuk membuat keputusan tentang setiap penilaian.⁴¹ Menurut Arikunto analisis soal merupakan prosedur sistematis yang akan memberikan informasi-informasi yang sangat khusus terhadap butir yang disusun.⁴²

Analisis butir soal dilakukan dengan tujuan untuk meningkatkan kualitas soal, yaitu apakah soal dapat diterima karena telah didukung data statistik yang memadai. Diperbaiki karena terbukti beberapa kelemahan. Atau bahkan tidak digunakan sama sekali karena secara empiris tidak berfungsi sama sekali. Analisis dilakukan dengan validitas isi dan reliabilitas.⁴³

⁴⁰ Purwanto, *Evaluasi Hasil Belajar...*, hlm. 102.

⁴¹ *Ibid.*, hlm. 117.

⁴² Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, (Jakarta: PT. Bumi aksara, 2000), hlm. 205.

⁴³ Mahanani, *Analisis Soal Internasional...*, hlm. 41–42.

Menurut Sudaryono terdapat butir soal yang memiliki ciri dapat dijawab dengan betul oleh kebanyakan peserta tes yang berkemampuan tinggi dan tidak dapat dijawab dengan betul oleh peserta tes yang berkemampuan rendah, butir soal yang memiliki ciri diatas, memiliki beberapa kategori dalam pembuatannya, sebagai berikut:⁴⁴ Butir demikian memiliki daya untuk membedakan responden berdasarkan kemampuan peserta tes dan Butir memiliki parameter berupa daya beda butir soal. Daya beda butir tes adalah kemampuan suatu butir soal dapat membedakan peserta tes yang telah menguasai materi yang ditanyakan dan peserta tes yang belum menguasai materi yang ditanyakan.⁴⁵

2. Validitas Isi/Konten

Instrument evaluasi mempunyai karakteristik yang sangat penting yaitu valid (*validity*). Menurut Gronlund valid dapat diartikan sebagai ketetapan intepretasi yang dihasilkan dari skor tes atau instrument evaluasi. Validitas suatu instrument evaluasi, tidak lain adalah derajat yang menunjukkan dimana suatu tes mengukur apa yang hendak diukur.⁴⁶

Validitas adalah alat ukur untuk mengetahui sejauhmana tes tersebut dapat mengukur apa yang hendak diukur. Secara umum ada tiga pendekatan dalam meneliti validitas yaitu: validitas isi dan validitas konstruk. Validitas isi yaitu validitas yang fokus pada elemen apa yang ada dalam ukuran sehingga analisis rasional adalah proses utama yang dilakukan dalam analisis validitas isi. Validitas konstruk

⁴⁴ Sudaryono, *Teori Respon Butir*, (Yogyakarta: Graha Ilmu, 2013), hlm. 3.

⁴⁵ Hasnah, "Analisis Kualitas...", hlm. 30.

⁴⁶ Sukardi, *Evaluasi Pendidikan Prinsip & Operasionalnya*, (Yogyakarta: PT Bumi Aksara, 2012), hlm. 1.

mengandung arti bahwa suatu alat ukur (dikatakan valid apabila telah cocok dengan konstruksi teoritik di mana tes itu dibuat). Sebuah tes dikatakan memiliki validitas konstruk apabila soal–soalnya mengukur setiap aspek berpikir seperti yang diuraikan dalam standar kompetensi, kompetensi dasar, maupun indikator yang terdapat dalam kurikulum.

dalam hal ini validitas ini juga mengukur sejauh mana elemen instrument asesmen yang relevan dan mewakili konstruk alat ukur yang ditargetkan untuk tujuan tertentu. Instrumen asesmen adalah alat yang akan digunakan untuk menerapkan semua metode validitas isi. Elemen dari instrument asesmen adalah semua aspek proses pengukuran yang dapat mempengaruhi data yang diperoleh. Misalnya, elemen kuesioner seperti item pernyataan, format jawaban, dan instruksi. Elemen observasi perilaku seperti kode observasi, parameter sampling-waktu, dan situasi dimana observasi terjadi.⁴⁷

Hasil akhir dari validitas isi adalah penilaian tentang kelayakan isi tes. Lawshe mengajukan cara yang disebut rasio validitas isi sebagai hasil ukur penilaian ahli atas kesetujuannya terhadap validitas isi yang menggunakan teknik statistik kesamapahaman antar penilai, sejauhmana juga bisa ditunjukkan dengan seberapa detailnya penjelasan tentang alat ukur tersebut, semakin detail maka akan semakin diyakini bahwa alat ukur itu makin tepat.⁴⁸ validitas isi dalam permasalahan ini sangat penting menggunakan instrument menggunakan formula validitas yang tepat. Formula

⁴⁷ Helli Ihsan, “Validitas Isi Alat Ukur Penelitian: Konsep Dan Panduan Penilaiannya,” dalam *Jurnal Pedagogia*, vol. 13, no. 3, 2015, hlm. 173.

⁴⁸ *Ibid.*, hlm. 174.

instrument validitas menggunakan formula Gregory hal ini dikarenakan validator yang dilibatkan hanya 2 orang yaitu pakar dibidang pendidikan. Mekanisme perhitungan validitas isi menurut Gregory adalah sebagai berikut:

- a) Setiap butir instrument dinilai dengan menggunakan skor 1,2,3 dan 4.
- b) Dilakukan pengelompokkan skor, yaitu skor 1 dan 2 dikelompokkan dalam kurang relevan, sedangkan skor 3 dan 4 dikelompokkan dalam sangat relevan.

Hasil penilaian ditabulasikan dalam bentuk matriks. Berikut formula validitas isi dilakukan tabulasi silang yang terdiri dari empat kolom yang ditunjukkan table berikut:⁴⁹

| | | Validator 1 | |
|-------------|---------------------------|---------------------------|---------------------------|
| | | Kurang relevan (skor 1-2) | Sangat relevan (Skor 3-4) |
| Validator 2 | Kurang relevan (skor 1-2) | A | B |
| | Sangat relevan (Skor 3-4) | C | D |

Tabel 2. Tabulasi Silang Hasil Penilaian Kedua Validator

Rumus untuk menghitung validitas isi instrument validasi isi adalah:

$$\text{Validitas Isi: } D / (A+B+C+D)$$

Keterangan:

A = sel yang menunjukkan ketidak setujuan antara kedua penilaian

B dan C = sel yang menunjukkan perbedaan pandangan antara penilaian

D= sel yang menunjukkan persetujuan yang valid antara kedua penilaian.

⁴⁹ Maade Candiasa, "Statistik Univariat Dan Bivariat Disertai Aplikasi SPSS", (Singaraja: Unit Penerbitan Universitas Pendidikan Ganesha, 2010), hlm. 110.

| Nilai | Interpretasi |
|------------|---------------|
| 0,00-0,19 | Sangat Rendah |
| 0,20 -0,39 | Rendah |
| 0,40- 0,59 | Sedang |
| 0,60 -0,79 | Tinggi |
| 0,80-1 | Sangat Tinggi |

Tabel 3. Kriteria Validitas

3. Taksonomi Bloom

Bloom mengklasifikasikan hasil belajar kedalam tiga bagian besar yaitu ranah kognitif, efektif dan psikomotorik. Bloom kemudian membagikan ranah kognitif ke dalam jenjang atau level proses berfikir, bloom membagi ranah tersebut menjadi enam bagian dari jenjang yang terendah sampai dengan yang paling tinggi, pengetahuan (*Knowledge*), pemahaman (*Comprehension*), penerapan (*Application*), analisis (*Analysis*), sentesis (*Synthesis*) dan evaluasi (*evaluation*).⁵⁰ Taksonomi Bloom dalam ranah kognitif selanjutnya mengalami perubahan dan penyempurnaan setelah dilakukan revisi oleh Anderson dan Karthwohl pada tahun 2000. Pada awalnya Bloom menggunakan kata benda dalam merumuskan jenjang proses kognitif yaitu pengetahuan, pemahaman, terapan, analisis, sintesis, dan evaluasi. Setelah mengalami revisi dan penyempurnaan menjadi mengingat, memahami, menerapkan, menganalisis, mengevaluasi, dan mencipta.⁵¹

⁵⁰ Nursalam Nursalam, Pengukuran Dalam..., hlm. 11.

⁵¹ Basuki dan Hariyanto, *Asemen Pembelajaran*, (Bandung: PT Remaja Rosdakarya, 2016), hlm.

Dalam revisi tersebut pengetahuan menjadi dimensi baru dalam ranah kognitif yang terpisah dengan dimensi proses kognitif.

| Tingkatan | Taksonomi Bloom (1956) | Anderson dan Karthwohl (2000) |
|-----------|------------------------|-------------------------------|
| C1 | Pengetahuan | Mengingat |
| C2 | Komprehensi | Memahami |
| C3 | Aplikasi | Menerapkan |
| C4 | Analisis | Menganalisis |
| C5 | Sintesis | Mengevaluasi |
| C6 | Evaluasi | Mengkreasi/mencipta |

Tabel 4. Revisi Taksonomi Bloom

Anderson dan Karthwohl dalam merevisi ranah kognitif taksonomi Bloom mengklasifikasikan keenam jenjang proses kognitif menjadi tiga kategori yaitu kemampuan berpikir tingkat rendah (*Lower order thinking skill*) meliputi proses berpikir dalam jenjang mengingat, kemampuan berpikir tingkat menengah (*Middle order thinking skill*) meliputi jenjang proses berpikir memahami dan menerapkan dan yang terakhir kemampuan berpikir tingkat tinggi (*Higher order thinking skill*) yang meliputi jenjang proses berpikir menganalisis, mengevaluasi dan mencipta.

Anderson dan Krathwohl mengemukakan masing-masing indikator dalam taksonomi Bloom (revisi) sebagai berikut:⁵²

⁵² Imam Gunawan dan Anggarini Retno, Taksonomi Bloom- Revisi Ranah Kognitif: Kerangka Landasan Untuk Pembelajaran, Pengajaran, dan Penilaian, dalam *Jurnal Pendidikan Dasar dan Pembelajaran*, Vol. 2, No. 02, 2012, hlm. 98.

1) Mengingat

Proses mengingat adalah proses mengambil pengetahuan yang dibutuhkan dari memori jangka panjang. Jika yang menjadi tujuan pembelajarannya adalah menumbuhkan kemampuan untuk meretensi materi pelajaran sama seperti materi yang diajarkan, maka mengingat adalah kategori kognitif yang tepat.

2) Memahami

Memahami merupakan proses mengkonstruksi makna dari pesan-pesan pembelajaran, yang disampaikan melalui pengajaran, buku, atau layar komputer. Peserta didik memahami ketika mereka menghubungkan pengetahuan baru dan pengetahuan lama atau pengetahuan baru dipadukan dengan kerangka kognitif yang telah ada.

3) Mengaplikasikan

Proses kognitif mengaplikasikan melibatkan penggunaan prosedur-prosedur tertentu untuk mengerjakan soal latihan atau menyelesaikan masalah. Kategori ini terdiri dari dua proses kognitif, yaitu mengeksekusi untuk tugas yang hanya berbentuk soal latihan dan mengimplementasikan untuk tugas yang merupakan masalah yang tidak familier.

4) Menganalisis

Menganalisis melibatkan proses memecah materi menjadi bagian-bagian kecil dan menentukan bagaimana hubungan antar bagian-bagian dan struktur keseluruhannya. Kategori proses menganalisis ini meliputi proses kognitif membedakan, mengorganisasi, dan mengatribusikan.

5) Mengevaluasi

Mengevaluasi didefinisikan sebagai membuat keputusan berdasar kriteria dan standar. Kriteria-kriteria yang sering digunakan adalah kualitas, efektivitas, efisiensi, dan konsistensi. Masing-masing dari kriteria tersebut ditentukan oleh peserta didik. Standar yang digunakan bisa bersifat kuantitatif maupun kualitatif. Kategori mengevaluasi mencakup proses kognitif memeriksa (keputusan yang diambil berdasarkan kriteria internal) dan mengkritik (keputusan yang diambil berdasarkan kriteria eksternal).

6) Mencipta

Mencipta melibatkan proses menyusun elemen-elemen menjadi sebuah keseluruhan yang koheren atau fungsional. Tujuan yang diklasifikasikan dalam proses mencipta menuntut peserta didik membuat produk baru dengan mereorganisasi sejumlah elemen atau bagian menjadi suatu pola atau struktur yang tidak pernah ada sebelumnya. Proses kognitif yang terlibat dalam mencipta pada umumnya sejalan dengan pengalaman belajar yang telah dimiliki sebelumnya. Proses kognitif tersebut yaitu merumuskan, merencanakan, dan memproduksi.

| Kategori Kognitif | Sub Kategori |
|-------------------|------------------------------------------|
| 1. Mengingat | 1.1. Mengenali 1.2. Mengingat kembali |

| | |
|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| 2. Memahami | 2.1. Menafsirkan 2.2. Mencontohkan 2.3. Mengklasifikasikan 2.4. Merangkum 2.5. Menyimpulkan 2.6. Membandingkan 2.7. Menjelaskan |
| 3. Menerapkan | 3.1. Mengeksekuis 3.2. Mengimplementasikan |
| 4. Menganalisis | 4.1. Membedakan 4.2. Mengorganisasi 4.3. Mengatribusi |
| 5. Mengevaluasi | 5.1. Memeriksa 5.2. Mengkritik |
| 6. Mencipta | 6.1. Merumuskan 6.2. Merencanakan 6.3. Memproduksi |

Tabel 5. Kategori dalam Proses Kognitif

4. Validitas Empiris

Istilah “validitas empiris” memuat kata “empiris” yang artinya “pengalaman”. Validitas empiris tidak dapat diperoleh hanya dengan menyusun instrument berdasarkan ketentuan seperti halnya validitas logis, tetapi harus dibuktikan melalui pengalaman.⁵³ Uji validitas empiris *product moment* dengan SPSS, penelitian yang dilakukan dengan menggunakan metode ini perlu diuji validitas. Uji validitas berguna untuk mengetahui kevalidan atau kesesuaian angket yang digunakan oleh peneliti untuk memperoleh data dari para responden atau sampel penelitian. Uji validitas *product moment pearson correlation* menggunakan prinsip mengkorelasikan atau menghubungkan antara masing-

⁵³ Suharsimi Arikunto, *Dasar-Dasar Evaluasi...*, hlm. 28.

masing skor item atau soal dengan skor total yang diperoleh dari jawaban responden. .⁵⁴

Dasar pengambilan keputusan dilakukan melalui beberapa cara yaitu: Membandingkan nilai r hitung dengan r tabel.⁵⁵

- Jika nilai r hitung $>$ r tabel, maka item soal angket tersebut dinyatakan valid
- Jika nilai r hitung $<$ r tabel, maka item soal angket tersebut dinyatakan tidak valid.

Membandingkan nilai Sig. (2-tailed) dengan Probabilitas 0,05

- Jika nilai Sig. (2-tailed) $<$ 0,05 dan Pearson Correlation bernilai positif, maka item soal angket tersebut valid
- Jika nilai Sig. (2-tailed), 0,05 dan Pearson Correlation bernilai negative, maka item soal angket tersebut tidak valid
- Jika nilai Sig. (2-tailed) $>$ 0,05, maka item soal angket tersebut tidak valid.

5. Reliabilitas

Reliabilitas merupakan metode yang mempelajari, mengidentifikasi, dan mengestimasi keajegan atau ketidakajegan skor tes.⁵⁶ Menurut Zainul Arifin reliabilitas merupakan tingkat atau derajat konsistensi dari suatu instrumen. Reliabilitas tes berkenaan dengan pertanyaan apakah suatu tes teliti dan dapat dipercaya sesuai dengan kriteria yang ditetapkan. Suatu tes dapat dikatakan reliabel

⁵⁴ Zulkifli Matondang, "Validitas Dan Reliabilitas Suatu Instrumen Penelitian", dalam *Jurnal Tabularasa*, vol. 6, no. 1, 2009, hlm. 87–97.

⁵⁵ *Ibid.*, hlm. 90.

⁵⁶ M. Arif Hidayatulloh, *Analisis Kualitas Butir...*, hlm. 61.

jika selalu memberikan hasil yang sama bila diteskan pada kelompok yang sama pada waktu atau kesempatan yang berbeda.⁵⁷

Reliabilitas dapat dinyatakan sebagai tingkat keajegan atau kemantapan hasil dari pengukuran terhadap hal yang sama. Reliabilitas mempunyai dua keajegan, keajegan yang pertama adalah keajegan internal yakni tingkat sejauhmana butir soal itu homogen baik dari tingkat kesukaran maupun bentuk soalnya, yang kedua adalah keajegan eksternal yaitu tingkat sejauhmana skor yang dihasilkan tetap sama sepanjang kemampuan siswa yang diukur belum berubah.⁵⁸

Sehingga suatu instrumen evaluasi, dikatakan mempunyai nilai reliabilitas tinggi, apabila tes yang dibuat mempunyai hasil yang konsisten dalam mengukur yang hendak diukur. Reliabilitas memberikan konsistensi yang membuat terpenuhinya syarat utama, yaitu validnya suatu hasil skor instrumen. Semakin reliabel sebuah tes maka semakin dikatakan bahwa hasil tes tersebut akan mempunyai hasil yang sama dan ketika dipakai kembali.⁵⁹

Azwar menyatakan bahwa reliabilitas merupakan salah satu ciri atau karakter utama instrumen pengukuran yang baik.⁶⁰ Arifin menyatakan bahwa suatu tes dikatakan reliabel jika selalu memberikan hasil yang sama bila diteskan pada

⁵⁷ Zainal Arifin, *Evaluasi Pembelajaran Prinsip Teknik Prosedur*, (Bandung: PT Remaja Rosdakarya, 2011), hlm. 258.

⁵⁸ Achmad Amru, "Penerapan Metode Kuder-Richardson (KR 20) Dan Naive Bayes Classifier Dalam Analisis Butir Soal Hasil Ujian Tengah Semester" *Tesis*, (Malang: Universitas Islam Negeri Maulana Malik Ibrahim, 2014), hlm. 22.

⁵⁹ Djaali dkk, *Pengukuran Dalam Bidang Pendidikan*, (Jakarta: Program Pascasarjana, 2000), hlm. 81.

⁶⁰ Saifudin Azwar, *Sikap Manusia Terori Dan Pengukurannya*, (Yogyakarta: Pustaka Pelajar, 2003), hlm. 172.

kelompok yang sama pada waktu atau kesempatan yang berbeda.⁶¹ Reliabilitas sebagai alat penilaian dilihat dari ketepatan atau keajegan alat tersebut dalam menilai apa yang dinilainya, artinya kapanpun alat penilaian tersebut digunakan akan memberikan hasil yang relatif sama.⁶²

Uji reliabilitas menggunakan uji Kuder-Richardson (KR 20) tinggi rendahnya reliabilitas secara empiris ditunjukkan oleh suatu angka yang disebut koefisien reliabilitas yang berkisar antara 0.00 – 1,00. Metode Kuder- Richardson 20 (K-R 20) merupakan koefisien reliabilitas yang dapat menggambarkan butir-butir untuk jawaban benar/salah yang diberi skor 0-1. Koefisien reliabilitas Kuder-Richardson 20 dapat dihitung dengan menggunakan rumus berikut.⁶³

$$r_{11} = \frac{K}{K-1} \left(\frac{S^2 - \sum pq}{S^2} \right)$$

r_{11} = Reliabilitas tes

p = Proporsi peserta tes menjawab benar

q' = Proporsi peserta tes menjawab salah ($q=1-p$)

$\sum pq$ = Jumlah perkalian antara p dan q

k = Banyaknya butir

s = Standar deviasi atau simpangan baku

⁶¹ Zainal Arifin, *Evaluasi Pembelajaran ...*, hlm. 116.

⁶² Nana Sudjana, *Penilaian Hasil Proses Belajar Mengajar*, (Bandung: Remaja Risdakarya, 2004), hlm. 16.

⁶³ Edi Istiyono, *Penilaian Dan Analisis Hasil Belajar Fisika Dengan Teori Tes Klasik Dan Modern* (Yogyakarta: UNY Press, 2020), hlm. 403.

Terdapat beberapa langkah yang harus ditempuh dalam perhitungan reliabilitas menggunakan Kuder-Richardson 20 yaitu:

- a) Membuat tabel persiapan perhitungan reabilitas KR-20.
- b) Menentukan proporsi peserta tes menjawab benar (p) untuk setiap butir .
- c) Menentukan proporsi peserta tes menjawab salah (q) dengan rumus $q = 1-p$.
- d) Menentukan varian total
- e) Memasukkan semua bilangan yang telah diperoleh pada langkah-langkah di atas kedalam persamaan KR-20.
- f) Menginterpretasikan hasil koefisien reliabilitas yang diperoleh

Dalam hal ini nilai (r_i) diartikan sebagai nilai reliabilitas, sehingga oleh Guilford kriteriannya disajikan dalam tabel berikut:

| Nilai | Interpretasi |
|---------------------------|---------------|
| $r_i < 0,20$ | Sangat Rendah |
| $0,20 \leq r_i < 0,40$ | Rendah |
| $0,40 \leq r_i < 0,60$ | Sedang |
| $0,60 \leq r_i < 0,80$ | Tinggi |
| $0,80 \leq r_i \leq 1,00$ | Sangat Tinggi |

Tabel 6. Kriteria Reliabilitas

6. Teori Respon Butir (*Item Response Theory*)

Teori Respon Butir (*Item Respon Theory*) bertujuan untuk mengatasi kelemahan dari teori klasik, pada teori respon butir peluang menjawab benar yang diberikan siswa, parameter butri, dan parameter peserta tes dihubungkan menggunakan metode formula yang harus ditaati baik oleh kelompok butir tes maupun siswa tes, yang artinya pada proses invariansi di antara butir tes dan peserta tes adalah butir yang

sama terhadap peserta tes yang berbeda harus patuh pada rumus.⁶⁴ Asal mula Teori Respon Butir (*Item Response Theory*) adalah kombinasi suatu versi hukum phi-gamma dengan suatu analisis faktor butir soal (item faktor analisis) kemudian bernama teori trait latent (*Latent trait theory*), kemudian sekarang secara umum dikenal menjadi teori jawaban butir soal (*Item Response Theory*).⁶⁵

Perbedaan yang mendasar antara teori tes klasik dengan teori respon butir terletak pada invariansi penskoran, dimana untuk penskoran teori respon butir adalah invarians (tidak berubah atau tetap) terhadap butir tes soal serta terhadap peserta tes.⁶⁶ Invariansi parameter butir soal melalui kelompok peserta tes merupakan karakteristik yang paling penting dari teori respon butir, banyak yang berpendapat bahwasannya indeks kesukaran butir tes sebagai proporsi jawaban yang benar sehingga sukar untuk membayangkan bagaimana indeks kesukaran tes dapat menjadi invariant terhadap kelompok peserta tes dari tingkat kemampuan yang berbeda.⁶⁷ Pada pengukuran teori respon butir, taraf sukar butir dikaitkan langsung dengan karakteristik butir, kemampuan tinggi dan rendahnya memiliki taraf sukar butir yang sama, kemampuan responden dan taraf sukar butir menjadi independent dan pengukuran teori respon butir dapat digunakan untuk pencocokan kemampuan responden dengan taraf sukar.⁶⁸ Analisis teori respon butir pada umumnya

⁶⁴ Ibid., hlm. 289.

⁶⁵ Wahidmurni, *Evaluasi Pembelajaran...*, hlm. 18.

⁶⁶ La Ode, "Analisis Perangkat Tes Kimia Ujian Sekolah SMA Tahun Ajaran 2004-2005 di kab. Buton", *Tesis*, (Yogyakarta: Pasca Sarjanan Universitas Negeri Yogyakarta, 2006), hlm. 79.

⁶⁷ Ibid., hlm. 78.

⁶⁸ Sudaryono, "Implementasi Teori Responsi Butir (*Item Response Theory*) Pada Penilaian Hasil Belajar Akhir Di Sekolah, hlm. 719.

dimaksudkan untuk mengetahui besar kecilnya indeks tingkat kesukaran, tebak semu dan daya beda pada butir soal yang bersangkutan.⁶⁹

Asal mula Teori Respon Butir (*Item Response Theory*) adalah kombinasi suatu versi hukum phi-gamma dengan suatu analisis faktor butir soal (item faktor analisis) kemudian bernama teori trait latent (*Latent trait theory*), kemudian sekarang secara umum dikenal menjadi teori jawaban butir soal (*Item Response Theory*).⁷⁰ Menurut Hambleton Swaminathan dan Rogers, pemikiran teori respon butir didasarkan pada dua buah postulat yaitu:⁷¹

- a) Prestasi subyek pada suatu butir soal dapat diprediksikan dengan seperangkat faktor yang disebut kemampuan laten (*latent traits*),
- b) Hubungan antara prestasi subyek pada suatu butir dan perangkat kemampuan yang mendasari sesuai dengan grafik naik monoton tertentu yang disebut kurva karakteristik butir (*item characteristic curve, ICC*). Kurva karakteristik butir ini menggambarkan bahwa semakin tinggi level kemampuan peserta tes, semakin meningkat pula peluang menjawab benar suatu butir.

Adapun kelebihan dari Teori Respon Butir (*Item Response Theory*) adalah:

- a) Teori Respon Butir (*Item Response Theory*) tidak berdasarkan grup dependent.
- b) Skor peserta didik dideskripsikan bukan test dependent.
- c) Model ini menekankan pada tingkat butir soal bukan tes.

⁶⁹ Sumadi Suryabrata, "Pengembangan Alat Ukur Psikologis", (Yogyakarta: Andi Offset, 2000), hlm. 32.

⁷⁰ Wahidmurni, *Evaluasi Pembelajaran...*, hlm. 129.

⁷¹ M. Arif Hidayatulloh, *Analisis Kualitas Butir...*, hlm. 59.

- d) Teori Respon Butir (*Item Response Theory*) tidak memerlukan perarel tes untuk menentukan reliabilitas tes.
- e) Teori Respon Butir (*Item Response Theory*) suatu model yang memerlukan suatu pengukuran ketetapan untuk setiap skor tingkat kemampuan.

Tujuan utama Teori Respon Butir (*Item Response Theory*) adalah memberikan kesamaan adalah 1) asumsi banyak soal yang diukur banyak soal yang diukur pada trait yang sama, perkiraan tingkat kemampuan peserta didik adalah independent, (2) asumsi pada populasi tingkat kesukaran, daya pembeda merupakan independent sampel yang menggambarkan untuk tujuan kalibrasi soal, (3) dan statistik yang digunakan untuk menghitung tingkat kemampuan peserta didik diperkirakan dapat terlaksana.⁷² Jadi Teori Respon Butir (*Item Response Theory*) merupakan hubungan antara probabilitas jawaban suatu butir soal yang benar dan kemampuan siswa atau tingkatan/level prestasi siswa. Namun kelemahan bekerja dengan model Teori Respon Butir (*Item Response Theory*) adalah bekerja melalui suatu proses yang sulit karena kelebihan Teori Respon Butir (*Item Response Theory*) adalah: (1) tanpa varian pada parameter butir soal, (2) tanpa varian pada parameter abilitas, (3) adanya ketepatan pada pengukuran lokal.

Lord dalam Sudaryono menyatakan invariansi pensekoran pada teori respon butir adalah invarians (tidak berubah atau tetap) terhadap butir tes serta terhadap

⁷² Sudaryono, "Implementasi Teori Responsi...., hlm. 722.

peserta tes. Invariansi parameter-parameter butir tes melalui kelompok peserta tes merupakan karakteristik yang paling penting dari teori respon butir.⁷³

Model logistik satu parameter (1-PL-Model), model ini dikembangkan oleh rash sehingga model ini disebut sebagai model rasch. Model ini hanya memiliki satu parameter butir untuk menunjukkan karakteristik butir soal yaitu parameter tingkat kesulitan butir atau indeks kesulitan butir (b_i). parameter inilah yang akan mempengaruhi atau menentukan karakteristik (kemampuan) peserta tes. Dengan kata lain, tingkat kemampuan seorang siswa dapat diukur atau ditentukan dengan parameter tingkat kesulitan butir soal tersebut.⁷⁴

Perhitungan tingkat kesukaran butir soal adalah pengukuran seberapa besar derajat kesukaran soal, jika soal memiliki tingkat kesukaran yang seimbang maka dapat dikatakan bahwa soal tersebut baik.⁷⁵ Soal dengan tingkat kesukaran yang mudah tidak merangsang siswa untuk mempertinggi usaha untuk memecahkan soal tersebut, sebaliknya jika soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena diluar kemampuan dari siswa.⁷⁶ Bilangan yang menunjukkan sukar dan mudah suatu butir soal disebut indeks kesukaran (*difficulty index*), indeks kesukaran antara 0,00 sampai 1,0. Indeks kesukaran ini menunjukkan taraf kesukaran soal. Soal dengan indeks kesukaran 0,0 menunjukkan bahwa soal itu terlalu sukar, sebaliknya jika

⁷³ Ibid., hlm. 722.

⁷⁴ Ibid., hlm. 723.

⁷⁵ Zainal Arifin, *Evaluasi Pembelajaran...*, hlm. 45.

⁷⁶ Ibid., hlm. 46.

indeks 1,0 menunjukkan bahwa soal terlalu mudah. Dalam evaluasi indeks kesukaran diberi simbol p .⁷⁷

Model parameter ini dinyatakan dengan persamaan.⁷⁸

$$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1+e^{D(\theta-b_i)}} \dots\dots\dots$$

Keterangan:

$P_i(\theta)$ = Probabilitas peserta tes yang berkemampuan θ menjawab butir ke- i dengan benar.

θ = Parameter kemampuan yang terletak diantara -4 sampai +4.

b_i = Parameter tingkat kesulitan butir ke - i .

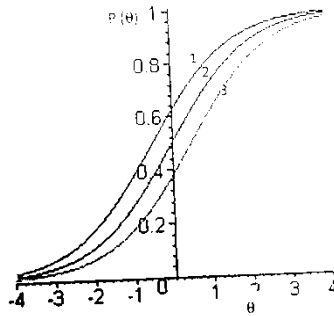
D = Parameter bernilai 1,7

Hubungan peluang menjawab benar $P_i(\theta)$ dengan tingkat kemampuan peserta (θ) dapat digambarkan sebagai kurva karakteristik butir (*item characteristic curve, ICC*). Gambar 1 dibawah ini merupakan ilustrasi kurva karakteristik butir untuk model 1 PL, dengan tingkat kesulitan butir 1 nilai $b = -10$, butir 2 nilai $b = 0$, dan butir 3 nilai $b = 1,0$.



⁷⁷ Suharsimi Arikunto, *Dasar-Dasar Evaluasi...*, hlm. 207.

⁷⁸ Edi Istiyono, *Penilaian Dan Analisis...*, hlm. 289.



Gambar 1 Kurva Karakteristik Butir Model 1PL

$$b_i = -1, b_2 = 0, \text{ dan } b_3 = 1$$

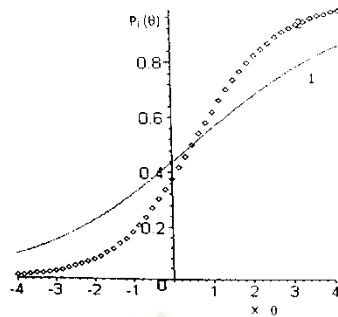
Gambar 1 menunjukkan bahwa perbedaan kurva diatas hanya terdapat pada posisi skala kemampuan saja (θ) artinya tingkat kesulitan tiap butir berbeda, butir ke -1,2,3 secara berturut-turut semakin sulit. Parameter b_i merupakan suatu titik pada skala kemampuan agar peluang menjawab benar sebesar 0,5, misal pada kasus di atas butir 3 memiliki parameter $b_3 = 1,0$, artinya diperlukan kemampuan minimal 1,0 pada skala untuk dapat menjawab benar dengan peluang 0,5. Butir dikatakan baik jika tingkat kesulitan antara -2 dan 2 ($-2 \leq b \leq 2$).⁷⁹

Model logistik 2 parameter butir yaitu indeks kesukaran butir (b) dan indeks daya beda butir (a), yang memenuhi model 2 logistik. Sebagai ilustrasi, urva karakteristik butir 1 ($a=0,5; b=0,5$) dan butir 2 ($a=1; b=0,5$) disajikan pada gambar berikut.⁸⁰

$$P_i(\theta) = \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}}, \text{ dengan } i : 1, 2, 3, \dots, n. \dots \dots \dots (2.9)$$

⁷⁹ Ibid., hlm. 290.

⁸⁰ Heri retnawati, *Teori Respon Butir...*, hlm. 16.

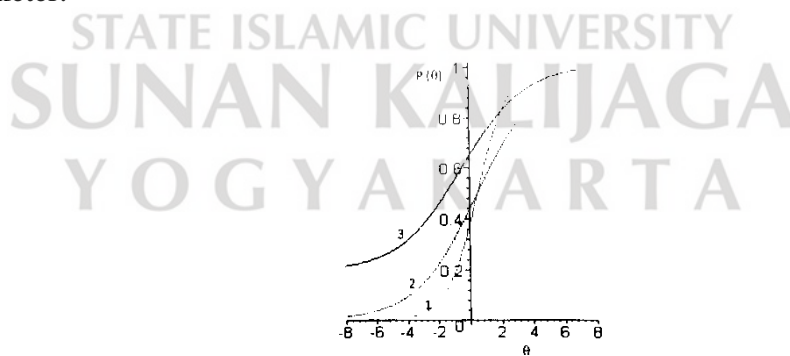


Gambar 2. Kurva karakteristik butir model 2P butir 1 (a=0,5 ; b=0,5) dan butir 2 (a=1 ; b=0,5).

Pada gambar tersebut, asimtot kiri (untuk $0 \rightarrow \sim$) adalah 0, jika bukan 0 maka nilai ini merupakan parameter tebakan semu atau c_i , sehingga model logistik menjadi model 3 parameter (3P), dengan adanya tebakan semu pada model logistik 3 parameter dapat dinyatakan sebagai berikut:⁸¹

$$P_i(\theta) = C_i + (1 - C_i) \frac{e^{ai(\theta - b_i)}}{1 + e^{ai(\theta - b_i)}} \dots\dots\dots (2.10)$$

Sebagai ilustrasi, gambar berikut merupakan kurva karakteristik butir 3 parameter:



Gambar 3. Kurva karakteristik butir model 3P

⁸¹ Ibid., hlm. 16.

Nilai kemampuan peserta (θ) terletak diantara -4 dan $+4$, sesuai dengan daerah asal distribusi normal. Pernyataan ini merupakan asumsi yang mendasari besar nilai b_i . Suatu butir dikatakan baik jika nilai berkisaran antara -2 dan $+2$.⁸² Jika nilai b_i mendekati -2 , maka indeks kesukaran butir rendah, sedangkan jika nilai b_i mendekati $+2$ maka indeks kesukaran butir sangat tinggi ntuk suatu kelompok peserta tes.

Parameter a_i merupakan daya pembeda yang dimiliki butir ke- i . Pada kurva karakteristik a_i merupakan kemiringan dari kurva titik b_i pada skala kemampuan tertentu, karena merupakan kemiringan, diperoleh semakin besar kemiringannya. Maka semakin besar daya pembeda butir tersebut. Pada butir yang baik nilai ini mempunyai hubungan positif dengan performen pada butir dengan kemampuan yang diukur dan a_i terletak 0 dan 2 .⁸³

Peluang menjawab benar dengan memberikan jawaban tebakan semu dilambangkan dengan c_i yang disebut dengan tebakan semu. Parameter ini memberikan suatu kemungkinan asimtot bahwa yang tidak nol pada kurva karakteristik butir (ICC). Parameter ini menggambarkan probabilitas peserta dengan kemampuan rendah menjawab dengan benar pada suatu butir yang mempunyai indeks kesukaran yang tidak sesuai dengan kemampuan peserta tersebut. Besarnya harga c_i diasusimkan lebih kecil dari pada nilai yang akan

⁸² Ibid., hlm. 17.

⁸³ Ibid., hlm. 19.

dihasilkan jika peserta tes menebak secara acak jawaban pada suatu butir soal. Butir soal tes nilai c_i ini berkisar antara 0 dan 1, suatu butir dikatakan baik jika nilai c_i tidak lebih dari 1.

7. Uji Asumsi Item Response Theory

IRT mengharuskan terpenuhinya tiga asumsi: unidimensionalitas, kemandirian lokal, dan invarian parameter. Hanya satu konstruk laten yang diukur oleh serangkaian item dalam tes, menurut asumsi unidimensional. Kemandirian lokal, di sisi lain, mensyaratkan bahwa reaksi setiap subjek terhadap suatu item secara statistik tidak terkait dengan respons lainnya. Selain premis bahwa parameter invarians adalah item independen dalam karakteristik sampel yang bervariasi dari suatu populasi.⁸⁴ Secara lebih rinci ketiga asumsi IRT dijelaskan dibawah ini:

a. Unidimensi

Unidimensi artinya substansi yang diukur adalah satu dimensi.⁸⁵ Unidimensi, artinya setiap butir tes hanya mengukur satu kemampuan. Asumsi ini merujuk pada abilitas yang diukur dalam suatu perangkat soal adalah tunggal. Idealnya, setiap butir tes yang dibuat hanya mengukur salah satu dari kemampuan peserta tes, bukan mengukur dua atau lebih kemampuan peserta tes. Contohnya pada tes prestasi belajar bidang studi matematika, butir-butir yang termuat di dalamnya hanya mengukur kemampuan siswa dalam bidang studi matematika

⁸⁴ Sudaryono, *Implementasi Teori Respon...*, hlm. 722.

⁸⁵ Mardapi, *Pengukuran Penilaian Dan Evaluasi Pendidikan*, hlm. 99.

saja, bukan bidang yang lainnya. Pada praktiknya, asumsi unidimensi tidak dapat dipenuhi secara ketat karena adanya faktor-faktor kognitif, kepribadian dan faktor-faktor pelaksanaan tes, seperti kecemasan, motivasi, dan tendensi untuk menebak. Oleh karena itu, asumsi unidimensi dapat ditunjukkan hanya jika tes mengandung satu saja komponen dominan yang mengukur prestasi subjek.⁸⁶

Jika suatu butir mengukur hal yang bersifat multidimensi, maka skor pada butir tersebut merupakan kombinasi dari berbagai kemampuan subjek. Meskipun begitu pada praktiknya asumsi unidimensi ini tidak dapat secara ketat diterapkan karena adanya faktor-faktor yang mempengaruhi seperti kognitif, kepribadian, dan faktor yang berkaitan dengan aspek administrasi tes. Hal yang paling penting dalam asumsi ini adalah adanya satu komponen dominan yang mempengaruhi performansi subjek. Terdapat tiga macam cara yang dapat digunakan untuk menguji asumsi unidimensi menurut De Mars sebagaimana dikutip oleh ratnawati, yaitu analisis nilai eigen dari matriks korelasi interbutir, uji-Stout pada uji asumsi unidimensi, dan indeks berdasarkan residual pada penyelesaian unidimensi.⁸⁷

b. Independensi Lokal

Jika faktor-faktor yang mempengaruhi prestasi konstan, maka respans subjek terhadap pasangan butir yang manapun akan independen secara statistik satu sarna lain. Kondisi ini disebut dengan independensi lokal.⁸⁸ Independensi lokal terhadap butir dapat diartikan bahwa respon subjek pada butir satu tidak

⁸⁶ Heri retnawati, *Teori Respon Butir...*, hlm. 1.

⁸⁷ *Ibid.*, hlm. 2.

⁸⁸ *Ibid.*, hlm. 3.

berpengaruh terhadap respon pada butir lain atau secara sederhana dapat dikatakan bahwa asumsi independensi lokal akan terpenuhi apabila jawaban peserta terhadap suatu butir soal tidak bergantung pada jawaban peserta terhadap butir soal yang lain.

Tes untuk memenuhi asumsi independensi lokal dapat dilakukan dengan membuktikan bahwa peluang dari pola jawaban setiap peserta tes sama dengan hasil kali peluang jawaban peserta tes pada setiap butir soal. Dengan kata lain, korelasi antara pasangan butir terjadi hanya jika kemampuan utama yang diukur dengan sekumpulan butir tidak dipengaruhi oleh suatu kemampuan yang tidak dimodelkan atau kemampuan yang mempengaruhi oleh kedua kumpulan butir tersebut. De Mars membuktikan asumsi ini adalah dengan membuktikan asumsi unidimensional.⁸⁹ Secara matematis dapat dinyatakan sebagai berikut.

Keterangan:

I : 1, 2, 3, ... n

N : banyaknya butir tes

$P(u_i|\theta)$: Probabilitas responden tes yang memiliki kemampuan dapat menjawab butir ke- i dengan benar

$P(u_1, u_2, \dots, u_n|\theta)$: probabilitas responden tes yang memiliki kemampuan dapat menjawab butir ke- i sampai ke- n dengan benar

⁸⁹ Ibid., hlm. 3.

c. Invariansi Paramater

Invariansi parameter artinya karakteristik butir soal tidak tergantung pada distribusi parameter kemampuan peserta tes dan parameter yang menjadi ciri peserta tes tidak bergantung dari ciri butir soal. Kemampuan seseorang tidak akan berubah hanya karena mengerjakan tes yang berbeda tingkat kesulitannya dan parameter butir tes tidak akan berubah hanya karena diujikan pada kelompok peserta tes yang berbeda tingkat kemampuannya.

Menurut Hambleton, Swaminathan, & Rogers sebagaimana dikutip oleh Retnawati menyatakan bahwa invariansi parameter kemampuan dapat diselidiki dengan mengajukan dua perangkat tes atau lebih yang memiliki tingkat kesukaran yang berbeda pada sekelompok peserta tes. Invariansi parameter kemampuan akan terbukti jika hasil estimasi kemampuan peserta tes tidak berbeda, walaupun tes yang dikerjakan berbeda tingkat kesulitannya. Invariansi parameter butir dapat diselidiki dengan mengujikan tes pada kelompok peserta yang berbeda. Invariansi parameter butir terbukti jika hasil estimasi parameter butir tidak berbeda walaupun diujikan pada kelompok peserta yang berbeda tingkat kemampuannya.⁹⁰

8. Estimasi Kemampuan dengan Metode *Maximum Likelihood*

Kemampuan peserta tes merupakan parameter yang menjadi dasar perhitungan dalam teori respons item. Parameter kemampuan peserta tes dilambangkan dengan θ (*theta*). Menurut Hambleton dalam, nilai dari θ tidak

⁹⁰ Ibid., hlm. 3.

memiliki batas, sehingga dapat berkisar dari $-\infty$ sampai ∞ . Akan tetapi, nilai dapat ditentukan dalam kisaran standar, yaitu -3 sampai 3.⁹¹ Nilai negatif pada estimasi kemampuan menunjukkan bahwa kemampuan siswa dalam menjawab soal tes kurang baik. Nilai positif menunjukkan bahwa kemampuan siswa dalam menjawab butir soal adalah baik. Estimasi kemampuan biasanya digunakan untuk mengetahui kemampuan siswa dalam menjawab pertanyaan.⁹²

Metode *maximum likelihood* adalah metode yang digunakan untuk mencari titik tertentu dalam memaksimalkan fungsi kemungkinan (probabilitas). Selain itu, metode *maximum likelihood* juga diartikan sebagai metode yang digunakan untuk mencari kemampuan yang sesuai untuk setiap pola respon peserta didik dalam menjawab butir soal.⁹³ Distribusi ini digunakan pada data waktu hidup dengan kegagalan konstan. Selain itu, metode *maximum likelihood* juga diartikan sebagai metode yang digunakan untuk mencari kemampuan yang sesuai untuk setiap pola respons peserta didik dalam menjawab butir soal.⁹⁴ Metode *maximum likelihood* merupakan metode yang diyakini efisien untuk semua model respon butir dan untuk tes yang panjang maupun yang pendek, metode ini mengansumsikan adanya respon yang berbeda dari kemampuan 0 yang sama, asumsi ini memungkinkan untuk menghitung probabilitas dari sebagian pola skor butir.⁹⁵ Langkah untuk mengestimasi kemampuan peserta didik menggunakan

⁹¹ Istiyono, *Pengembangan Instrumen...*, hlm. 57.

⁹² Mardapi, *Pengukuran Penilaian...*, hlm. 4.

⁹³ Mardapi, *Teknik Penyusunan...*, hlm. 21.

⁹⁴ Mardapi, *Pengukuran Penilaian...*, hlm. 26.

⁹⁵ Heri retnawati, *Teori Respon Butir...*, hlm. 23.

metode *maximum likelihood* adalah dengan memasukkan parameter daya beda, tingkat kesulitan butir, dan tebakan semu (*pseudo guessing*) dalam persamaan model IRT (1PL, 2PL, 3PL).

Secara umum, persamaan model IRT sebagai berikut:

$$P_i(\theta) = c_1 + (1 - c_1) \frac{e^{D a_i(\theta - b_i)}}{1 + e^{D a_i(\theta - b_i)}}$$

Keterangan:

$P_i(\theta)$ = Probabilitas menjawab benar butir ke-i peserta didik

a_i = daya beda butir ke-i

b_i = tingkat kesulitan butir ke-i

c_1 = tebakan semu (*pseudo guessing*) butir ke-i

θ = kemampuan (*ability*) testee

$D = 1,702$

Berdasarkan persamaan di atas, maka untuk 1 PL, c bernilai 0 dan a bernilai sama untuk seluruh item atau sama dengan 1, sedangkan untuk 2 PL berarti c bernilai 0. Semua unsur berupa daya beda, tingkat kesulitan butir, tebakan semu (*pseudo guessing*) dimasukkan hingga terbentuk nilai-nilai probabilitas. Secara matematis, estimasi kemampuan dengan menggunakan metode *maximum likelihood* berikut:

Peluang peserta didik dengan kemampuan (θ) memberi respons U_i pada item ke-I:

$$U_i = \begin{cases} 1 & \text{untuk jawaban benar} \\ 0 & \text{untuk jawaban salah} \end{cases}$$

Sementara itu, persamaan matematis untuk estimasi kemampuan (θ) sebagai fungsi *maximum likelihood* adalah:⁹⁶

$$L(U_1, U_2, \dots, U_n|\theta) = \prod_{i=1}^n P_i^{U_i} (1 - P_i)^{1-U_i}$$

$$L(U_1, U_2, \dots, U_n|\theta) = \prod_{i=1}^n P_i^{U_i} Q_i^{(1-U_i)}$$

$$L(U_i|\theta) = \prod P_i Q_i$$

Keterangan:

L = Fungsi *maximum likelihood*

U_i = Pola respon peserta didik dalam menjawab butir soal

θ = Tingkat kemampuan peserta didik dalam menjawab butir soal

P_i = Probabilitas menjawab benar butir ke- i

Q_i = Probabilitas menjawab salah butir ke- i

Misalnya ada lima butir soal dengan respon sebagai berikut:

$$U = (U_1, U_2, U_3, U_4, U_5) = (1 \ 0 \ 1 \ 1 \ 0)$$

Jika respon pada lima butir seperti dinyatakan dengan fungsi *likelihood* maka dapat ditulis:

$$L(u|\theta) = P_1 Q_2 P_3 Q_4 P_5$$

Prosedur estimasi kemampuan peserta tes dengan menggunakan pendekatan MLE, apabila skor dan pola respon serta parameter butir diketahui adalah sebagai berikut:

⁹⁶ Ibid., hlm. 3.

- a. Cari besarnya peluang menjawab benar untuk tiap butir dan tiap level kemampuan.
- b. Hitung skor sebenarnya (*true score*) untuk tiap level kemampuan, mulai dari -4 sampai 4.
- c. Hitung besarnya MLE untuk tiap respon peserta didik.
- d. Nilai terbesar untuk probabilitas respon merupakan estimasi kemampuan peserta didik.

9. Kecocokan Model

Model ogive normal merupakan model bentuk kurve karakteristik item yang sangat dominan pada masa awal penelitian mengenai Teori Respon Butir, namun pada saat ini mulai banyak digantikan oleh model-model logistik. Ada tiga macam model logistik. Perbedaan mendasar antara berbagai model yang populer adalah terletak pada jumlah parameter yang digunakan dalam mendeskripsikan butir. Oleh karena itu penyusunan model dalam perkembangannya merujuk pada jumlah parameter yang digunakan, yaitu parameter indeks kesukaran, daya diskriminasi butir, dan probabilitas tebakan semu.⁹⁷

Ada 3 model yang dapat digunakan untuk melakukan analisis dengan menggunakan teori respon butir, yakni model 1PL, 2PL dan 3PL. Model 1PL ini disebut juga sebagai model Rasch. Pada penyusunan model 1PL ini, karakteristik butir yang dilibatkan hanya satu, yaitu parameter taraf kesukaran

⁹⁷ Saifuddin Azwar, "Dasar-Dasar Psikometri", (Yogyakarta: Pustaka Pelajar, 1999), hlm. 66.

butir (bi). Parameter (bi) digambarkan dalam suatu kurva ICC sebagai titik dalam skala kemampuan yang memiliki probabilitas menjawab benar sebesar 0,5. Semakin tinggi taraf kesukaran butir, maka dibutuhkan abilitas yang lebih besar untuk mempertahankan probabilitas 0,5 dalam kurva. Oleh karena itu pada kurva ICC butir-butir yang sulit selalu berada di sebelah kanan butir yang mudah.⁹⁸ Model logistik dua parameter (2PL) menggunakan dua parameter dalam penyusunan model, yaitu parameter taraf kesukaran butir (bi) dan daya beda butir (ai). Model logistik tiga parameter menggunakan semua parameter butir yang ada, yaitu taraf kesukaran butir (bi) daya beda butir (ai), dan probabilitas tebakan semu (ci).

Terdapat dua cara yang dapat digunakan untuk menentukan kecocokan model analisis yang akan digunakan, yakni dengan kecocokan model secara statistik dan dengan plot kurva karakteristik butir.⁹⁹ Pemilihan model dilakukan secara statistik, dari ketiga model dibuat kecocokan butir berdasarkan nilai khi-kuadratnya. Kecocokan model ini dapat diketahui, dengan membandingkan khi-kuadrat hasil perhitungan dengan khi kuadrat tabel dengan derajat kebebasan tertentu.

Kualitas butir secara empiris ditelaah berdasarkan kecocokan data dengan model nilai parameter butir. Kecocokan suatu butir dengan model dapat dilihat dari khi kuadrat butir dibandingkan dengan harga kritik distribusi khi kuadrat

⁹⁸ Ibid., hlm.45.

⁹⁹ Heri retnawati, *Teori Respon Butir...*, hlm 24.

sesuai dengan butir yang bersangkutan pada tingkat signifikansi α . Suatu butir dikatakan tidak cocok modelnya jika nilai butir lebih besar dari harga distribusi pada nilai kritisnya, sebaliknya butir cocok dengan model jika nilai item lebih kecil atau sama dengan nilai distribusi atau dikatakan cocok model jika probabilitas lebih besar dari 0,01.¹⁰⁰

F. Metode Penelitian

1. Desain Penelitian

Penelitian yang digunakan adalah pendekatan penelitian kuantitatif. Menurut Creswell penelitian kuantitatif merupakan salah satu jenis penelitian Pendidikan dimana peneliti memutuskan apa yang akan diteliti, menyusun pertanyaan spesifik, membatasi pertanyaan, mengumpulkan data terukur dari partisipan, menganalisis, angka-angka dengan menggunakan statistik, melakukan penyelidikan yang tidak memihak dengan cara-cara yang obyektif.¹⁰¹ Penelitian kuantitatif memerlukan studi pada sampel dari populasi dan sangat bergantung pada data numerik dan analisis statistik.¹⁰²

Penelitian ini bertujuan mengetahui kualitas butir soal pada soal akhir tahun pelajaran SKI kelas X di MAN 2 Bantul Yogyakarta. Analisis dalam penelitian ini dilakukan secara kuantitatif. Penganalisisan butir soal dilakukan

¹⁰⁰ Yance Manoppo, "Karakteristik Kemampuan Siswa Melalui Ujian Sekolah Mata Pelajaran Kimia Pada SMA Di Kecamatan Teluk Ambon Baguala Menggunakan *Classical Test Theory* (CTT) Dan *Item Respons Theory* (IRT) Model Rasch," dalam *Molluca Journal of Chemistry Education (MJoCE)*, vol. 11, no. 1 2021, hlm. 27.

¹⁰¹ Suharsimi Arikunto, *Dasar-Dasar Evaluasi...*, hlm. 207.

¹⁰² Zainal Arifin, *Evaluasi Pembelajaran...*, hlm. 45.

secara kualitatif dengan menelaah soal berdasarkan kaidah penulisan soal ditinjau dari materi, konstruksi, dan bahasa.

2. Variabel Penelitian

Variabel utama dalam penelitian ini adalah kualitas butir soal, meliputi validitas dan indeks kendala tes. Sedangkan sub variable adalah aspek kognitif yang terukur dalam soal akhir tahun pelajaran SKI kelas X di MAN 2 Bantul Yogyakarta.

Sub variabel teori respon butir adalah:

- a. Indeks kesukaran butir, indeks yang menyatakan tingkat kesulitan untuk menjawab suatu butir yang tampak pada skala kemampuan agar peluang menjawab benar sebesar 0.5 pada item characteristic curve.
- b. Kecocokan dengan model adalah menunjukkan butir tes berperilaku secara konsisten dengan apa yang diharapkan model.
- c. Estimasi kemampuan responden memberikan informasi kemampuan dari responden.

3. Objek Penelitian

Objek penelitian adalah hal yang menjadi sasaran penelitian.¹⁰³ Menurut Supranto objek penelitian adalah elemen yang dapat berupa orang, organisasi atau data yang akan diteliti.¹⁰⁴ Dalam melakukan penelitian yang pertama kali harus diperhatikan adalah objek penelitian yang akan di teliti, dimana objek

¹⁰³ Depdikbud, *Kamus Besar Bahasa Indonesia, Buku Satu*, (Jakarta: Balai Pustaka Utama, 1989), hlm. 622.

¹⁰⁴ Josep Supranto, *Statistik Teori Dan Aplikasi Edisi Keenam*, (Jakarta: Erlangga, 2000), hlm. 21.

penelitian tersebut terkandung masalah yang akan dijadikan bahan penelitian untuk dicari pemecahannya.

Menurut Husein Umar objek penelitian menjelaskan tentang apa atau siapa yang menjadi objek penelitian, juga dimana dan kapan penelitian dilakukan. Bisa juga ditambahkan hal hal lain yang di anggap perlu.¹⁰⁵ Objek penelitian adalah butir soal ujian akhir tahun pelajaran SKI kelas X tahun ajaran 2020/2021 di MAN 2 Bantul Yogyakarta dan seluruh respon jawaban siswa peserta ujian akhir tahun pelajaran SKI kelas X tahun ajaran 2020/2021 di MAN 2 Bantul Yogyakarta.

4. Instrumen Penelitian

Instrumen adalah alat pada waktu peneliti menggunakan suatu metode. Menurut Suharsimi Arikunto “instrument pengumpulan data adalah alat bantu yang dipilih dan digunakan oleh peneliti dalam kegiatan mengumpulkan data agar kegiatan tersebut menjadi sistematis dan mudah bagi peneliti”.¹⁰⁶ Instrumen diperlukan agar pekerjaan yang dilakukan lebih mudah dan hasilnya lebih baik, dalam artian cermat, lengkap dan sistematis sehingga data lebih mudah diolah. Instrumen atau alat yang digunakan dalam penelitian ini berupa lembar jawab siswa ujian akhir tahun pelajaran SKI kelas X tahun ajaran 2020/2021 di MAN 2 Bantul Yogyakarta. Instrumen Penelitian yang digunakan dalam penelitian ini adalah:

¹⁰⁵ Husein Umar, *Metode Penelitian Untuk Skripsi Dan Tesis*, (Jakarta: Rajawali, 2013), hlm. 18.

¹⁰⁶ Arikunto Suharsimi, “*Manajemen Penelitian*,” (Jakarta: Rineka Cipta, 2005), hlm. 101.

- a. Dokumen lembar jawab siswa ujian akhir tahun pelajaran SKI kelas X tahun ajaran 2020/2021 di MAN 2 Bantul Yogyakarta.

5. Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan dengan instrument daftar cek dan dokumentasi:

a. Dokumentasi

Dalam KBBI Dokumentasi didefinisikan sebagai pengumpulan, pemilihan, pengolahan, dan penyimpanan informasi di bidang pengetahuan, pemberian atau pengumpulan bukti dan keterangan seperti gambar, kutipan, kliping, dan bahan referensi lainnya. Sementara menurut Sugiyono Dokumentasi merupakan catatan peristiwa yang sudah berlalu. Dokumen bisa berbentuk tulisan, gambar, atau karya-karya monumental dari seorang. Dokumen yang berbentuk tulisan misalnya catatan harian, sejarah kehidupan (*life histories*), ceritera, biografi, peraturan, kebijakan. Dokumen yang berbentuk gambar misalnya foto, gambar hidup, sketsa dan lain-lain.¹⁰⁷

Dalam penelitian ini Dokumentasi yang digunakan untuk memperoleh data respon peserta didik terhadap butir tes yang berbentuk lembar jawab siswa ujian akhir tahun pelajaran SKI kelas X tahun ajaran 2020/2021 di MAN 2 Bantul Yogyakarta. Data diperoleh langsung dari fotokopi lembar jawab pelajaran SKI tersebut sehingga benar-benar objektif dan akurat, serta keaslian data terjamin. Selain itu dokumentasi digunakan untuk memperoleh

¹⁰⁷ Sugiyono, *Metode Penelitian Pendidikan...*, hlm. 14.

data profil MAN 2 Bantu Yogyakarta. Dokumen tambahan berupa naskah soal dan kunci jawaban soal ujian akhir tahun.

b. Angket (*Kuesioner*)

Kuesioner merupakan teknik pengumpulan data yang dilakukan dengan cara memberi seperangkat pertanyaan atau pernyataan tertulis kepada responden untuk dijawab. Kuesioner digunakan untuk menyelidiki pendapat atau subjek mengenai suatu hal atau untuk mengungkapkan kepada responden. Angket atau kuesioner adalah sejumlah pertanyaan atau pernyataan yang digunakan untuk memperoleh informasi sampel dalam arti laporan pribadinya, atau hal-hal lain yang ingin diketahui.¹⁰⁸ Menurut Sugiyono kuesioner merupakan teknik pengumpulan data yang dilakukan dengan cara memberi seperangkat pertanyaan atau pernyataan tertulis kepada responden untuk menjawabnya.¹⁰⁹ Menurut Hadi Sabari Yunus angket tidak lain juga merupakan alat pengumpul yang berupa daftar pertanyaan, namun diisi sendiri oleh responden.¹¹⁰

Menurut Sugiyono tipe pertanyaan dalam angket dapat terbuka atau tertutup. Pertanyaan terbuka adalah pertanyaan yang diharapkan responden untuk menuliskan jawabannya berbentuk uraian tentang suatu hal, sedangkan pertanyaan tertutup akan membantu responden untuk menjawab dengan cepat

¹⁰⁸ Suharsimi Arikunto, *Prosedur Penelitian: Suatu Pendekatan Praktek*, (Jakarta: Rineka cipta, 2002), hlm. 128.

¹⁰⁹ Sugiyono, *Metode Penelitian Pendidikan...*, hlm. 142.

¹¹⁰ Hadi Sabari Yunus, *Metodologi Penelitian Wilayah Kontemporer*, (Yogyakarta: Pustaka Pelajar, 2010), hlm. 372.

dan juga memudahkan peneliti dalam melakukan analisis data terhadap seluruh angket tertutup jadi responden tinggal memilih jawaban yang sesuai dengan keinginannya.¹¹¹

Menurut Suharsimi Arikunto keuntungan menggunakan angket adalah:¹¹²

- 1) Tidak memerlukan kehadiran peneliti.
- 2) Dapat dibagikan serentak kepada responden.
- 3) Dapat dijawab oleh responden sesuai dengan waktu yang diinginkan.
- 4) Dapat dibuat anonym sehingga responden bebas jujur.
- 5) Pertanyaan yang dibuat sama dengan responden lainnya.

Sedang kelemahan penggunaan angket adalah:

- 1) Responden dalam menjawab sering tidak teliti sehingga ada yang terlewat.
- 2) Seringkali sukar dicarivalidasinya.
- 3) Terkadang responden menjawab tidak jujur.
- 4) Sering tidak Kembali jika dikirimkan melalui pos.
- 5) Waktu pengembalian tidak bersamaan.

Dalam penyusunan instrument peneliti harus memperhatikan Langkah-langkah seperti mendefinisikan konstrak, menyidik faktor dan menyusun butir pertanyaan.¹¹³

¹¹¹ Sugiyono, *Metode Penelitian Kuantitatif Kualitatif Dan R&D* (Bandung: Alfabeta, 2009), hlm. 143.

¹¹² Arikunto, *Prosedur Penelitian...*, hlm. 129.

¹¹³ Sutrisno Hadi, *Statistik Dalam Basica*, (jakarta: Penerbit Andi, 1991), hlm. 11.

Pengumpulan data yang digunakan dalam penelitian ini dikembangkan berdasarkan variabel dan permasalahan yang diteliti. Hal ini dimaksudkan agar semua aspek yang diteliti dapat diamati secara menyeluruh dan tujuan penelitian yang ditetapkan data dicapai. Oleh karena itu pengumpulan data dikembangkan melalui angket atau kuesioner yang berbentuk skala likert dengan menggunakan empat alternatif jawaban untuk setiap pertanyaan. Pembobotan setiap alternatif jawaban menggunakan skala 1-4. Penyusunan angket/kuesioner dilakukan dengan pedoman kuesioner diberikan kepada para ahli untuk menentukan kualitas dari soal SKI kelas X tahun ajaran 2020/2021 di MAN 2 Bantul Yogyakarta.¹¹⁴

6. Teknik Analisis Data

a. Validitas Isi/Konten

Teknik Analisis data dengan uji validitas adalah pengujian dengan tujuan menguji ketepatan dalam menggunakan suatu alat ukur, hal ini dilakukan untuk melihat apakah alat ukur yang digunakan dapat mengukur dengan cermat atau tidak.¹¹⁵ Formula instrument validitas menggunakan formula Gregory hal ini dikarenakan validator yang dilibatkan hanya 2 orang yaitu pakar dibidang pendidikan. Berikut formula validitas isi dilakukan tabulasi silang yang terdiri dari empat kolom yang ditunjukkan tabel berikut.

| |
|-------------|
| Validator 1 |
|-------------|

¹¹⁴ Ibid., hlm. 19.

¹¹⁵ Sugiyono, *Metode Penelitian Pendidikan...* hlm. 125.

| | | | |
|----------------|------------------------------|------------------------------|------------------------------|
| | | Kurang relevan (skor 1-2) | Sangat relevan (Skor 3-4) |
| Validator 2 | Kurang relevan (skor 1-2) | A | B |
| | Sangat relevan (Skor 3-4) | C | D |

Tabel 7. Tabulasi Silang Hasil Penilaian Kedua Validator

Rumus untuk menghitung validitas isi instrument validasi isi adalah:

Validitas Isi: $D / (A+B+C+D)$

Keterangan:

A = sel yang menunjukkan ketidak setujuan antara kedua penilaian

B dan C = sel yang menunjukkan perbedaan pandangan antara penilaian

D= sel yang menunjukkan persetujuan yang valid antara kedua penilaian.

| Nilai | Interpretasi |
|------------|---------------|
| 0,00-0,19 | Sangat Rendah |
| 0,20 -0,39 | Rendah |
| 0,40- 0,59 | Sedang |
| 0,60 -0,79 | Tinggi |
| 0,80-1 | Sangat Tinggi |

Tabel 8. Kriteria Validitas

b. Validitas Empiris

Uji validitas empiris product moment dengan SPSS, penelitian yang dilakukan dengan menggunakan metode ini perlu diuji validitas. Uji validitas berguna untuk mengetahui kevalidan atau kesesuaian angket yang digunakan oleh peneliti untuk memperoleh data dari para responden atau sampel penelitian. Uji validitas procut moment pearson correlation menggunakan prinsip mengkorelasikan atau

menghubungkan antara masing-masing skor item atau soal dengan skor total yang diperoleh dari jawaban responden.¹¹⁶

Dasar pengambilan keputusan dilakukan melalui beberapa cara yaitu: Membandingkan nilai r hitung dengan r tabel.¹¹⁷

- Jika nilai r hitung $>$ r tabel, maka item soal angket tersebut dinyatakan valid
- Jika nilai r hitung $<$ r tabel, maka item soal angket tersebut dinyatakan tidak valid.

Membandingkan nilai Sig. (2-tailed) dengan Probabilitas 0,05

- Jika nilai Sig. (2-tailed) $<$ 0,05 dan Pearson Correlation bernilai positif, maka item soal angket tersebut valid
- Jika nilai Sig. (2-tailed), 0,05 dan Pearson Correlation bernilai negative, maka item soal angket tersebut tidak valid
- Jika nilai Sig. (2-tailed) $>$ 0,05, maka item soal angket tersebut tidak valid.

c. Realibilitas

Reliabilitas merupakan metode yang mempelajari, mengidentifikasi, dan mengestimasi kejelasan atau ketidakejelasan skor tes.¹¹⁸ Menurut Zainul Arifin reliabilitas merupakan tingkat atau derajat konsistensi dari suatu instrumen. Reliabilitas tes berkenaan dengan pertanyaan apakah suatu tes teliti dan dapat dipercaya sesuai dengan kriteria yang ditetapkan. Suatu tes dapat dikatakan reliabel jika selalu memberikan hasil yang sama bila diteskan pada kelompok

¹¹⁶ Zulkifli Matondang, "Validitas Dan Reliabilitas...", hlm. 91.

¹¹⁷ Ibid., hlm. 92.

¹¹⁸ M. Arif Hidayatulloh, "Analisis Kualitas Butir...", hlm. 61.

yang sama pada waktu atau kesempatan yang berbeda.¹¹⁹ Sehingga suatu instrumen evaluasi, dikatakan mempunyai nilai reliabilitas tinggi, apabila tes yang dibuat mempunyai hasil yang konsisten dalam mengukur yang hendak diukur. Reliabilitas memberikan konsistensi yang membuat terpenuhinya syarat utama, yaitu validnya suatu hasil skor instrumen. Semakin reliabel sebuah tes maka semakin dikatakan bahwa hasil tes tersebut akan mempunyai hasil yang sama dan ketika dipakai kembali.

Terdapat lima cara untuk mengetahui reliabilitas suatu tes yaitu:

- 1) Metode dua tes/ metode paralel. Metode ini menggunakan dua tes yang paralel dan setaraf diberikan kepada sekelompok siswa yang sama, kemudian kedua hasilnya dikorelasikan.
- 2) Metode satu tes/ tes ulang (*test-retest metode*) metode ini adalah sebuah tes yang memberikan dua kali tes kepada kelompok siswa yang sama tapi dalam waktu yang berbeda, kemudian hasilnya dikorelasikannya.
- 3) Metode *split-half*, suatu tes dibagi menjadi dua bagian yang sama tingkat kesukarannya. Sama isi dan bentuknya. Kemudian dilihat skor masing-masing bagian paruhan tes tersebut dan di cari korelasinya. Termasuk *split-half method* dengan cara lain yang tidak memerlukan perhitungan korelasi, yaitu menggunakan deviasi standar masing-masing dari kedua bagian tes dan deviasi standar seluruh tes.

¹¹⁹ Zainal Arifin, *Evaluasi Pembelajaran...*, hlm. 258.

Metode Kuder-Richardson digunakan jika suatu instrumen mengukur satu gejala psikologis atau perilaku yang sama, artinya tes tersebut dapat dikatakan reliabel bila terbukti konsistensi jawaban antara soal yang satu dengan soal yang lain.¹²⁰ Suatu instrument evaluasi dikatakan valid apabila instrument yang digunakan dapat mengukur apa yang hendaknya diukur. Validitas suatu instrument evaluasi mempunyai beberapa makna penting seperti:

- 1) Validitas berhubungan dengan ketetapan interpretasi tes atau instrument evaluasi.
- 2) Validitas diartikan sebagai derajat yang menunjukkan kategori yang mencakup rendah, menengah, dan tinggi.
- 3) Prinsip suatu tes valid, tidak universal. Validitas suatu tes hanya valid untuk suatu tujuan tertentu saja.¹²¹

Wahidmurni, dkk menyatakan bahwa validitas merupakan seberapa besar derajat diperolehnya bukti yang empiric guna mendukung kebenaran dan kesesuaian pertanyaan tersebut, pengukuran disebut valid apabila betul-betul mengukur apa yang hendak diukur. Dengan demikian validitas merupakan aspek paling penting dalam menentukan kualitas hasil pengukuran.¹²²

Tinggi rendahnya reliabilitas secara empiris ditunjukkan oleh suatu angka yang disebut koefisien reliabilitas yang berkisar antara 0.00 – 1,00. Metode Kuder- Richardson 20 (K-R 20) merupakan koefisien reliabilitas yang

¹²⁰ M Ngalim Purwanto, *Prinsip-Prinsip...*, hlm. 139.

¹²¹ Sukardi, *Evaluasi Pendidikan Prinsip...*, hlm. 31.

¹²² Wahidmurni, *Evaluasi Pembelajaran...*, hlm. 88.

dapat menggambarkan butir-butir untuk jawaban benar/salah yang diberi skor 0-1. Koefisien reliabilitas Kuder-Richardson 20 dapat dihitung dengan menggunakan rumus berikut.

$$r_{11} = \frac{K}{K-1} \left(\frac{S^2 - \sum pq}{S^2} \right)$$

r_{11} = Reliabilitas tes

p = Proporsi peserta tes menjawab benar

q = Proporsi peserta tes menjawab salah ($q=1-p$)

$\sum pq$ = Jumlah perkalian antara p dan q

k = Banyaknya butir

s = Standar deviasi atau simpangan baku

Terdapat beberapa langkah yang harus ditempuh dalam perhitungan reliabilitas menggunakan Kuder-Richardson 20 yaitu:

g) Membuat tabel persiapan perhitungan reabilitas KR-20

h) Menentukan proporsi peserta tes menjawab benar (p) untuk setiap butir

i) Menentukan proporsi peserta tes menjawab salah (q) dengan rumus $q = 1-p$

j) Menentukan varian total

k) Memasukkan semua bilangan yang telah diperoleh pada langkah-langkah di atas kedalam persamaan KR-20

l) Menginterpretasikan hasil koefisien reliabilitas yang diperoleh

Dalam hal ini nilai (r_i) diartikan sebagai nilai reliabilitas, sehingga oleh Guilford kriteriannya disajikan dalam tabel berikut:

| Nilai | Interpretasi |
|---------------------------|---------------|
| $r_i < 0,20$ | Sangat Rendah |
| $0,20 \leq r_i < 0,40$ | Rendah |
| $0,40 \leq r_i < 0,60$ | Sedang |
| $0,60 \leq r_i < 0,80$ | Tinggi |
| $0,80 \leq r_i \leq 1,00$ | Sangat Tinggi |

Tabel 9. Kriteria Reliabilitas

d. Analisis Butir Soal dengan Teori Respon Butir (*Item Response Theory*)

Estimasi butir dan responden dilakukan dengan prosedur *PROX (normal approximation estimation)*. Kecocokan antara kemampuan responden (θ) dan indeks kesukaran (b) akan menghasilkan akurasi dalam pengukuran. Akurasi maksimal terjadi saat $P(\theta) = 0,5$. Estimasi parameter dilakukan dengan membuang responden yang benar dan salah semua. Estimasi parameter responden dan butir dilakukan serentak karena keduanya belum diketahui. Estimasi terus dilakukan sampai nilai parameter responden dan butir konstan.¹²³

a) Kualitas Butir

Ditentukan dari kecocokan butir dengan model Rasch dan indeks kesukaran butir. Butir yang baik harus memenuhi syarat dari teori respon butir. Butir dianalisis kecocokannya dengan nilai *infit meansquare*.

| Nilai <i>infit Meansquare</i> | Keterangan |
|-------------------------------|-------------|
| >1,33 | Tidak cocok |

¹²³ Dadan Rosana, "Analisis Butir Dan Identifikasi Ketidakwajaran Skor Ujian Akhir Sekolah Untuk Standarisasi Penilaian," dalam *Jurnal Kependidikan: Penelitian Inovasi Pembelajaran*, vol. 45, no. 2, 2015, hlm. 136.

| | |
|-------------|-------------|
| 0,77 – 1,33 | Cocok |
| <0,77 | Tidak cocok |

Tabel 10. Kriteria kecocokan butir dengan model rasch

Tahap selanjutnya melihat nilai outfit t butir dengan kriteria sebagai berikut:

| Kriteria | Keterangan |
|----------------------|------------|
| Outfir $t \leq 2,00$ | Soal Lolos |
| Outfir $t > 2,00$ | Soal Gugur |

Tabel 11. Kriteria Lolos tidaknya suatu butir soal

Tahap ketiga adalah menganalisis indeks kesukaran butir dengan melihat nilai delta. Indeks kesukaran butir menurut teori respon butir adalah sebagai berikut:

| Nilai <i>threshold</i> | Keterangan |
|------------------------|--------------|
| $b > 2$ | Sangat sukar |
| $1 < b \leq 2$ | Sukar |
| $-1 \leq b \leq 1$ | Sedang |
| $-1 > b \geq -2$ | Mudah |
| $b < -2$ | Sangat mudah |

Tabel 12. Kriteria Indeks Kesukaran Butir

Berdasarkan hasil analisis tiap butir tes, maka kualitas perangkat tes dapat ditentukan berdasarkan presentasi jumlah butir tes yang tidak baik.

Kriteria kualitas tes ditentukan: ¹²⁴

| No | Presentase jumlah butir tes yang tidak baik | Kriteria |
|----|---------------------------------------------|-------------|
| 1 | 0% s.d 10% | Sangat baik |
| 2 | 11% s.d 20% | Baik |
| 3 | 21% s.d 30% | Cukup baik |
| 4 | 31% s.d 40% | Kurang baik |
| 5 | >40% | Tidak baik |

¹²⁴ Ibid., hlm. 85.

Tabel 13. Kriteria kualitas butir menurut pendekatan teori respon butir

Baik tidaknya perangkat tes secara kuantitatif dapat dilihat dari besaran statistic butir maupun responden yang memiliki rata-rata, ralat (*error*), realibilitas, *infit mean squares*, *outfit mean squares* baik jika nilainya mendekati 1 sedangkan *infit t* dan *outfit t* dikatakan baik jika nilainya mendekati 0.

b) Estimasi Kemampuan Responden

Kemampuan responden dilihat dari banyaknya butir yang dapat dijawab dengan benar. Semakin banyak butir yang dijawab benar maka kemampuan responden semakin tinggi. Estimasi kemampuan responden dilihat pada file teta pada nilai estimate. Kriteriaanya adalah:

| Nilai <i>Estimate</i> | Keterangan |
|-----------------------|------------------|
| >+1,00 | Kemampuan tinggi |
| -1,00 s.d +1,00 | Kemampuan sedang |
| <-1,00 | Kemampuan rendah |

Tabel 14. Kriteria kemampuan responden

c) Peta butir responden

Peta distribusi tingkat kesulitan butir dan kemamouan responden dapat dilihat pada file map. File ini menyajikan persebaran responden menurut tingkat kesukarannya dalam skala logit -4,0 sampai +4,0. Dari peta ini dapat dilihat tingkat kesukaran butir jika dibandingkan dengan kemampuan responden.

d) Kurva Butir

Kurva karakteristik butir dibuat dengan menggunakan nilai delta. Delta merupakan tingkat kesukaran butir atau sering disebut dengan parameter b. nilai delta disubstitusikan ke dalam persamaan.

$$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1+e^{D(\theta-b_i)}} \dots\dots\dots$$

Keterangan:

$P_i(\theta)$ = Probabilitas peserta tes yang berkemampuan θ menjawab butir ke-i dengan benar.

θ = Parameter kemampuan yang terletak diantara -4 sampai +4.

b_i = Parameter tingkat kesulitan butir ke – i.

D = Parameter bernilai 1,7

Hubungan peluang menjawab benar $P_i(\theta)$ dengan tingkat kemampuan peserta (θ) dapat digambarkan sebagai kurva karakteristik butir (*item characteristic curve, ICC*).¹²⁵

G. Sistemastika Pembahasan

Secara garis besar, bahasan Tesis ini terbagi menjadi tiga bagian, yaitu bagian awal, bagian inti dan bagian akhir. Bagian awal berisikan halaman judul, halaman surat pernyataan, halaman persetujuan pembimbing, halamn pengesahan, halaman motto, halaman persembahan, kata pengantar, abstrak, daftar isi, daftar tabel, daftar

¹²⁵ Edi Istiyono, *Penilaian Dan Analisis...*, hlm. 289.

gambar dan daftar lampiran. Kemudian bagian tengah berisikan uraian penelitian mulai dari pendahuluan sampai penutup yang tertuang dalam empat bab. Setiap bab terdapat sub-sub bab yang menjelaskan pokok pembahasan sebagaimana berikut:

Bab Pertama diawali dengan pendahuluan, mencakup: latar belakang masalah, rumusan masalah, tujuan dan kegunaan penelitian, kajian penelitian terdahulu, kerangka teori, metode penelitian dan sistematika pembahasan.

Bab Kedua, berisi gambaran umum MAN 2 Bantul Yogyakarta yang terdiri dari sejarah singkat, lokasi madrasah, visi, misi dan tujuan madrasah, struktur organisasi, keadaan guru pengajar dan keadaan siswa.

Bab Tiga, berisi tentang hasil validasi isi, empiris, reabilitas dan analisis kualitas butir soal akhir tahun pelajaran SKI kelas X di MAN 2 Bantul Yogyakarta tahun 2020/2021. Uji kecocokan model 1PL, 2PL dan 3PL.

Bab Empat, berisi penutup yang terdiri dari kesimpulan dan saran.

Pada bagian akhir tesis terdapat daftar pustaka, lampiran-lampiran, dan daftar riwayat hidup, angket, soal ujian dan kisi-kisi ujian akhir tahun.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

BAB IV

PENUTUP

A. Kesimpulan

Dari hasil analisis yang telah dilakukan terhadap 25 soal Mata pelajaran Sejarah Kebudayaan Islam MAN 2 Bantul dapat ditarik kesimpulan sebagai berikut:

1. Hasil analisis validitas isi menunjukkan bahwa kriteria validitas isi dikategorikan sangat tinggi dengan hasil 0,60. Hasil analisis validitas konten jika dilihat dari kompetensi dasar dan indikator soal sudah terwakili didalam soal, sehingga sudah dikatakan valid secara validitas konten. Tetapi jika dilihat dari validitas kontrak soal masih terdiri dari C1 dan C2.
2. Hasil analisis validitas empiris menunjukkan bahwa soal SKI ujian akhir tahun dengan jumlah siswa 157 dan 25 butir soal terdapat 24 butir soal yang valid dan 1 butir soal tidak valid pada no soal 4.
3. Hasil estimasi reliabilitas menggunakan teori Kuder-Richardson (KR20) diperoleh hasil sebesar 0,82748, maka dapat disimpulkan tingkat reliabilitas soal SKI sangat tinggi.
4. Hasil analisis parameter 1PL, diperoleh tingkat kesukaran dari 25 butir soal sudah cukup berimbang. Hal ini dapat dipahami dengan proporsi 7 butir soal dikategorikan sangat mudah, 5 butir soal mudah, 4 butir soal sukar dan 9 butir sedang.

5. Hasil analisis parameter butir dengan model 2PL, parameter tingkat kesukaran 2 butir masuk dalam kategori sangat mudah, 11 butir dengan kategori mudah, 11 butir dengan kategori sedang, dan 1 butir masuk dalam kategori sangat sukar. Dalam model 2PL ini tidak ada butir yang masuk dalam kategori sukar. 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.
6. Hasil analisis parameter butir dengan model 3PL menunjukkan bahwa tingkat kesukaran butir soal 1 butir sangat mudah, 9 butir mudah, 13 butir sedang, 2 butir sukar. Daya beda butir soal 24 butir baik, 1 cukup baik, tebak semu dari butir soal 8 butir tidak baik dan 17 butir baik. Hasil Analisis kecocokan model 1PL, 2PL, 3PL model yang menghasilkan butir yang cocok terdapat pada model 1PL.
7. Hasil estimasi kemampuan siswa dengan estimasi MLE, siswa dengan kemampuan tinggi dicontohkan siswa nomor 61, 72, dan 79 dengan kemampuan masing-masing 2.986200325, 2.174195569 dan 2.986200325. Sedangkan siswa dengan kemampuan sedang dicontohkan siswa nomor 1, 2 dan 3 dengan kemampuan masing-masing -0.727558224, 0.881563263, dan 1.646185231. Sedangkan siswa yang memiliki kemampuan rendah dicontohkan siswa nomor 123, 125, dan 127 dengan kemampuan masing-masing -2.721888673, -3.375643114, -2.173257804.

B. Rekomendasi

1. Untuk guru.

Guru dapat melakukan analisis sendiri untuk meningkatkan kualitas soal ujian. Dengan memperbaiki maupun menghapus dan mengganti soal yang memiliki karakteristik parameter yang kurang baik. Guru dapat melakukan pembuatan soal dengan tingkatan *Higher Order Thinking Skills* (HOTS).

2. Untuk mahasiswa

Ke depan dapat dilanjutkan melakukan uji tes soal menggunakan 3PL maupun 4PL. Langkah yang dapat dilakukan ialah dengan memperbanyak jumlah subjek agar model analisis butir dapat fit.

DAFTAR PUSTAKA

- Ahmad, Iqbal Faza, Nur Hidayanto Pancoro Setyo Putro, Zulkifli Syauqi Thontowi, Ahmad Syafii, and M Aldi Subakti. "Trends in the Implementation of Higher-Order Thinking Skills in Islamic Religious Education in Madrasahs and Schools: A Systematic Literature Review." *Jurnal Pendidikan Islam* 9, no. 2 (2020): 195–216.
- Ahmad, Iqbal Faza, and Sukiman Sukiman. "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, no. 2 (2019): 137–64.
- Amru, Achmad. "Penerapan Metode Kuder-Richardson (KR 20) Dan Naive Bayes Classifier Dalam Analisis Butir Soal Hasil Ujian Tengah Semester." Universitas Islam Negeri Maulana Malik Ibrahim, 2014.
- Ansori, Ari Hasan. "PENGARUH METODE PENYEKORAN DAN KECEMASAN TES TERHADAP INDEKS KETIDAKWAJARAN SEKOR TES MATEMATIKA SETELAH MENGONTROL PENGETAHUAN AWAL." *JMIE (Journal of Madrasah Ibtidaiyah Education)* 1, no. 1 (2017).
- Arikunto, Suharsimi. "Prosedur Penelitian Suatu Pendekatan Praktik," 2013.
- "Arsip Lembaga Pendidikan Madrasah Aliyah Negeri 2 Bantul Yogyakarta," n.d.
- Azwar, Saifuddin. "Dasar-Dasar Psikometri." *Yogyakarta: Pustaka Pelajar*, 1999.
- Badrun Kartowagiran. *Pengantar Teori Tes Klasik*. Yogyakarta: Pascasarjana UNY & Dinas Pendidikan Propinsi Yogyakarta, 2009.

- Basuki dan Hariyanto. *Asemen Pembelajaran*. Bandung: PT Remaja Rosdakarya, 2016.
- Candiasa, I Made. “Statistik Univariat Dan Bivariat Disertai Aplikasi SPSS.” *Singaraja: Unit Penerbitan Universitas Pendidikan Ganesha*, 2010.
- Creswell, John W. *Educational Research: Planning, Conducting, and Evaluating Quantitative*. Vol. 7. Prentice Hall Upper Saddle River, NJ, 2002.
- Depdikbud. *Kamus Besar Bahasa Indonesia, Buku Satu*. Jakarta: Balai Pustaka Utama, 1989.
- Djaali dkk. *Pengukuran Dalam Bidang Pendidikan*. Jakarta: Program Pascasarjana, 2000.
- Edi Istiyono. *Penilaian Dan Analisis Hasil Belajar Fisika Dengan Teori Tes Klasik Dan Modern*. Yogyakarta: UNY Press, 2020.
- Fatonah, Siti. “Evaluasi Pelaksanaan Asesmen Otentik Kurikulum 2013 Di MI Yogyakarta.” *Al-Bidayah: Jurnal Pendidikan Dasar Islam* 8, no. 2 (2016).
- Gall, Meredith D. *Applying Educational Research: How To Read, Do*, 2010.
- Hadi, Sutrisno. “Metodologi Research II Yogyakarta: Andi Offset, 1991,” 1992.
- Hambleton, R K, H Swaminathan, and H J Rogers. “Principles and Applications of Item Response Theory.” Boston, MA: Kluwer-Nijhoff Publishing Company, 1985.
- Hasnah, Hasnah. “Analisis Kualitas Soal Matematika Ujian Sekolah Kelas XII IPA SMA Negeri Di Watansoppeng Berdasarkan Teori Respon Butir.” *PEP Educational Assessment* 1, no. 1 (2017): 27–34.
- Heri retnawati. *Teori Respon Butir Dan Penerapannya Untuk Peneliti, Praktis Pengukuran Dan Penguji*. Edited by Nuha Medika. Yogyakarta, 2014.

- Hullin, Ch L, F Drasgow, and Ch K Parsons. "Item Response Theory: Application to Psychological Measurement." *Homewood, IL: Dow Jones-Irwin*, 1983.
- Ihsan, Helli. "Validitas Isi Alat Uukur Penelitian: Konsep Dan Panduan Penilaiannya." *Pedagogia* 13, no. 3 (2015): 173–79.
- Istiyono. *Pengembangan Instrumen Penilaian Dan Analisis Hasil Belajar Fisika Dengan Teori Tes Klasik Dan Modern*. Yogyakarta: UNY Press, 2020.
- Karkal, Yeshwanth Rao, and Ganesh Shenoy Kundapur. "Item Analysis of Multiple Choice Questions of Undergraduate Pharmacology Examinations in an International Medical School in India." *Journal of Dr. NTR University of Health Sciences* 5, no. 3 (2016): 183.
- Kemendikbud. *Panduan Penilaian Oleh Pendidik Dan Satuan Pendidik Untuk Sekolah Menengah Atas*. Jakarta: Direktorat Jenderal Pendidikan Dasar dan Menengah, 2017.
- Kereh, Cicylia T, Paulus C Tjiang, and Jozua Sabandar. "Validitas Dan Reliabilitas Instrumen Tes Matematika Dasar Yang Berkaitan Dengan Pendahuluan Fisika Inti." *Jurnal Pendidikan* 2, no. 01 (2015).
- Kheyami, Deena, Ahmed Jaradat, Tareq Al-Shibani, and Fuad A Ali. "Item Analysis of Multiple Choice Questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain." *Sultan Qaboos University Medical Journal* 18, no. 1 (2018): e68.
- Kurniawan, Riza Yonisa, Albrian Fiky Prakoso, Luqman Hakim, Retno Mustika Dewi, and Irin Widayanti. "Pemberian Pelatihan Analisis Butir Soal Bagi Guru Di Kabupaten Jombang: Efektif?" *Jurnal Pemberdayaan Masyarakat Madani*

(*JPMM*) 1, no. 2 (2017): 179–93.

Lord, Frederic M. *Applications of Item Response Theory to Practical Testing Problems*.
Routledge, 2012.

M. Arif Hidayatulloh. “Analisis Kualitas Butir Soal Ulangan Semester Genap Bahasa Arab Kelas VIII Tahun Pelajaran 2012/2013 Dengan Program Quest Di SMP Muhammadiyah 3 Depok Sleman Daerah Istimewa Yogyakarta.” Jurusan Pendidikan Bahasa Arab Fakultas Ilmu Tarbiyah dan Keguruan Universitas Islam Negeri Sunan Kalijaga Yogyakarta, 2013.

Mahanani. “Analisis Soal Internasional Competitions and Assessments for Schools (ICAS) Dengan Menggunakan Metode Item Response Theory (IRT) Dan Classical Test (CTT).” Biologi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang, 2015.

Manoppo, Yance. “Karakteristik Kemampuan Siswa Melalui Ujian Sekolah Mata Pelajaran Kimia Pada SMA Di Kecamatan Teluk Ambon Baguala Menggunakan Classical Test Theory (CTT) Dan Item Respons Theory (IRT) Model Rasch.” *Molluca Journal of Chemistry Education (MJoCE)* 11, no. 1 (2021): 24–31.

Mardapi. *Pengukuran Penilaian Dan Evaluasi Pendidikan*. Edited by Nuha Medika. Yogyakarta, 2012.

Mardapi, Djemari. “Teknik Penyusunan Instrumen Tes Dan Nontes.” Yogyakarta: Mitra Cendikia Press, 2008.

Matondang, Zulkifli. “Validitas Dan Reliabilitas Suatu Instrumen Penelitian.” *Jurnal Tabularasa* 6, no. 1 (2009): 87–97.

N.Sudjana. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja, 2012.

- Nana Sudjana. *Penilaian Hasil Proses Belajar Mengajar*. Edited by Remaja Risdakarya. Bandung, 2004.
- Niron, Maria Dominika. *Pendampingan Manajemen Pelaksanaan Kurikulum Berbasis Kompetensi di Sekolah se Propinsi DIY*. Yogyakarta: Departemen Pendidikan Nasional Universitas Negeri Yogyakarta, 2005.
- Nurchahyo, Firmanto Adi. "Aplikasi IRT Dalam Analisis Aitem Tes Kognitif." *Buletin Psikologi* 24, no. 2 (2017): 64–75.
- Nusantari, Diah Oga. "Meningkatkan Estimasi Reliabilitas Instrumen Pilihan Ganda Dengan Memperhatikan Probabilitas Guessing." *Formatif: Jurnal Ilmiah Pendidikan MIPA* 6, no. 2 (2016).
- Oktarina, Ani, and Siti Fatonah. "Pengamatan Tentang Pembelajaran Dan Penilaian Pada Anak Usia Dini Di Era Pandemi Covid-19." *Cakrawala Dini: Jurnal Pendidikan Anak Usia Dini* 12, no. 1 (2021): 31–40.
- Perdana, Sukma Adi. "Analisis Kualitas Instrumen Pengukuran Pemahaman Konsep Persamaan Kuadrat Melalui Teori Tes Klasik Dan Rasch Model." *Jurnal Kiprah* 6, no. 1 (2018): 41–48.
- Prilianti, Lya Deni, Dedeh Kurniasih, and Fitriani Fitriani. "ANALISIS KEVALIDAN LKS BERBASIS HIERARKI KONSEP PADA MATERI KELARUTAN DAN HASIL KALI KELARUTAN." *Jurnal Pendidikan Informatika Dan Sains* 7, no. 1 (2018): 68–77.
- Purwanto. *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar, 2010.
- Purwanto, M Ngilim. "Prinsip-Prinsip Dan Teknik Evaluasi Pengajaran," 2019.
- Ratnaningsih, Dewi Juliah, and Isfarudi Isfarudi. "Analisis Butir Tes Objektif Ujian

- Akhir Semester Mahasiswa Universitas Terbuka Berdasarkan Teori Tes Modern.”
Jurnal Pendidikan Terbuka Dan Jarak Jauh 14, no. 2 (2013): 98–109.
- Robb, Thomas N, and Jay Ercanbrack. “A Study of the Effect of Direct Test Preparation on the TOEIC Scores of Japanese University Students.” *TESL-EJ* 3, no. 4 (1999): 1–22.
- Rosana, Dadan. “Analisis Butir Dan Identifikasi Ketidakwajaran Skor Ujian Akhir Sekolah Untuk Standarisasi Penilaian.” *Jurnal Kependidikan: Penelitian Inovasi Pembelajaran* 45, no. 2 (2015).
- Rosidah, Nenden Annisa, Taufik Ramlan Ramalis, and Iyon Suyana. “Karakteristik Tes Keterampilan Berpikir Kritis (KBK) Berdasarkan Pendekatan Teori Respon Butir.” *Jurnal Inovasi Dan Pembelajaran Fisika* 5, no. 1 (2018).
- Saifudin Azwar. *Sikap Manusia Terori Dan Pengukurannya*. Edited by Pustaka Pelajar. Yogyakarta, 2003.
- Setiawati, Maulina Nurriska, Purwanto Purwanto, and Taufik Ramlan Ramalis. “Analisis Parameter Tes Penilaian Akhir Semester Fisika Kelas X Dengan Teori Respon Butir.” *WaPFI (Wahana Pendidikan Fisika)* 7, no. 1 (n.d.): 1–9.
- Snipes, Michael, and D Christopher Taylor. “Model Selection and Akaike Information Criteria: An Example from Wine Ratings and Prices.” *Wine Economics and Policy* 3, no. 1 (2014): 3–9.
- Sri Wening. *Materi Evaluasi Pembelajaran*. Yogyakarta: Fakultas Teknik Universitas Negeri Yogyakarta, 2012.
- Subali, Bambang, and Pujiyati Suyata. “Panduan Analisis Data Pengukuran Pendidikan Untuk Memperoleh Bukti Empirik Kesahihan Menggunakan Program Quest.”

- Yogyakarta: Lembaga Penelitian Dan Pengabdian Pada Masyarakat UNY, 2011.
- Sudaryono. *Teori Respon Butir*. Yogyakarta: Graha Ilmu, 2013.
- Sudaryono, Sudaryono. “Implementasi Teori Responsi Butir (Item Response Theory) Pada Penilaian Hasil Belajar Akhir Di Sekolah.” *Jurnal Pendidikan Dan Kebudayaan* 17, no. 6 (2011): 719–32.
- Sudijono, Anas. “Pengantar Evaluasi Pendidikan,” 2001.
- Sugiyono. *Metode Penelitian Kuantitatif Kualitatif Dan R&D*. Bandung: Alfabeta, 2009.
- Sugiyono, Dr. “Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif Dan R&D,” 2013.
- Suharsimi Arikunto. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: PT. Bumi aksara, 2000.
- Suharsimi, Arikunto. “Manajemen Penelitian.” *Jakarta: Rineka Cipta*, 2005.
- Suharsimi Arikunto. *Prosedur Penelitian: Suatu Pendekatan Praktek*. Jakarta: Rineka cipta, 2002.
- Sukardi. *Evaluasi Pendidikan Prinsip & Operasionalnya*. Yogyakarta: PT Bumi Aksara, 2012.
- Sumiati, Ati, Umi Widiastuti, and Usep Suhud. “Workshop Teknik Menganalisis Butir Soal Dalam Meningkatkan Kompetensi Guru Di SMK Cileungsi Bogor.” *Jurnal Pemberdayaan Masyarakat Madani (JPMM)* 2, no. 1 (2018): 136–53.
- Supranto, Josep. “Statistik Teori Dan Aplikasi Edisi Keenam (Jilid I).” *Erlangga. Jakarta. Hlm*, 2000, 1–27.

- Suryabrata, Sumadi. "Pengembangan Alat Ukur Psikologis." Yogyakarta: Andi Offset, 2000.
- Umar, Husein. "Metode Penelitian Untuk Skripsi Dan Tesis." *Jakarta: Rajawali* 42 (2013).
- Wahidmurni, dkk. *Evaluasi Pembelajaran, Kompetensi Dan Praktik*. Yogyakarta: Nuha Litera, 2010.
- Wilson, Leslie Owen. "Anderson and Krathwohl Blooms Taxonomy Revised." *Understanding the New Version of Bloom's Taxonomy*, 2016.
- Yang, Frances M. "Item Response Theory for Measurement Validity." *Shanghai Archives of Psychiatry* 26, no. 3 (2014): 171.
- Yunus, Hadi Sabari. "Metodologi Penelitian Wilayah Kontemporer." *Yogyakarta: Pustaka Pelajar* 251 (2010).
- Zainal Arifin. *Evaluasi Pembelajaran Prinsip Teknik Prosedur*. Bandung: PT Remaja Rosdakarya, 2011.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA