

**KUALITAS SOAL UJIAN SEKOLAH ATAU MADRASAH
(US/M) MATA PELAJARAN MATEMATIKA
TAHUN AJARAN 2016-2017 KOTA SURABAYA
BERDASARKAN TEORI RESPON BUTIR**

SKRIPSI

Oleh:

NURUL HASANAH

NIM. D04213028



**UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
FAKULTAS TARBIYAH DAN KEGURUAN
JURUSAN PENDIDIKAN MATEMATIKA DAN IPA
PRODI PENDIDIKAN MATEMATIKA
JANUARI 2018**

PERSETUJUAN PEMBIMBING SKRIPSI

Skripsi oleh:

Nama : Nurul Hasanah

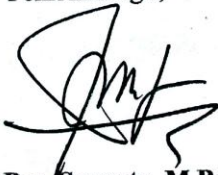
Nim : D04213028

Judul : KUALITAS SOAL UJIAN SEKOLAH ATAU MADRASAH
(US/M) MATA PELAJARAN MATEMATIKA TAHUN
AJARAN 2016-2017 KOTA SURABAYA BERDASARKAN
TEORI RESPON BUTIR

ini telah diperiksa dan disetujui untuk diujikan.

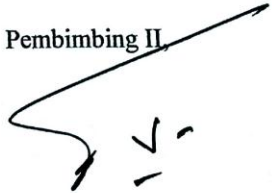
Surabaya, 24 Oktober 2017

Pembimbing I,



Drs. Suparto, M.Pd.I
NIP. 196904021995031002

Pembimbing II,



Dr. Kusaeri, M.pd
NIP. 197306052007012048

PENGESAHAN TIM PENGUJI SKRIPSI
Skripsi oleh Nurul Hasanah ini telah dipertahankan di depan Tim
Penguji Skripsi

Surabaya, 31 Januari 2018

Mengesahkan Fakultas Tarbiyah dan Keguruan
Universitas Islam Negeri Sunan Ampel Surabaya



Tim Penguji
Penguji I,

Maunah Setyawati, M.Si
NIP. 197411042008012008

Penguji II,

Agus Prasetyo Kurniawan, M.Pd
NIP. 198308212011011009

Penguji III,

Drs. Suparto, M.Pd.I
NIP. 196904021995031002

Penguji IV,

Dr. Kusaeri, M.Pd
NIP. 197306052007012048

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Nurul Hasanah
NIM : D04213028
Jurusan/Program Studi : Pendidikan Matematika dan
IPA/Pendidikan Matematika
Fakultas : Fakultas Tarbiyah dan Keguruan

Menyatakan dengan sesungguhnya bahwa skripsi yang saya tulis ini benar-benar tulisan saya, dan bukan merupakan plagiasi baik sebagian atau seluruhnya.

Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa skripsi ini hasil plagiasi, baik sebagian atau seluruhnya, maka saya bersedia menerima sanksi atas perbuatan tersebut sesuai dengan ketentuan yang berlaku.

Surabaya, 31 Januari 2018
Yang membuat pernyataan



Nurul Hasanah
NIM. D04213028



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertandatangan di bawah ini, saya:

Nama : Nurul Hasanah
NIM : D04213028
Fakultas/Jurusan : FTK / PMIPA
E-mail address : nurul.h3028@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi Tesis Desertasi Lain-lain (.....)

yang berjudul :

Kualitas Soal Ujian Sekolah atau Madrasah (US/M) Mata Pelajaran

Matematika Tahun Ajaran 2016-2017 Kota Surabaya Berdasarkan

Teori Respon Butir.

Beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara **fulltext** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 07 Februari 2018

Penulis


(Nurul Hasanah.)
Nama terang dan tandatangan

KUALITAS SOAL UJIAN SEKOLAH ATAU MADRASAH (US/M) MATA PELAJARAN MATEMATIKA TAHUN AJARAN 2016 - 2017 DI KOTA SURABAYA BERDASARKAN TEORI RESPON BUTIR

Oleh:
Nurul Hasanah

ABSTRAK

Kualitas soal menjadi salah satu bagian yang penting dalam evaluasi pendidikan. Kualitas soal yang digunakan untuk menguji siswa sangat mempengaruhi kualitas hasil pengukuran yang pada akhirnya juga mempengaruhi hasil evaluasi pendidikan. Untuk melihat kualitas soal, dapat dilakukan dengan menggunakan analisis kualitas soal, baik secara kualitatif (teoritik) maupun kuantitatif (empiris). Tujuan dari penelitian ini adalah untuk mengetahui kualitas soal secara kuantitatif dengan teori respon butir, indikator yang dianalisis adalah reliabilitas soal, daya beda soal, tingkat kesukaran soal, dan nilai peluang menebak atau *guessing* dari soal.

Jenis penelitian ini menggunakan penelitian deskriptif kuantitatif. Penelitian ini mendeskripsikan kualitas soal berdasarkan teori respon butir dengan menggunakan bantuan program BILOG MG. Sampel yang digunakan sebanyak 200 siswa dengan teknik random sampling. Penelitian ini dilakukan di SMP Negeri 3 Surabaya dan SMP Islam Al-Amal Surabaya. Metode pengumpulan datanya menggunakan metode dokumentasi dan metode tes.

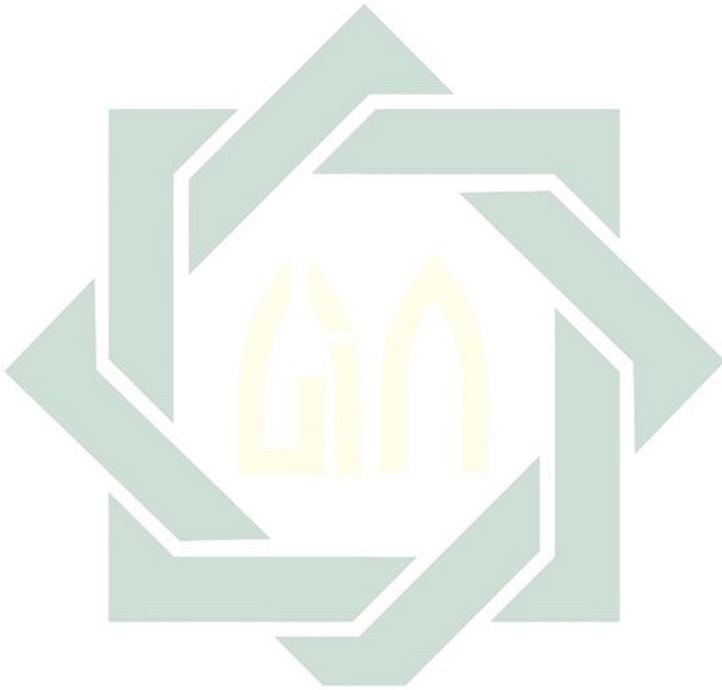
Hasil penelitian sebagai berikut: reliabilitas soal 0,8213 masuk dalam kategori reliabel. Tingkat kesukaran dari 40 butir soal secara keseluruhan adalah -2,089 masuk kedalam kategori kurang baik. Daya beda dari 40 butir soal secara keseluruhan adalah 1,063 masuk kedalam kategori baik. Nilai peluang menebak atau *guessing* dari 40 butir soal secara keseluruhan adalah 0,250 masuk dalam kategori baik.

Kata Kunci: Tes respon butir, tingkat kesukaran, daya beda, peluang menebak atau *guessing*.

DAFTAR ISI

SAMPUL DALAM	i
PERSETUJUAN PEMBIMBING	ii
PENGESAHAN TIM PENGUJI	iii
PERNYATAAN KEASLIAN TULISAN	iv
PERSEMBAHAN	v
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR GRAFIK	xv
DAFTAR LAMPIRAN	xvi
BAB I : PENDAHULUAN	1
A. Latar Belakang	1
B. Rumusan Masalah	5
C. Tujuan Penelitian	6
D. Manfaat Penelitian	6
E. Definisi Operasional	6
BAB II : KAJIAN PUSTAKA	8
A. Kualitas Soal	8
B. Teori Klasik	16
C. Teori Respon Butir	19
D. Analisis Kualitas Soal	31
BAB III : METODE PENELITIAN	34
A. Jenis Penelitian	34
B. Populasi dan Sampel Penelitian	34
C. Prosedur Penelitian	35
D. Metode Pengumpulan Data	35
E. Teknik Pengumpulan Data	36
BAB IV : HASIL PENELITIAN	40
A. Deskripsi Data	40
B. Analisis Data	41
C. Pembahasan	48

BAB VI : PENUTUP	60
A. Simpulan.....	60
B. Saran.....	60
DAFTAR PUSTAKA	61
LAMPIRAN	



BAB I

PENDAHULUAN

A. Latar Belakang

Pemerintah telah mengeluarkan Peraturan Menteri Pendidikan Nasional Nomor 23 Tahun 2016 tentang Standar Penilaian Pendidikan. Dalam peraturan tersebut disebutkan bahwa standar penilaian pendidikan adalah standar nasional pendidikan yang berkaitan dengan mekanisme, prosedur, dan instrumen penilaian hasil belajar peserta didik, dan penilaian pendidikan adalah proses pengumpulan dan pengolahan informasi untuk menentukan pencapaian hasil belajar peserta didik. Penilaian pendidikan melalui tiga cara yakni: penilaian oleh pendidik, penilaian oleh satuan pendidikan (sekolah), dan penilaian oleh pemerintah.¹

Penilaian hasil belajar oleh pemerintah dilakukan dalam bentuk pelaksanaan Ujian Nasional (UN) yang bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu dalam kelompok mata pelajaran ilmu pengetahuan dan teknologi. Hal ini sesuai dalam Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 pasal 57 yang menyatakan bahwa dalam rangka pengendalian mutu pendidikan secara nasional dilakukan evaluasi sebagai bentuk akuntabilitas penyelenggaraan pendidikan kepada pihak-pihak yang berkepentingan. UN digunakan sebagai standarisasi dari pemerintah untuk menguji kelayakan seorang siswa untuk dapat melanjutkan pendidikan ke jenjang yang lebih tinggi dan sebagai pemerataan pendidikan secara nasional atau dengan kata lain merupakan evaluasi akhir siswa.

Penilaian hasil belajar peserta didik pada jenjang Sekolah Dasar (SD) atau Madrasah Ibtidaiyah (MI) mengalami beberapa perubahan dalam setiap tahunnya. Dimulai dari tahun 1965 hingga 1971 disebut dengan Ujian Negara, pada tahun 1972 sampai dengan 1979 disebut dengan Ujian Sekolah, tahun 1980 sampai 2002 disebut dengan Evaluasi Belajar Tahap Nasional (EBTANAS),

¹Kemendikbud, *Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 23 Tahun 2016. Tentang Standar Penilaian Pendidikan Bab II Pasal 2*, (Jakarta : Kemendikbud, 2016)

tahun 2003 sampai 2004 disebut dengan Ujian Akhir Nasional (UAN), pada tahun 2005 hingga tahun 2013 disebut dengan Ujian Nasional, dan pada tahun 2014 hingga saat ini berganti nama menjadi Ujian Sekolah atau Madrasah (US/M).²

Tahun pelajaran 2016/2017 berdasarkan Keputusan Menteri Pendidikan Nasional Republik Indonesia Nomor 58 Tahun 2015 ditetapkan sistem penilaiannya dalam bentuk Ujian Sekolah/Madrasah (US/M)³. Khusus penyelenggaraan Ujian Sekolah/Madrasah (US/M) tahun pelajaran 2016/2017, komposisi butir soal, sebanyak 25% ditetapkan oleh Pemerintah Pusat dan berlaku secara nasional, serta 75% soal ditetapkan oleh penyelenggara US/M tingkat provinsi. Ujian Sekolah/ Madrasah (US/M) jenjang SD dalam pelaksanaannya berpedoman pada Keputusan Peraturan Kepala Badan Penelitian dan Pengembangan Kementerian Pendidikan dan Kebudayaan Nomor 007/H/EP/2017 tentang Prosedur Operasi Standar (POS) Ujian Sekolah/Madrasah (US/M) untuk Sekolah Dasar/Madrasah Ibtidaiyah/Sekolah dasar Luar Biasa (SD/MI/SDLB) Tahun Pelajaran 2016/2017.⁴ Bentuk tes yang digunakan diantaranya berupa tes tertulis (*paper and pencil test*).

Dalam evaluasi pendidikan yang berkaitan dengan penilaian oleh pemerintah, kualitas soal menjadi salah satu bagian yang penting. Kualitas soal yang digunakan untuk mengetes siswa sangat mempengaruhi kualitas hasil pengukuran yang pada akhirnya juga mempengaruhi hasil evaluasi pendidikan⁵. Dengan kata lain, untuk mendapatkan hasil evaluasi pembelajaran yang baik, tes yang digunakan harus terjamin kualitasnya. Untuk melihat kualitas sebuah

² https://id.wikipedia.org/wiki/Ujian_Nasional, diakses pada tanggal 9 Agustus 2017

³ Kemendikbud, *Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 102 Tahun 2013 Tentang Penyelenggaraan Ujian Sekolah/Madrasah Pada Sekolah Dasar/Madrasah Ibtidaiyah, Sekolah Dasar Luar Biasa, Dan Program Paket A/Ula*, (Jakarta : Kemendikbud, 2017)

⁴ Mendikbud, *Peraturan Kepala Badan Penelitian Dan Pengembangan Kementerian Pendidikan Dan Kebudayaan Nomor: 007 /H/Ep/2017 Tentang Prosedur Operasional Standar Ujian Sekolah/Madrasah Pada Sekolah Dasar/Madrasah Ibtidaiyah, Sekolah Dasar Luar Biasa, Dan Penyelenggara Program Paket A/Ula Tahun Pelajaran 2016/2017*,(Jakarta : Mendikbud, 2017)

⁵ Muh. Nurung, *Kualitas Tes Ujian Akhir Sekolah Berstandar Nasional (UASBN) Ipa Sd Tahun Pelajaran 2007/2008 Di Kota Kendari*,(Thesis. Yogyakarta: Graduate School, State University of Yogyakarta, 2008), hal 3

tes dapat dilakukan dengan menggunakan analisis kualitas soal, baik secara kualitatif (teoritik) maupun kuantitatif (empiris). Secara kualitatif tes dikatakan baik jika telah memenuhi persyaratan penyusunan dari sisi materi, konstruksi dan bahasa. Adapun secara kuantitatif dapat dilakukan dengan dua teknik yaitu teori tes klasik dan teori tes responsi butir.⁶

Analisis kualitas soal tes merupakan suatu tahap yang harus ditempuh untuk mengetahui derajat kualitas suatu tes. Dalam penilaian hasil belajar, tes diharapkan dapat menggambarkan perilaku peserta didik dan menghasilkan nilai yang objektif serta akurat. Jika tes yang digunakan guru kurang baik, maka hasil yang diperoleh tentunya kurang baik. Hal ini dapat merugikan peserta didik itu sendiri. Tes hendaknya disusun sesuai dengan prinsip dan prosedur penyusunan tes yang standar. Setelah digunakan perlu diketahui apakah tes tersebut berkualitas baik atau kurang baik sehingga harus dilakukan analisis kualitas tes.

Analisis secara empiris dapat menggunakan pendekatan tes klasik maupun pendekatan tes modern (teori respon butir). Pada teori tes klasik, besarnya skor tampak (X) peserta ujian merupakan penjumlahan dari skor murni (T) dan skor kesalahan pengukuran (E). Secara matematis, dapat dituliskan dalam bentuk formula $X = T + E$. Teori tes klasik memiliki beberapa asumsi seperti yaitu⁷: (1) tidak ada korelasi antara skor yang sebenarnya dan skor kesalahan, (2) rerata kesalahan acak pengukuran sama dengan nol, dan (3) skor kesalahan pada test paralel tidak berkorelasi. Dengan dasar asumsi tersebut, maka dikembangkan sejumlah formula untuk mengestimasi indeks keandalan, indeks kesahihan tes, dan indeks kesukaran serta daya beda dari suatu tes. Selanjutnya, sejumlah ahli mengembangkan program komputer untuk mengestimasi parameter butir berdasarkan teori tes klasik.

Sayangnya, teori tes klasik memiliki keterbatasan yang mendasar antara lain: (1) Hasil estimasi parameter tergantung pada karakteristik peserta ujian (*group dependent*). Hal ini berimplikasi

⁶ Syamsir Syainudin dan Muhammad Ilyas, *Karakteristik Butir Tes Matematika pada Tes Buatan MGMP Matematika Kota Palopo Berdasarkan Teori Klasik*, Jurnal Pedagogy, Volume:1 Nomor 1, (Palopo : Universitas Cokroaminoto, 2014), hal 126

⁷Djemari Mardapi, *Teknik penyusunan instrument tes dan nontes*, (Yogyakarta: Mitra Cendekia Press, 2008), hal 32

pada tingkat kesukaran soal akan menjadi rendah jika tes diujikan pada kelompok peserta tes berkemampuan tinggi, dan sebaliknya jika tes diujikan pada peserta dengan kemampuan rendah, maka tingkat kesukaran tes itu akan tinggi. (2) Hasil estimasi kemampuan peserta tergantung pada karakteristik butir soal (*item dependent*). Jika peserta dengan kemampuan tinggi mengerjakan butir yang mudah, maka peserta tersebut mempunyai peluang menjawab benar yang sangat tinggi. Sebaliknya, jika peserta dengan kemampuan rendah menjawab butir yang sulit, maka peluang dia menjawab dengan benar akan sangat kecil.⁸

Penelitian terdahulu telah banyak dilakukan dalam menguji suatu kualitas soal ujian. Diantaranya penelitian oleh; (1) Muh. Nurung dimana dalam menganalisis soal peneliti menggunakan teori klasik dengan poin- poin yang dianalisis diantaranya adalah tingkat kesukaran, daya beda, dan efektifitas pengecohnya dengan menggunakan program iteman.⁹(2) Nur Khoiriyah yang dalam penelitiannya juga menganalisis soal UASBN mata pelajaran matematika tahun 2009 di Kota Jombang dengan menggunakan teori klasik dimana yang dianalisis adalah validitas, reliabilitas, tingkat kesukaran, daya beda dan efektifitas pengecohnya dengan bantuan program iteman¹⁰, dan (3) Syamsir Sainuddin yang menganalisis karakteristik butir tes matematika pada tes buatan MGMP matematika di Kota Palopo berdasarkan teori klasik yang mengukur validitas, reliabilitas, tingkat kesukaran, daya beda dan efektifitas pengecohnya dengan menggunakan microsoft excel dan SPSS.¹¹ Dari ketiga penelitian di atas para peneliti menguji kualitas suatu soal tertentu dengan menggunakan teori yang sama yaitu teori tes klasik dan masih menggunakan program-program manual.

Mengatasi kelemahan pada teori tes klasik, maka para ahli pengukuran mengembangkan model yang tidak terikat dengan sampel (*sample free*). Model ini selanjutnya dikenal tes modern atau teori respon butir. Menurut teori respon butir, perilaku seseorang

⁸ Dewi Julia Ratnaningsih, *Analisis Butir Tes Objektif Ujian Akhir Semester Mahasiswa Universitas Terbuka Berdasarkan Teori Tes Modern*, Jurnal Pendidikan Terbuka dan Jarak Jauh Volume:14 Nomor , (September 2013), hal 99

⁹ Muh. Nurung, Op. Cit., hal 25

¹⁰ Nur Khoiriyah, *Kualitas Tes UASBN Mata Pelajaran Matematika SD/MI Di Kabupaten Jombang Tahun 2009*, (Undergraduate Thesis, UIN Sunan Ampel Surabaya, 2010), hal 14.

¹¹ Syamsir Syainudin dan Muhammad Ilyas, Op. Cit., hal 43

dapat dijelaskan oleh karakteristik orang yang bersangkutan sampai pada batas-batas tertentu. Teori respons butir (IRT) merupakan salah satu cara untuk menilai kelayakan butir dengan membandingkan rerata penampilan butir terhadap tampilan bukti kemampuan kelompok yang diramalkan oleh model. Tujuan utama teori respons butir dikembangkan adalah untuk mengatasi kelemahan teori tes klasik yang tidak *independent* terhadap kelompok peserta yang mengerjakan tes maupun terhadap tes yang diujikan.¹²

Kondisi demikian menggugah peneliti yang kemudian menjadi pertimbangan sebagai bahan yang melatarbelakangi penyusunan skripsi yang berjudul **“Kualitas Soal Ujian Sekolah atau Madrasah (US/M) Mata Pelajaran Matematika Tahun Ajaran 2016 - 2017 di Kota Surabaya Berdasarkan Teori Respon Butir”**.

B. Rumusan Masalah

Berdasarkan pokok-pokok latar belakang yang telah diuraikan sebelumnya, rumusan masalah penelitian ini adalah “Bagaimanakah kualitas soal US/M mata pelajaran matematika tahun ajaran 2016-2017 Kota Surabaya berdasarkan teori respon butir?”.

Adapun kualitas soal dapat dilihat dari beberapa indikator, diantaranya ;

1. Bagaimanakah reliabilitas soal US/M mata pelajaran matematika tahun ajaran 2016-2017 di Kota Surabaya ?
2. Bagaimanakah daya beda soal US/M mata pelajaran matematika tahun ajaran 2016-2017 di Kota Surabaya ?
3. Bagaimanakah tingkat kesukaran soal US/M mata pelajaran matematika tahun ajaran 2016-2017 di Kota Surabaya ?
4. Bagaimanakah peluang menebak atau *guessing* dari soal US/M mata pelajaran matematika tahun ajaran 2016-2017 di Kota Surabaya?

¹² Ibid, hal 100

C. Tujuan Penelitian

Tujuan utama penelitian ini adalah untuk mengetahui kualitas soal ujian sekolah atau madrasah (US/M) mata pelajaran matematika tahun ajaran 2016-2017 Kota Surabaya berdasarkan teori respon butir.

Tujuan tersebut dirinci sebagai berikut:

1. Untuk mengetahui reliabilitas soal US/M mata pelajaran matematika tahun ajaran 2016-2017 di Kota Surabaya.
2. Untuk mengetahui daya beda soal US/M mata pelajaran matematika tahun ajaran 2016-2017 di Kota Surabaya.
3. Untuk mengetahui tingkat kesukaran soal US/M mata pelajaran matematika tahun ajaran 2016-2017 di Kota Surabaya.
4. Untuk mengetahui peluang menebak atau *guessing* dari soal US/M mata pelajaran matematika tahun ajaran 2016-2017 di Kota Surabaya

D. Manfaat Penelitian

Hasil penelitian ini diharapkan memiliki manfaat sebagai berikut:

1. Hasil analisis dalam penelitian ini dapat menjadi ukuran untuk menunjukkan tingkat kesahihan suatu tes US/M. Dengan diketahuinya tingkat kesahihan tes US/M dapat digunakan untuk memotret secara keseluruhan hasil US/M.
2. Hasil analisis dalam penelitian ini dapat digunakan untuk mengetahui fungsi soal.
3. Dapat menjadi masukan bagi satuan terkait sebagai bahan untuk menentukan kebijakan dan langkah-langkah yang dipandang efektif dibidang pendidikan, terutama yang berhubungan dengan evaluasi.
4. Dapat menjadi acuan bagi para peneliti lain untuk melakukan penelitian sejenis atau melanjutkan penelitian tersebut secara intensif dan mendalam.

E. Definisi Operasional

Agar penelitian ini dapat terarah dan tidak terlalu luas jangkauannya maka dibutuhkan pembatasan masalah. Adapun pembatasan masalah dalam penelitian ini adalah:

1. Soal US/M mata pelajaran matematika adalah seperangkat soal ujian akhir nasional tingkat SD di Kota Surabaya mata

pelajaran matematika yang berbentuk objektif pilihan ganda yang terdiri atas 40 item.

2. Kualitas soal secara kuantitatif adalah kategori butir soal US/M setelah dianalisis Adapun karakteristiknya adalah reliabilitas soal, tingkat kesukaran soal, daya beda soal dan peluang menebak atau *guessing* soal.
3. Analisis butir soal dengan teori respon butir model 3 PL merupakan analisis butir soal dengan teori respon butir yang menggunakan fungsi matematika untuk menghubungkan karakteristik butir soal dengan kemampuan responden yang digambarkan melalui kurva karakteristik butir. Dimana probabilitas peserta tes untuk dapat menjawab benar suatu butir soal ditentukan oleh tiga karakteristik butir yaitu tingkat kesukaran butir, daya beda butir, dan peluang menebak atau *guessing* butir.
4. Reliabilitas suatu tes adalah taraf sampai dimana suatu tes mampu menunjukkan konsisten hasil pengukurannya yang diperlihatkan dalam taraf ketetapan dan ketelitian hasil. Reliabel tes berhubungan dengan ketetapan hasil tes.
5. Tingkat kesukaran soal adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks kesukaran antara 0 dan 1.
6. Daya beda soal adalah kemampuan soal membedakan peserta yang berkemampuan tinggi dan berkemampuan rendah setelah menjawab soal yang diujikan.
7. Peluang Menebak atau *guessing* adalah ukuran peluang peserta untuk menebak dengan benar.

BAB II KAJIAN TEORI

A. Kualitas Soal

1. Pengertian

Menurut istilah, kata kualitas berarti mutu, yaitu tingkat baik buruknya sesuatu.¹ Pakar dan organisasi mencoba mendefinisikan kualitas (mutu) berdasarkan sudut pandangnya masing-masing seperti yang terurai di bawah ini²:

- a. Menurut Joseph Juran, kualitas adalah kesesuaian untuk penggunaan (*fitness for use*), ini berarti bahwa suatu produk atau jasa hendaklah sesuai dengan apa yang diperlukan atau diharapkan oleh pengguna.
- b. Menurut Edward Deming, suatu tingkat yang dapat diprediksi dari keseragaman dan keberagamaan pada biaya rendah dan sesuai dengan pasar.
- c. Welch Jr mengatakan bahwa kualitas adalah jaminan kesetiaan pelanggan, pertahanan terbaik melawan saingan dari luar, dan satu-satunya jalan menuju pertumbuhan dan pendapatan yang langgeng.
- d. Menurut ISO 2000, kualitas adalah totalitas kerakteristik suatu produk (barang dan jasa) yang menunjang kemampuannya untuk memuaskan kebutuhan yang dispesifikan atau ditetapkan.
- e. Menurut Soewarso Hardjosudarmo, bahwa yang dimaksud kualitas adalah penilaian subyektif daripada “costumer” penentuan ini ditentukan oleh persepsi “costumer” terhadap produk dan jasa.

Dari beberapa pendapat tokoh di atas, terdapat beberapa kesamaan yaitu dalam elemen-elemen sebagai berikut:

- a. Kualitas meliputi usaha memenuhi atau melebihi harapan pelanggan.
- b. Kualitas menyangkut produk, jasa, manusia, proses dan lingkungan.

¹Departemen Pendidikan Nasional, *Kamus Besar Bahasa Indonesia*, (Jakarta: Balai Pustaka.2002), hal 603

²Nur Khoiriyah, *Kualitas Tes Uasbn Mata Pelajaran Matematika Sd/Mi Di Kabupaten Jombang Tahun 2009*, (Undergraduate Thesis, UIN Sunan Ampel Surabaya.2010), hal 13

- c. Kualitas merupakan kondisi yang selalu berubah (misalnya apa yang dianggap kualitas saat ini, mungkin dianggap kurang berkualitas pada masa mendatang)

Kualitas dalam konteks pendidikan adalah mengacu pada prestasi yang dicapai oleh anak didik atau sekolah pada setiap kurun waktu tertentu. Prestasi yang dicapai atau hasil pendidikan (*student achievement*) dapat berupa hasil tes kemampuan akademis, (misalnya ulangan umum, UAS, EBTA dan UNAS). Dapat pula prestasi dibidang lain, seperti prestasi disuatu cabang olahraga, seni atau ketrampilan tambahan tertentu.³ Mutu atau kualitas jasa pendidikan bersifat relatif (sesuai dengan kebutuhan pelanggan), dan bukan bersifat absolute. Dengan kata lain, mutu pendidikan akan baik dan memuaskan jika sesuai atau melebihi kebutuhan para pelanggan yang bersangkutan. Dalam pendidikan, yang dimaksud dengan pelanggan atau klien (*client*) dibagi menjadi dua, yakni pelanggan internal dan pelanggan eksternal⁴.

- a) Pelanggan internal (*internal customer*) adalah orang-orang yang berada dalam organisasi sekolah, yaitu guru, staf tata usaha, pesuruh (*office boys*) *cleaning service*, pelayan teknis dan komponen lainnya.
- b) Pelanggan eksternal (*external customer*) adalah orang-orang yang berada di luar organisasi sekolah yang memperoleh layanan dari sekolah.

Pelayanan eksternal dibagi menjadi dua macam, yakni:

- 1) Pelanggan primer (*primary customer*) adalah pelanggan utama, yakni orang-orang yang langsung bersentuhan dengan jasa-jasa pendidikan yang diberikan oleh sekolah, seperti peserta didik.
- 2) Pelanggan sekunder (*secondary customer*) adalah pihak-pihak lain yang secara tidak langsung terimbas dari layanan pendidikan yang diberikan oleh sekolah, yaitu orang tua siswa, masyarakat, pemerintah dan dunia usaha dan industri sebagai pengguna tenaga kerja.

³Nanang Hanifah Dan Cucu Suhana, *Konsep Strategi Pembelajaran*, (Bandung: Rafika Aditama, 2009) hal. 83-86

⁴Ibid, hal 81-83

2. Indikator Kualitas

Seperti jelaskan di atas, bahwa para pakar telah mendefinisikan kualitas secara beragam menurut pendapatnya masing-masing, begitu juga dengan indikator kualitas. David A Gavin mengemukakan delapan dimensi atau kategori kritis dari kualitas, yaitu⁵:

- a. *Performance* (kinerja). Karakteristik kinerja utama produk.
- b. *Feature* (profil). Aspek sekunder dari kinerja, atau kinerja tambahan dari suatu produk.
- c. *Reliability* (dapat dipercaya). Kemungkinan produk malfungsi atau tidak berfungsi dengan baik, dengan konteks ini produk atau jasa dapat dipercaya dalam menjalankan fungsinya.
- d. *Conformance* (kesesuaian). Kesesuaian atau cocok dengan keinginan atau kebutuhan konsumen.
- e. *Durability* (daya tahan). Daya tahan produk atau masa hidup produk, baik secara ekonomis maupun teknis.
- f. *Serviceability* (kepelayanan). Kecepatan, kesopanan, kompetensi, mudah diperbaiki.
- g. *Aesthetics* (keindahan). Keindahan produk dalam desain, rasa, suara atau bau dari produk, dan ini bersifat subyektif.
- h. *Perceived quality* (kualitas yang dipersepsi). Kualitas dalam pandangan pelanggan atau konsumen.

3. Prinsip-Prinsip Kualitas

Prinsip kualitas adalah sejumlah asumsi yang dinilai dan diyakini memiliki kekuatan untuk mewujudkan mutu. Akan hal ini, beberapa ahli dan organisasi mencoba merumuskan prinsip-prinsip yang paling tepat untuk mewujudkan kualitas dalam organisasi atau kelembagaan.

Menurut Deming ada empat belas prinsip kualitas yang harus dilakukan jika menghendaki tercapainya suatu kualitas, yaitu⁶:

⁵Nur Khoiriyah, *Kualitas Tes Uasbn Mata Pelajaran Matematika Sd/Mi Di Kabupaten Jombang Tahun 2009*, (Undergraduate Thesis, UIN Sunan Ampel Surabaya, 2010), hal 15

⁶Tim Dosen Administrasi Pendidikan UPI, *Manajemen Pendidikan*, (Bandung: Alfabeta, 2009), hal. 296-302

- a. Menciptakan konsistensi tujuan untuk pengembangan produk dan jasa dengan adanya tujuan suasana bisnis yang kompetitif.
- b. Adopsi filosofi baru.
- c. Menghentikan ketergantungan pada adanya dengan upaya pencapaian kualitas.
- d. Menghentikan anggapan bahwa penghargaan dalam bisnis adalah terletak pada harga.
- e. Peningkatan sistem produksi dan layanan secara terus menerus guna peningkatan kualitas dan produktivitas.
- f. Pelatihan dalam pekerjaan
- g. Kepemimpinan kelembagaan
- h. Menghilangkan rasa takut
- i. Menghilangkan penghalang antar departemen.
- j. Mengurangi slogan peringatan-peringatan dan target, dan mengganti dengan pematapan metode-metode yang dapat meningkatkan kualitas kerja.
- k. Kurangi standar kerja yang menentukan kuota berdasarkan jumlah. Hilangkan penghambat yang dapat menghilangkan hak asasi manusia untuk merasa bangga terhadap kecakapan kerjanya.
- l. Lembagakan suatu program pendidikan dan peningkatan diri yang penuh semangat.
- m. Setiap orang dalam perusahaan bekerja sama dalam mendukung proses transformasi.

4. Karakteristik Kualitas Soal

Tes yang digunakan untuk mengukur kemampuan peserta didik haruslah tes yang berkualitas baik. Tes yang berkualitas baik tentunya mampu menjadi tolak ukur yang baik untuk mengukur kemampuan peserta didik⁷. Untuk mengetahui bagaimanakah kualitas tes, maka dilakukan analisis kualitas tes. Dengan analisis kualitas tes kita dapat mengetahui bagaimana kondisi soal yang digunakan untuk tes. Setidaknya terdapat empat karakteristik yang harus dimiliki oleh tes hasil belajar, sehingga tes tersebut dapat dinyatakan sebagai tes yang baik.

⁷ Ani Purwanti dan dan Irni Wulandari, *Studi Kualitas UASBN Mata Pelajaran Matematika Wilayah Jakarta Timur Tahun Ajaran 2007/2008*, www.lpmpdki.web.id/pdf/ani%20-%20irni.pdf Diakses pada tanggal 2 September 2017.

Keempat karakteristik tersebut adalah validitas, reliabilitas, objektif dan praktis⁸.

Dari keempat karakteristik tersebut, peneliti hanya menggunakan karakteristik kedua yaitu reliabilitas untuk penelitian terkait kualitas soal US/M mata pelajaran matematika tahun ajaran 2016-2017 Kota Surabaya. Dengan diadakannya penelitian tersebut diharapkan kualitas soal pada tes tersebut dapat diketahui.

a. Validitas

Valid sering diartikan sebagai kesahihan. Suatu alat ukur disebut valid bilamana alat ukur tersebut memiliki isi yang layak mengukur objek yang harus diukur dan sesuai dengan kriteria tertentu⁹. Suatu alat ukur dipandang valid apabila sesuai dengan isi kurikulum yang hendak diukur¹⁰. Dalam hal ini ada kesesuaian antara alat ukur dengan fungsi pengukuran dan sarana pengukuran.

Validitas memiliki beberapa karakteristik. Pertama, validitas merujuk kepada ketepatan interpretasi hasil suatu penilaian yang dikenakan terhadap siswa, bukan merujuk pada instrumen penilaian tersebut. Kedua, validitas berkaitan dengan pengategorian derajat tertentu, seperti validitasnya tinggi, sedang, atau rendah. Ketiga, validitas senantiasa berkaitan dengan kondisi khusus. Artinya tidak ada penilaian yang valid untuk semua tujuan.¹¹

Validitas dapat dibagi menjadi 3 macam yaitu validitas isi, validitas konstruksi, validitas ramalan atau validitas prediksi.¹²

1) Validitas isi (*content validity*)

Sebuah tes dikatakan memiliki validitas isi jika mengukur tujuan khusus tertentu yang sejajar dengan

⁸Anas Sudijono, Pengantar Evaluasi Pendidikan, (Jakarta : Raja Grafindo Persada, 2011), hlm. 93.

⁹toswari.staff.gunadarma.ac.id/Downloads/files/8165/Uji+Validitas+dan+Reliabilitas.pdf, diakses pada tanggal 4 September 2017

¹⁰ Sumarna Surapranata, *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes Implementasi Kurikulum 2004*, (Bandung: Remaja Rosdakarya. 2005), hal 51

¹¹ Anas Sudijono, *Pengantar Evaluasi Pendidikan*, (Jakarta : Raja Grafindo Persada, 2011), hal 52

¹² Ibid, hal 54

materi atau isi pelajaran yang diberikan. Oleh karena materi yang diajarkan tertera dalam kurikulum maka validitas isi ini sering disebut validitas kurikuler.

2) Validitas konstruksi (*construct validity*)

Sebuah tes dikatakan memiliki validitas konstruksi jika butir-butir soal tersebut mengukur setiap aspek berpikir seperti yang disebutkan dalam tujuan instruksional khusus.

3) Validitas prediksi (*predictive validity*)

Memprediksi artinya meramal, dan meramal selalu mengenai hal yang akan datang. Sebuah tes dikatakan memiliki validitas prediksi apabila mempunyai kemampuan untuk meramalkan apa yang akan terjadi pada masa yang akan datang.

b. Reliabilitas

Reliabel berarti dapat dipercaya. Reliabilitas (keandalan) adalah ketetapan atau ketelitian suatu alat evaluasi¹³. Suatu tes atau alat evaluasi dikatakan andal jika dapat dipercaya, konsisten atau stabil dan produktif¹⁴. Tes reliabel dapat tidak valid, tetapi tes yang valid biasanya reliabel¹⁵. Diantara hal yang mempengaruhi hasil tes adalah banyaknya soal dalam tes dan kualitas butir soal. Semakin panjang tes biasanya makin reliabel. Cara mencari reliabilitas antara lain¹⁶:

1) Metode paralel (*equivalent*)

Tes paralel atau tes ekuivalen adalah dua buah tes yang mempunyai kesamaan tujuan, tingkat kesukaran, dan susunan, tetapi butir-butir soalnya berbeda. Dalam menggunakan metode tes paralel ini pengetes harus menyiapkan dua buah tes, dan masing-masing dicobakan pada kelompok siswa yang sama.

2) Metode tes ulang (*test-retest method*)

¹³ <http://id.wikipedia.org/wiki/Reliabilitas>, diakses pada tanggal 4 September 2017.

¹⁴ Arrynda R, *Analisis Kualitas Soal-Soal Try Out Ujian Nasional Mata Pelajaran Ipa Smp Di Kabupaten Banjarnegara*, (Skripsi : Universitas Negeri Malang, 2012), hal 15

¹⁵ M. Chabib Thoha, *Teknik Evaluasi Pendidikan*, (Jakarta: PT. RajaGrafindo Persada. 1996), h. 118

¹⁶ Arrynda R, Op., Cit, hal 16

Dalam menggunakan teknik atau metode ini pengetes hanya memiliki satu seri tes tetapi dicobakan dua kali, kemudian hasil dari kedua kali tes tersebut dihitung korelasinya.

3) Metode belah dua (*split-half method*)

Dalam metode ini item atau butir soal dibelah menjadi dua. Ada dua cara membelah butir soal yaitu, membelah atas item-item genap dan item-item ganjil yang selanjutnya disebut belahan ganjil-genap, dan membelah atas item-item awal dan item-item akhir yaitu separo jumlah pada nomor-nomor awal dan separo pada nomor-nomor akhir yang selanjutnya disebut belahan awal-akhir. Nilai dari reliabilitas diberi lambang r yang dapat dicari besarnya dengan menggunakan rumus KR- 21 atau Spearman-Brown.¹⁷

Kriteria lain yang dapat digunakan untuk menentukan kualitas tes adalah indeks kesukaran dan daya pembeda.

1) Tingkat Kesukaran

Tingkat kesukaran suatu butir soal, yang disimbolkan dengan p_i , merupakan salah satu parameter butir soal yang sangat berguna dalam penganalisan suatu tes. Hal ini disebabkan karena dengan melihat parameter butir ini, akan diketahui seberapa baiknya kualitas suatu butir soal. Jika p_i mendekati 0, maka soal tersebut terlalu sukar, sedangkan jika p_i mendekati 1, maka soal tersebut terlalu mudah, sehingga perlu dibuang. Hal ini disebabkan karena butir tersebut tidak dapat membedakan kemampuan seorang siswa dengan siswa lainnya.

Secara umum indeks kesukaran suatu butir sebaiknya terletak pada interval 0,3 – 0,7. Pada interval ini, informasi tentang kemampuan siswa akan diperoleh secara maksimal. Dalam merancang indeks kesukaran suatu perangkat tes, perlu

¹⁷Kusaeri, *Acuan dan Teknik Penilaian Proses dan Hasil Belajar dalam Kurikulum 2013*, (Yogyakarta : Ar-Ruzz Media, 2014), hal 65

dipertimbangkan tujuan penyusunan perangkat tes tersebut. Untuk menentukan indeks kesukaran dari suatu butir pada perangkat tes pilihan ganda, digunakan persamaan sebagai berikut ¹⁸:

$$p_i = \frac{\sum B}{N}$$

p_i = proporsi menjawab benar pada butir soal tertentu.

$\sum B$ = banyaknya peserta tes yang menjawab benar.

N = jumlah peserta tes yang menjawab.

Namun, dalam teori respon butir Tingkat kesukaran butir dinilai dalam skala $-\infty \leq b \leq \infty$. Butir yang dinyatakan baik adalah butir yang memiliki tingkat kesukaran (b_i) berkisar antara $-2 \leq b \leq 2$.¹⁹ Butir yang memiliki tingkat kesukaran mendekati +2 menunjukkan bahwa butir tersebut memiliki tingkat kesukaran yang tinggi. Begitupun dengan sebaliknya, butir yang memiliki tingkat kesukaran mendekati -2 menunjukkan bahwa butir tersebut memiliki tingkat kesukaran yang rendah atau soal tersebut termasuk kedalam katagori mudah.

4) Daya Pembeda

Untuk menentukan daya pembeda, dapat digunakan indeks diskriminasi, indeks korelasi biserial, indeks korelasi point biserial, dan indeks keselarasan. Pada analisis butir dalam penelitian ini, hanya digunakan indeks korelasi point biserial. Koefisien korelasinya untuk suatu butir tes ditentukan dengan rumus²⁰:

$$D = \frac{BA - BB}{\frac{1}{2} N} \text{ atau } D = \frac{2 (BA - BB)}{N}$$

Keterangan :

D = Daya pembeda

¹⁸Ibid, hal 106

¹⁹ R.K. Hamblen & Swaminathan, Hariharan & Rogers, *Fundamental of Item Response Theory*, (Newbury Park : Sage Publication Inc, 1991), hal 13

²⁰Ibid, hal 108

BA = Jumlah jawaban benar pada kelompok atas
 BB = Jumlah jawaban benar pada kelompok bawah
 N = Jumlah siswa yang mengerjakan tes

Pada suatu butir soal, indeks daya beda dikatakan baik jika lebih besar atau sama dengan 0,3. Indeks daya pembeda suatu butir yang kecil nilainya akan menyebabkan butir tersebut tidak dapat membedakan siswa yang kemampuannya tinggi dan siswa yang kemampuannya rendah.

Daya beda butir dalam teori respon butir secara teoritik parameternya didefinisikan pada skala $(-\infty, +\infty)$. Nilai yang negatif menunjukkan bahwa butir tersebut tidak baik. Nilai yang positif menunjukkan bahwa tes tersebut berfungsi dengan baik dalam mengukur kemampuan yang hendak diukur. Semakin kecil koefisien daya beda menunjukkan soal tersebut berdaya beda kurang baik, sedangkan semakin besar koefisien daya beda menunjukkan soal berdaya beda baik. Nilai daya beda yang baik berkisar antara 0 sampai dengan 2.²¹

B. Teori Klasik

Salah satu teori pengukuran yang tertua didunia pengukuran *behavioral* adalah *classical true-score theory*. Teori ini dalam bahasa Indonesia sering disebut dengan teori tes klasik. Teori tes klasik merupakan sebuah teori yang mudah dalam penerapannya serta model yang cukup berguna dalam mendeskripsikan bagaimana kesalahan dalam pengukuran dapat mempengaruhi skor pengamatan.

Inti teori klasik adalah asumsi-asumsi yang dirumuskan secara sistematis serta dalam jangka waktu yang lama. Dari asumsi-asumsi tersebut kemudian dijabarkan dalam beberapa kesimpulan. Ada tujuh asumsi yang ada dalam teori tes klasik ini. Allen & Yen menguraikan asumsi-asumsi teori klasik sebagai berikut²²:

a) $X = T+E$

²¹ R.K. Hamblton & Swaminathan, Hariharan & Rogers, Op. Cit., hal 15

²² M. J Allen & W. M Yen, *Introduction to measurement theory*, (California: Brookd/Cole Publishing Company, 1979), hal 60-67

Dengan :

$X = \text{Observed Score}$ (Skor tampak). Skor yang diperoleh dari pengukuran, masih mengandung komponen:

- 1) T: True score, menggambarkan yang sesungguhnya pada subjek (tidak diketahui)
- 2) E: error pengukuran (tidak diketahui)

Skor yang diperoleh subjek dalam suatu tes merupakan hasil penjumlahan dari 2 komponen: skor murni (T) yang bersifat konstan dan skor kesalahan (E) yang bersifat random yang bervariasi dalam berbagai peristiwa administrasi tes, sehingga X pun bervariasi.

b) $\epsilon(X) = T$

Asumsi kedua adalah bahwa skor murni (T) merupakan nilai harapan $\epsilon(X)$. Dengan demikian skor murni adalah nilai rata-rata skor perolehan teoretis sekiranya dilakukan pengukuran berulang-ulang (sampai tak terhingga) terhadap seseorang dengan menggunakan alat ukur.

c) $\rho_{et} = 0$.

Skor kesalahan E dan skor murni T yang dicapai oleh suatu populasi subjek dalam pengtesan tidak saling berkorelasi. Subjek yang mendapat skor T tinggi dalam tes tidak secara sistematis mendapat skor E lebih besar/positif atau lebih kecil/negatif dibanding subjek lain yang mendapat skor T lebih rendah dengan tes yang sama.

d) $\rho_{e1e2} = 0$

Asumsi keempat menyatakan bahwa korelasi antara kesalahan pada pengukuran pertama dan kesalahan pada pengukuran kedua adalah nol. Artinya bahwa skor-skor kesalahan pada dua tes untuk mengukur hal yang sama tidak memiliki korelasi (hubungan). Dengan demikian besarnya kesalahan pada suatu tes tidak bergantung kesalahan pada tes lain.

e) $\rho_{e1t2} = 0$

Asumsi kelima menyatakan bahwa jika terdapat dua tes untuk mengukur atribut yang sama maka skor kesalahan pada tes pertama tidak berkorelasi dengan skor murni pada tes kedua. Asumsi ini akan gugur jika salah satu tes tersebut ternyata mengukur aspek yang berpengaruh terhadap terjadinya kesalahan pada pengukuran yang lain.

f) Tes-tes yang Pararel

Asumsi keenam teori tes klasik adalah menyajikan tentang pengertian tes yang pararel. Dua perangkat tes dapat dikatakan sebagai tes-tes yang pararel jika skor-skor populasi yang menempuh kedua tes tersebut mendapat skor murni yang sama ($T = T'$) dan varian skor-skor kesalahannya sama ($\sigma_e^2 = \sigma_e'^2$). Dalam prakteknya, asumsi keenam teori ini sulit terpenuhi.

g) Tes-tes dengan τ yang secara esensial ekuivalen

Jika 2 tes menghasilkan skor tampak x_1 dan x_2 dan skor tersebut memenuhi asumsi 1 sampai dengan asumsi 5, dan untuk semua populasi subjek $T_1 = T_2 + C$, dimana C adalah suatu konstanta, maka tes-tes tersebut *essentially τ equivalent test*.

Sumintono & Widhiarso memberikan catatan penggunaan skor mentah/*raw score* sebagai ukuran prestasi yang memiliki beberapa kelemahan, di antaranya sebagai berikut :²³

1. Skor mentah pada dasarnya bukanlah hasil pengukuran. Lebih tepatnya skor mentah adalah jumlah jawaban benar dari soal yang dikerjakan siswa.
2. Skor mentah adalah informasi awal. Skor mentah juga biasanya dinyatakan dalam persentase (%) yang tidak lain hanyalah ringkasan data berupa angka, tetapi tidak memberikan data dari suatu pengukuran.
3. Skor mentah memiliki makna kuantitatif yang lemah. Makna kuantitatif dari skor mentah yang didapat akan berbeda, bergantung pada banyaknya soal, sedangkan persentase jawaban betul selalu bergantung pada tingkat kesulitan soal.
4. Skor mentah tidak menunjukkan kemampuan seseorang terhadap tugas tertentu. Skor mentah juga tidak bisa banyak menjelaskan tingkat kesulitan soalnya.
5. Skor mentah dan persentase jawaban benar tidak selalu bersifat linier. Dalam sebuah tes yang bersifat linier, siswa yang memiliki skor 15 (skala 0 hingga 100) selalu memiliki kemampuan lebih tinggi dibanding yang memiliki skor 10.

²³ <http://www.semestapsikometrika.com/2017/07/sekilas-tentang-teori-tes-klasik.html> , diakses pada tanggal 30 Oktober 2017

Namun, secara empirik terkadang keduanya memungkinkan memiliki kemampuan yang sama.

Meskipun secara sekilas teori tes klasik memiliki banyak keterbatasan, namun bukan berarti pendekatan ini tidak memiliki kelebihan. Beberapa kelebihan teori tes klasik dibandingkan teori respon butir diantaranya adalah terkait dengan analisis yang lebih sederhana dan mudah dipahami. Teori ini memiliki nilai praktis yang tinggi dalam menerangkan masalah reliabilitas dan validitas, disamping pemahamannya yang tidak menuntut pengetahuan mendalam mengenai fungsi distribusi statistik dan model-model matematikanya.²⁴ Selain itu dalam melakukan analisis, sampel yang diperlukan bisa lebih sedikit jika dibandingkan dengan pendekatan teori respon butir, sehingga sangat cocok dilakukan untuk penilaian di kelas dan perbaikan evaluasi pembelajaran.

C. Teori Respon Butir

Untuk mengatasi kelemahan-kelemahan yang ada pada teori klasik, para ahli pengukuran berusaha untuk mencari alternatif. Model yang diinginkan harus mempunyai sifat-sifat²⁵: (1) karakteristik butir tidak tergantung kepada kelompok peserta tes yang dikenai butir soal tersebut, (2) skor yang menyatakan kemampuan peserta tes tidak tergantung pada tes, (3) model dinyatakan dalam tingkatan (level) butir soal, tidak dalam tingkatan tes, (4) model tingkat tidak memerlukan tes paralel untuk menghitung koefisien reliabilitas, dan (5) model menyediakan ukuran yang tepat untuk setiap skor kemampuan. Model alternatif yang dapat mempunyai ciri-ciri itu adalah model pengukuran yang disebut teori respon butir (*Item Response Theory*).

Teori tes modern atau biasa disebut dengan teori respon butir (*item response theory*) dikembangkan oleh para ahli pengukuran bidang psikologi dan pendidikan sebagai upaya meminimalkan kekurangan-kekurangan yang ada dalam teori tes klasik.

²⁴ S. Azwar, *Dasar-dasar Psikometri*, (Yogyakarta: Pustaka Pelajar, 2015), hal 35

²⁵ Suwanto, *Teori Tes Klasik dan Teori Tes Modern*, (Sukoharjo : Universitas Veteran Bangun Nusantara Sukoharjo, 2011), Jurnal Pendidikan Volume 1 No.1, hal 75

Perhitungan dalam analisis butir berdasarkan teori ini dapat dilakukan dengan menggunakan bantuan program komputer.²⁶

Teori respon butir ini menyatakan bahwa perilaku seseorang dapat dijelaskan sampai derajat tertentu dengan karakteristik orang tersebut.²⁷ Karakteristik ini bermacam-macam, misalnya : kemampuan verbal, kemampuan *numeric reasoning*, psikomotor dan lain-lain. Karakteristik ini disebut *trait*. Posisi seseorang pada suatu trait dapat digunakan untuk mengestimasi besarnya kemampuannya. Trait ini sering dinyatakan sebagai dimensi kemampuan seseorang.

Jumlah trait pada diri seseorang pada umumnya lebih dari satu dan setiap trait diasumsikan menentukan unjuk kerja seseorang dalam mengerjakan tes. Trait yang banyak disebut latent trait atau variabel laten. Misalkan suatu tes yang terdiri dari n butir dan mengukur k trait, maka variabel laten tersebut dapat dinyatakan dengan vector $\theta = (\theta_1, \theta_2, \theta_3, \dots, \theta_k)^{28}$.

1. Asumsi Teori Respon Butir

a) Unidimensi

Asumsi untuk model teori respon butir secara mendalam digunakan, sehingga hanya satu kemampuan yang diukur dengan butir-butir tes tersebut. Asumsi ini dinamakan unidimensi. Sebuah tes dikatakan unidimensi jika butir tes tersebut secara statistik adalah tidak bebas di dalam populasi yang dilibatkan. Suatu ciri atau sifat terpendam tunggal ada sehingga butir-butir tersebut secara statistik adalah bebas di dalam masing-masing subpopulasi dari peserta tes yang homogen.

b) Independen lokal

Asumsi lain dalam teori respon butir adalah asumsi independen lokal. Independen lokal dibagi dua yakni independensi lokal terhadap respons peserta tes dan independensi lokal terhadap butir tes. Independen lokal terhadap respon peserta tes, memiliki arti bahwa betul

²⁶ Ibid, hal 76

²⁷ Djemari Mardapi, *Teknik Penyusunan Instrumen Tes dan Non Tes*, (Yogyakarta : Mitra Cendikia Press, 2008), hal 145

²⁸ Herwelis, *Analisis Kualitas Soal Ujian Akhir Sekolah (Uas) Mata Pelajaran Matematika Tingkat Smk Kabupaten Takalar Tahun Pelajaran 2012/2013*, Thesis : Program Pascasarjana Universitas Negeri Makasar, 2013), hal 53

salahnya peserta menjawab sebuah butir tidak berpengaruh dan tidak dipengaruhi oleh betul salahnya peserta lain dalam menjawab betul butir tersebut. Sedangkan independensi lokal terhadap butir, memiliki arti bahwa betul salahnya seorang peserta tes menjawab sebuah butir tidak terpengaruh dan tidak dipengaruhi oleh betul salahnya peserta tes dalam menjawab betul butir yang lain.

c) Invarian

Asumsi berikutnya adalah invarian. Invariansi pensekoran pada teori tes modern adalah invarians (tidak berubah) terhadap butir tes serta terhadap peserta tes. Invariansi parameter-parameter butir tes melalui kelompok peserta tes merupakan karakteristik yang paling penting dalam teori respon butir.²⁹

Dari penjelasan tentang asumsi-asumsi model teori respon butir, maka ada tiga asumsi yang mendasarinya yakni ; unidimensi, independen lokal, dan invariansi parameter. Unidimensi jika setiap butir soal hanya mengukur satu kemampuan. Independen lokal terjadi jika faktor-faktor yang memengaruhi kemampuan konstan, maka respon seorang peserta tes terhadap pasangan butir manapun akan independen secara statistik satu sama lain. Artinya, jawaban seorang peserta tes terhadap suatu butir tidak memengaruhi jawabannya terhadap butir yang lain. Sedangkan Invarian parameter artinya karakteristik butir soal tidak tergantung pada karakteristik kemampuan peserta tes dan karakteristik kemampuan peserta tes tidak tergantung pada karakteristik butir soal.

Menurut Sudaryono karakteristik butir dalam teori respon butir terdiri dari daya beda butir, taraf sukar butir dan faktor kebetulan menjawab betul (peluang menebak gussing) pada butir dinyatakan berturut-turut dengan huruf a, b, dan c. Parameter peserta tes adalah kemampuan peserta tes yang dinyatakan dengan θ . Kemampuan peserta tes terhadap butir

²⁹ Devi Dwi Kurniawan, *Analisis Kualitas Soal Ujian Akhir Semester Matematika Berdasarkan Teori Respon Butir*, Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS, 2015, hal 124

ke- j dinyatakan dalam bentuk probabilitas jawaban betul $P_i(\theta)$. Sekor responden mencerminkan kemampuan responden sehingga sekor responden dan kemampuan responden merupakan parameter responden. Kemampuan responden merupakan suatu kontinum dari rendah ke tinggi. Biasanya sekor responden tinggi menunjukkan kemampuan tinggi dan sekor responden rendah menunjukkan kemampuan responden rendah. Biasanya pada sekor responden tinggi atau kemampuan tinggi, proporsi jawaban betul juga tinggi.³⁰

Pada karakteristik butir, proporsi jawaban betul dikenal sebagai probabilitas jawaban betul.

Skor responden (θ) = kemampuan responden

Proporsi jawaban betul $P(\theta)$ = probabilitas jawaban betul

Untuk butir ke- i , probabilitas jawaban betul berkaitan dengan kemampuan responden θ . Makin tinggi kemampuan responden θ , makin besar pula probabilitas jawaban betul. Hubungan di antara probabilitas jawaban betul pada butir ke- i dengan kemampuan responden θ adalah $P_i(\theta) = f(q)$.³¹

2. Model- Model Logistik Teori Respon Butir

Dalam teori respon butir ada tiga model logistik perhitungan dalam karakteristik kurva butir, yakni model logistik satu parameter (1PL), model logistik dua parameter (2PL) dan model logistik tiga parameter (3PL). Model-model ini akan menyajikan persamaan hitungan hubungan antara kemungkinan jawaban benar terhadap kemampuan. Model-model karakteristik butir tergantung pada bentuk matematis fungsi karakteristik butirnya dan banyaknya parameter yang dilibatkan dalam model yang digunakan.³²

a) Model Logistik Satu Parameter / 1PL (*Rasch Model*)

Model logistik satu parameter (model rasch), yaitu untuk menganalisis data yang hanya menitikberatkan

³⁰Sudaryono, *Dasar-Dasar Evaluasi Pembelajaran*, (Yogyakarta: Graha Ilmu,2012), hal 175

³¹ Ibid, hal 176

³² Heri Retnawati, *Teori Respons Butir dan Penerapannya*, (Yogyakarta : Nuha Medika, 2014), hal 18

pada parameter tingkat kesukaran. Frank B. Baker merumuskan Model-IPL sebagai berikut³³.

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

Keterangan:

$P_i(\theta)$ = Peluang menjawab benar peserta yang berkemampuan θ pada butir ke- i

b_i = Tingkat kesukaran butir ke- i

e = Bilangan natural yang bernilai 2,712

n = Jumlah butir soal

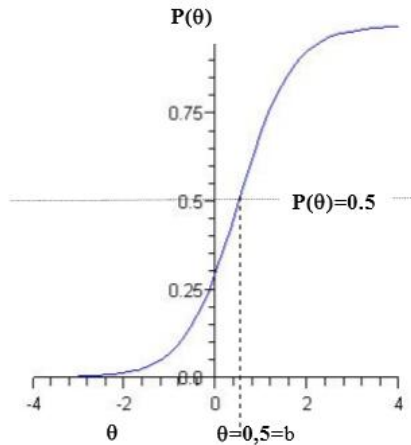
D = Konstanta bernilai 1,7 sebagai simpangan baku distribusi logistic

Parameter b_i merupakan suatu titik pada skala kemampuan agar peluang menjawab benar sebesar 50%. Misalkan suatu butir tes mempunyai parameter $b_i = 0,3$, artinya diperlukan kemampuan minimal 0,3 pada skala untuk dapat menjawab benar dengan peluang 50%. Semakin besar nilai parameter b_i , maka semakin besar kemampuan yang diperlukan untuk menjawab benar dengan peluang 50%. Dengan kata lain, semakin besar nilai parameter b_i , maka semakin sulit butir soal tersebut.

Hubungan peluang menjawab benar $P_i(\theta)$ dengan tingkat kemampuan peserta (θ) dapat digambarkan sebagai kurva karakteristik butir. Di bawah ini karakteristik kurva yang didapatkan dengan $b = 0,5$.³⁴

³³ Frank B. Baker, *The Basic Of Item Response Theori*, (America :ERIC Clearinghouse On Assesment and Evaluation, 2001), hal 25

³⁴ Heri Retnawati, *Teori Respons Butir dan Penerapannya*, (Yogyakarta : Nuha Medika, 2014), hal 20



Gambar 2.1
Kurva Karakteristik Butir untuk Model Satu
Parameter dengan $b = 0,5$

- b) Model Logistik Dua Parameter 2PL (*Birnbaum Model*)
 Model logistik dua parameter, yaitu untuk menganalisis data yang hanya menitikberatkan pada parameter tingkat kesukaran dan daya pembeda soal. Persamaan fungsi logistik untuk dua parameternya adalah³⁵:

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

$$i = 1, 2, 3, \dots, n$$

θ = Tingkat kemampuan peserta tes

$P_i(\theta)$ = Peluang menjawab benar peserta yang berkemampuan θ pada butir ke- i

a_i = Parameter daya beda butir

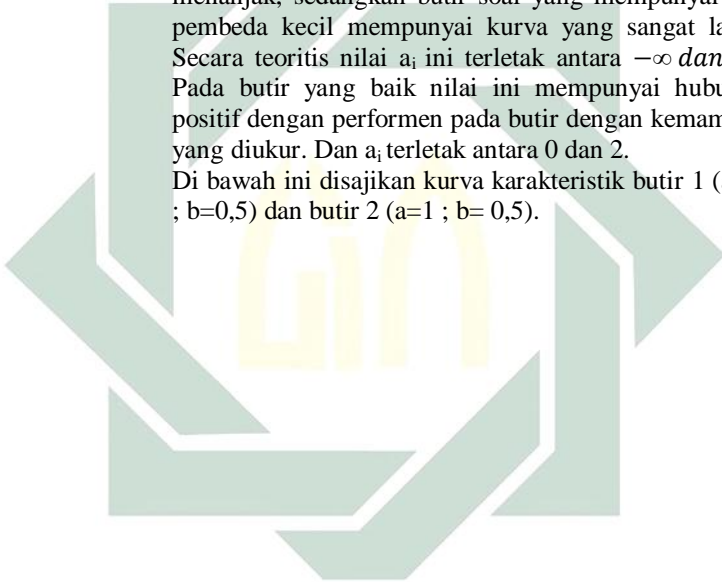
b_i = Tingkat kesukaran butir ke- i

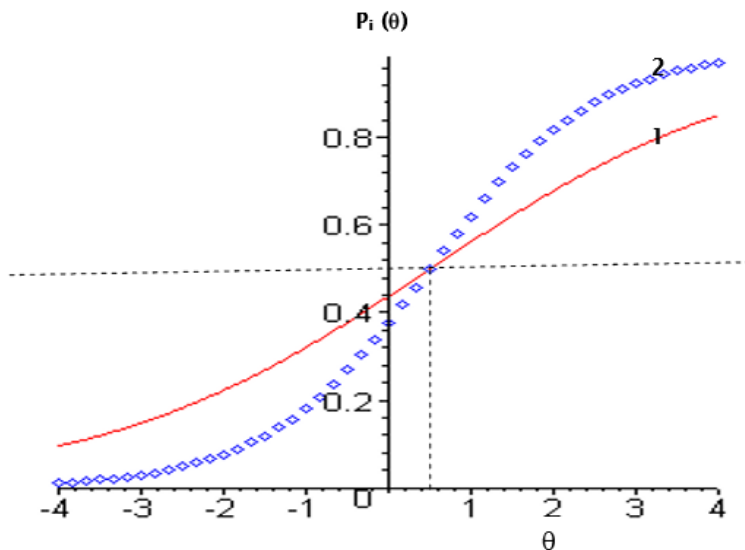
³⁵ Frank B. Baker, Op., Cit, hal 26

e = Bilangan natural yang bernilai 2,712
 n = Jumlah butir soal
 D = Konstanta bernilai 1,7 sebagai simpangan baku distribusi Logistic

Parameter a_i merupakan indeks daya pembeda yang dimiliki butir ke- i . Pada kurva karakteristik, a_i proporsional terhadap koefisien arah garis singgung (*Slope*) pada titik $\theta = b$. Butir soal yang memiliki daya pembeda yang besar mempunyai kurva yang sangat menanjak, sedangkan butir soal yang mempunyai daya pembeda kecil mempunyai kurva yang sangat landai. Secara teoritis nilai a_i ini terletak antara $-\infty$ dan $+\infty$. Pada butir yang baik nilai ini mempunyai hubungan positif dengan performen pada butir dengan kemampuan yang diukur. Dan a_i terletak antara 0 dan 2.

Di bawah ini disajikan kurva karakteristik butir 1 ($a=0,5$; $b=0,5$) dan butir 2 ($a=1$; $b=0,5$).





Gambar 2.2
Kurva Karakteristik Butir untuk Model 2 Parameter
dengan Karakteristik Butir 1 ($a=0,5$; $b=0,5$) dan
butir 2 ($a=1$; $b= 0,5$)

Berdasarkan gambar tersebut, jika indeks daya pembeda butir 1 lebih rendah dibandingkan butir 2, maka akan nampak bahwa kurva karakteristik butir 1 lebih landai dibandingkan butir 2.

- c) Model Logistik Tiga Parameter 3PL (*Birnbaum Model*)
 Model logistik tiga parameter, yaitu untuk menganalisis data yang menitikberatkan pada parameter tingkat kesukaran soal, daya pembeda soal dan peluang menebak (guessing). Dengan adanya peluang menebak pada model logistik tiga parameter, memungkinkan subjek yang memiliki kemampuan rendah mempunyai

peluang untuk menjawab butir soal dengan benar. Adapun persamaannya adalah sebagai berikut³⁶:

$$\text{Rumus: } P_i(\theta) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta - b_i)}}$$

Dimana:

b_i = Parameter tingkat kesukaran soal

a_i = Parameter daya pembeda soal

c_i = Parameter/peluang tebakan benar

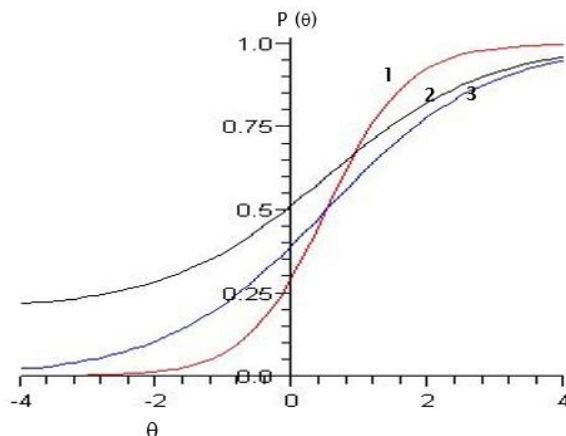
θ = Level/tingkat kemampuan

Nilai kemampuan peserta (θ) terletak diantara -3 dan +3, sesuai dengan daerah asal distribusi normal. Pernyataan ini merupakan asumsi yang mendasari besar nilai b_i . Secara teoritis, nilai b_i terletak di $-\infty$ dan $+\infty$. Suatu butir dikatakan baik jika nilai ini berkisar antara -2 dan +2.³⁷ Jika nilai b_i mendekati -2, maka indeks kesukaran butir sangat rendah, sedangkan jika nilai b_i mendekati +2, maka indeks kesukaran butir sangat tinggi untuk suatu kelompok peserta tes.

Peluang menjawab benar pada saat kemampuan peserta tes sangat rendah dilambangkan dengan c_i , yang disebut dengan peluang tebakan. Parameter ini merupakan suatu kemungkinan asimtot bahwa yang tidak nol (*nonzero lower asymptote*) 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 diasumsikan lebih kecil daripada nilai yang akan dihasilkan jika peserta tes menebak secara acak jawaban pada suatu butir. Gambar dibawah ini menyajikan suatu kurva karakteristik butir 1 ($a=1, b=0,5, c=0$), butir 2 ($a=0,5, b=0,5, c=0$) dan butir 3 ($a=0,5, b=0,5, c=0,2$).

³⁶ Ibid, hal 28

³⁷ Heri Retnawati, "Teori Respons Butir dan Penerapannya", (Yogyakarta : Parama Publishing, 2014), hal 16



Gambar 2.3.
Kurva Karakteristik Butir untuk Model Tiga-Parameter dengan Butir 1 ($a=1, b=0,5, c=0$), Butir 2 ($a=0,5, b=0,5, c=0$) dan Butir 3 ($a=0,5, b=0,5, c=0,2$)

Mencermati gambar tersebut, nampak bahwa pada skala kemampuan peserta tes yang sangat rendah ($\theta = -4$), peluang menjawab benar butir 3 sebesar 0,2 sedangkan pada butir 1 dan butir 2 mendekati 0.

Pada suatu butir tes, nilai c_i ini berkisar antara 0 dan 1. Suatu butir dikatakan baik jika nilai c_i tidak lebih dari $1/k$, dengan k banyaknya pilihan. Jadi misalkan pada suatu perangkat tes pilihan ganda dengan 4 pilihan untuk setiap butir tesnya, butir ini dikatakan baik jika nilai c_i tidak lebih dari 0,25.³⁸

d) Fungsi Informasi

Pada teori respon butir, ketidakpastian butir berkaitan dengan parameter ciri peserta (θ). Jika peserta dengan kemampuan tinggi mengerjakan butir yang mudah, maka

³⁸N, Hidayanto, *Karakteristik Butir Soal Ulangan Kenaikan Kelas Persiapan Bank Soal Bahasa Inggris*, (Yogyakarta : Jurnal Penelitian dan Evaluasi Pendidikan, Nomor 1,2011), hal 23.

peserta tersebut mempunyai peluang menjawab benar yang sangat tinggi. Sebaliknya, jika peserta dengan kemampuan rendah menjawab butir yang sulit, maka peluang dia menjawab dengan benar akan sangat kecil.³⁹ Pada kasus ini, tidak diketahui seberapa besar kemampuan peserta sehingga ketidakpastian tentang kemampuan peserta sangat tinggi dan fungsi informasi menjadi rendah. Tetapi, jika peserta dengan kemampuan tinggi menjawab butir yang sukar, peserta dengan kemampuan sedang menjawab butir yang sedang tingkat kesulitannya, serta peserta dengan kemampuan rendah menjawab butir yang rendah tingkat kesulitannya, maka jawaban yang diperoleh akan bervariasi. Dalam hal ini, ketidakpastian tentang kemampuan peserta menjadi lebih rendah karena besar kemampuan peserta diketahui sehingga nilai fungsi informasi menjadi lebih tinggi

Dalam teori klasik, melalui koefisien reliabilitas, kita menilai mutu perangkat ujites yang terdiri atas sejumlah butir sedangkan dalam teori respon butir, kita mencapai tujuan yang sama melalui fungsi informasi ujites. Pada respon butir skor butir yang membentuk fungsi informasi ujites adalah independen satu dengan yang lainnya. Perubahan pada satu sektor hanya mengubah jumlah nilai pada fungsi informasi ujites dan tidak mengganggu nilai lainnya. Pada teori respon butir menggunakan istilah informasi untuk menyatakan reliabilitas dan validitas butir soal ataupun tes. Fungsi informasi sangat berguna untuk ; konstruksi tes, seleksi butir, penilaian presisi pengukuran, komparasi sejumlah tes, penentuan bobot dalam penskoran dan komparasi berbagai metode penskoran. Setiap kurva karakteristik butir mempunyai fungsi informasi butir. Besarnya fungsi informasi tergantung pada daya pembeda, tingkat kesukaran dan dugaan semu⁴⁰.

³⁹ Sudaryono, *Implementasi Teori Respon Butir (Item Response Theory) pada Penilaian Hasil Belajar Akhir di Sekolah*, Jurnal Pendidikan dan Kebudayaan Volume 17 Nomor 6 Tahun 2011, hal 726

⁴⁰ Heri Retnawati, *Validitas, Reliabilitas dan Karakteristik Butir*, (Yogyakarta : Parama Publishing, 2016) , hal 178

Fungsi informasi butir, dilambangkan dengan $I_i(\theta)$, dinyatakan dalam persamaan⁴¹:

$$I_i(\theta) = \frac{\left[P_i'(\theta) \right]^2}{P_i(\theta) Q_i(\theta)},$$

$$i = 1, 2, 3, \dots, n$$

Dimana:

$I_i(\theta)$ = Informasi yg diberikan butir i

$P_i'(\theta)$ = Turunan pertama dari $P_i(\theta)$

$Q_i(\theta)$ = $1 - P_i(\theta)$

Fungsi informasi butir tergantung pada kemiringan (*slope*) fungsi respon butir dan variansi bersyarat pada masing-masing tingkat kemampuan (θ). Semakin besar nilai *slope* dan semakin kecil nilai variansi, maka semakin banyak informasi yang dihasilkan, dan akan memperkecil kesalahan pengukuran.

Selanjutnya dalam teori respon butir juga dikenal fungsi informasi tes. Fungsi informasi tes merupakan jumlah dari fungsi informasi butir penyusun tes tersebut, yaitu⁴²:

$$I(\theta) = \sum_{i=1}^n I_i(\theta) = \sum_{i=1}^n \frac{\left[P_i'(\theta) \right]^2}{P_i(\theta) Q_i(\theta)}$$

Banyaknya informasi yang dikandung oleh fungsi informasi tes dipengaruhi oleh kualitas dan jumlah butir penyusun tes tersebut. Semakin curam *slope* fungsi respon butir, semakin banyak informasi yang dikandung. Dan, semakin kecil variansi butir semakin besar pula informasi yang diberikan oleh tes. Fungsi informasi tes tidak tergantung pada kombinasi tertentu dari butir-butir penyusun tes.

⁴¹Devi Dwi Kurniawan, *Analisis Kualitas Soal Ujian Akhir Semester Matematika Berdasarkan Teori Respon Butir*, Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS, 2015, hal 126

⁴² Ibid, hal 126

Masing-masing butir penyusun tes independen satu sama lain.⁴³

D. Analisis Butir Soal

Analisis butir tes hasil belajar dilakukan untuk menguji bagaimanakah keadaan butir-butir soal yang digunakan untuk suatu tes. Pada analisis butir soal, kita menguji keadaan soal-soal tersebut tiap butirnya. Analisis terhadap butir-butir tes yang telah dijawab oleh peserta didik mempunyai dua tujuan penting. Pertama, jawaban soal itu merupakan informasi diagnostik untuk meneliti pelajaran dari kelas itu dan kegagalan-kegagalan belajarnya, serta selanjutnya untuk membimbing ke arah cara belajar yang lebih baik. Kedua, jawaban-jawaban terhadap soal yang terpisah dan perbaikan soal-soal yang didasarkan atas jawaban-jawaban itu merupakan basis bagi persiapan yang lebih baik untuk tes-tes tahun berikutnya.⁴⁴

Tujuan utama analisis soal adalah untuk mendapatkan informasi tentang karakteristik setiap butir soal, baik secara kualitatif, kuantitatif. Hasil suatu analisis soal dapat digunakan untuk mengetahui apakah soal akan berfungsi dengan baik atau tidak. Pada dasarnya cara yang ditempuh dalam analisis soal adalah melalui telaah (analisis kualitatif) dan analisis berdasarkan data hasil uji coba (analisis kuantitatif).

a. Analisis Secara Kualitatif

Analisis secara kualitatif terhadap butir soal pada dasarnya adalah penelaahan butir soal ditinjau dari segi kaidah penulisan soal yaitu⁴⁵:

- 1) Isi atau materi
- 2) Konstruksi
- 3) Bahasa.

Soal yang baik harus memenuhi ketiga kaidah penulisan soal tersebut. Sedangkan kaidah penulisan soal pilihan ganda menurut Kusaeri antara lain⁴⁶:

⁴³ Herwelis, *Analisis Kualitas Soal Ujian Akhir Sekolah (Uas) Mata Pelajaran Matematika Tingkat Smk Kabupaten Takalar Tahun Pelajaran 2012/2013*, Thesis : Program Pascasarjana Universitas Negeri Makasar, 2013), hal 64

⁴⁴ Mulyadi, *Evaluasi Pendidikan*, (Malang: UIN Maliki Press, 2010), hlm. 109.

⁴⁵ Suwanto, *Teori Tes Klasik dan Teori Tes Modern*, Jurnal Widyatama Volume:20 Nomor 1, (Sukoharjo : Universitas Veteran Bangun Nusantara, 2011), hal 24

- 1) Soal sesuai dengan indikator.
 - 2) Pilihan jawaban harus homogen dan logis ditinjau dari segi materi
 - 3) Setiap soal harus mempunyai satu jawaban yang benar atau paling benar
 - 4) Pokok soal harus dirumuskan secara jelas dan tegas
 - 5) Rumusan pokok soal dan pilihan jawaban harus berupa pernyataan yang diperlukan saja
 - 6) Pokok soal jangan memberi petunjuk kearah jawaban benar
 - 7) Panjang rumusan pilihan jawaban harus relative sama
 - 8) Pilihan jawaban jangan megandung “Semua pilihan jawaban diatas salah”, atau “Semua pilihan jawaban diatas benar”.
 - 9) Pilihan jawaban yang berbentuk angka atau waktu harus disusun berdasarkan urutan besar kecilnya nilai angka tersebut, atau kronologis waktunya
 - 10) Gambar geafik, table, diagram, dan sejenisnyayang terdapat pada soal harus jelas dan berfungsi
 - 11) Butir soal jangan bergantung pada jawaban soal sebelumnya
 - 12) Rumusan soal harus menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia
 - 13) Pilihan jawaban jangan mengulang kata atau frase yang buka merupakan satu kesatuan pengertian. Letakkan kata atau frase tersebut pada pokok soal
- Analisis soal secara kualitatif dilakukan sebelum suatu soal diujikan. Analisis ini belum memberikan gambaran tentang karakteristik psikometri soal. Oleh karena itu, untuk membuktikan bahwa soal-soal itu sudah baik perlu diuji cobakan terhadap sejumlah siswa.
- b. Analisis Secara Kuantitatif
- Penelaahan soal secara kuantitatif maksudnya adalah penelaahan butir soal didasarkan pada data empirik dari butir soal yang bersangkutan. Data empiric ini diperoleh dari soal yang telah diujikan. Ada dua pendekatan dalam

⁴⁶ Kusaeri, *Acuan dan Teknik Penilaian Proses dan Hasil Belajar dalam Kurikulum 2013*, (Yogyakarta : Ar-Ruzz Media, 2014), hal 71-82

analisis secara kuantitatif, yaitu pendekatan secara klasik dan modern (Respon Butir).⁴⁷

Dalam penelitian ini, kami hanya akan menganalisis soal secara kuantitatif berdasarkan teori respon butir model logistik tiga parameter. Adapun hasil yang diperlukan adalah ⁴⁸:

- a. Estimasi 3 PL : Dalam model logistik 3 parameter, estimasi parameter butir dinyatakan dalam bentuk daya beda (a_i), tingkat kesukaran butir (b_i), dan peluang menebak/ *guessing* (c_i).
- b. Fungsi informasi tes : Dalam mencari fungsi informasi tes tergantung pada kemiringan (*slope*), fungsi respon butir dan variansi bersyarat pada masing-masing tingkat kemampuan (θ). Semakin besar nilai *slope* dan semakin kecil nilai variansi, maka semakin banyak informasi yang dihasilkan, dan akan memperkecil kesalahan pengukuran.

⁴⁷www.gurupembaharu.com/home/download/panduan-analisis-butir-soal.pdf , diakses pada tanggal 14 Juni 2017

⁴⁸ Herwelis, *Analisis Kualitas Soal Ujian Akhir Sekolah (Uas) Mata Pelajaran Matematika Tingkat Smk Kabupaten Takalar Tahun Pelajaran 2012/2013*, Thesis : Program Pascasarjana Universitas Negeri Makasar, 2013), hal 100

BAB III

METODE PENELITIAN

A. Jenis Penelitian

Jenis penelitian ini menggunakan penelitian deskriptif kuantitatif. Penelitian deskriptif kuantitatif merupakan metode menggambarkan dan menginterpretasikan objek sesuai apa adanya. Penelitian ini menggunakan analisis data secara kuantitatif. Analisis secara kuantitatif digunakan untuk mendeskripsikan kualitas soal dalam segi empiris, yaitu reliabilitas, tingkat kesukaran, daya pembeda dan peluang menjawab benar (peluang menebak atau *guessing*). Jawaban siswa dalam US/M mata pelajaran matematika, akan dicermati, ditelaah, diobservasi dan kemudian dianalisis.

B. Populasi dan Sampel Penelitian

1. Populasi (Subjek Penelitian)

Subjek penelitian ini adalah jawaban US/M siswa tahun ajaran 2016-2017 di Kota Surabaya yang diperoleh dengan mengujikan kembali soal US/M kepada siswa SMP Kelas VII .

2. Teknik Pengambilan Sampel

Dalam penelitian ini digunakan teknik *random sampling* dalam pengambilan sampel. Teknik *random sampling* ini adalah pengambilan anggota-anggota sampel yang dilakukan dengan cara mengacak individu-individu anggota populasi¹.

3. Jumlah Sampel

Secara umum, ada dua cara dalam menentukan jumlah anggota sampel penelitian, yaitu cara proporsi dan cara ketelitian estimasi. Dalam penelitian ini kami menggunakan cara proporsi. Untuk cara proporsi , *Isaac* telah membuat perhitungan akhir dengan formula empiris sampel yang digunakan adalah >100 sampel.²

¹ Zaenal Arifin, *Metodologi Penelitian Pendidika Filosofi, Teori & Aplikasinya*, (Surabaya: Lentera Cendikia, 2009), hal 67

² *Ibid*, 75-76

C. Prosedur Penelitian

Adapun rancangan penelitian yang akan dilakukan oleh peneliti, dari awal hingga akhir adalah sebagai berikut:

1. Tahap Persiapan
Kegiatan yang dilakukan pada tahap ini meliputi:
 - a. Mengidentifikasi masalah.
 - b. Menyusun proposal penelitian
 - c. Melakukan seminar proposal
 - d. Melakukan perizinan pengambilan data
2. Tahap Pelaksanaan
Kegiatan yang dilakukan pada tahap ini meliputi:
 - a. Mengambil data penelitian di dinas pendidikan Kota Surabaya. Jenis data penelitian ini adalah berupa soal US/M mata pelajaran matematika tahun ajaran 2016-2017 Kota Surabaya
 - b. Menguji soal US/M mata pelajaran matematika tahun ajaran 2016-2017 kepada siswa kelas VII SMP.
 - c. Menganalisis subjek penelitian secara kuantitatif. Analisis ini meliputi reliabilitas soal, tingkat kesukaran soal dan daya pembeda soal.
 - d. Menganalisis subjek penelitian berdasarkan teori respon butir model 3 PL.
3. Tahap Akhir
Kegiatan yang dilakukan pada tahap ini yaitu menyusun laporan akhir penelitian berdasarkan data dan analisis data. Hasil yang diharapkan adalah mengetahui kualitas soal US/M mata pelajaran Matematika tahun 2017 se-tingkat SD Kota Surabaya.

D. Metode Pengumpulan Data

Pengumpulan data dalam penelitian ini adalah :

1. Metode dokumentasi
Metode pengumpulan data dengan dokumentasi ialah pengambilan data yang diperoleh melalui dokumen-dokumen. Metode dokumentasi diperoleh melalui peninggalan tertulis seperti arsip-arsip dari Dinas Pendidikan Kota Surabaya yang berupa seperangkat soal US/M tahun ajaran 2016-2017.

2. Metode Tes

Dikarenakan keterbatasan peneliti dan tidak adanya data berupa jawaban siswa di Dinas Pendidikan Kota Surabaya, maka dengan pertimbangan beberapa pihak, penelitian ini tetap dilanjutkan dengan mengujikan kembali soal US/M tahun ajaran 2016-2017 Kota Surabaya kepada siswa kelas 7 di sekolah tertentu yakni di SMP Negeri 3 Surabaya dan di SMP Islam Al Amal Surabaya. Adapun alasan memilih 2 sekolah tersebut adalah untuk meratakan data. Dimana SMPN 3 adalah sekolah dengan *great* yang tinggi dan SMP Islam Al Amal Surabaya dengan *great* yang rendah. Metode pengumpulan data dengan metode tes adalah pengambilan data yang diperoleh dengan mengujikan suatu perangkat tes.

E. Teknik Analisis Data

1. Reliabilitas Soal

Tujuan utama menghitung reliabilitas skor tes adalah untuk mengetahui tingkat ketepatan (*precision*) dan keajegan (*consistency*) skor tes. Untuk mengetahui koefisien reliabilitas tes soal bentuk pilihan ganda digunakan rumus Kuder Richardson 20 (KR-20) seperti berikut ini³.

$$KR - 21 = \left\{ \frac{k}{k-1} \right\} \left\{ 1 - \frac{M(K-M)}{K \cdot (St)^2} \right\}$$

Keterangan :

k = Banyaknya butir tes

M = Mean Skor Total

St^2 = Varians Total

Ada beberapa kriteria dalam menentukan reliabilitas suatu soal:⁴

- a. Jika tes digunakan untuk sesuatu yang penting yang cenderung dapat mempengaruhi individu secara signifikan, maka tingkat reliabilitas yang diperlukan adalah 0,90 bahkan sampai mencapai 0,95

³ Heri Retnawati, *Validitas Reliabilitas & Karakteristik Butir*, (Yogyakarta : Parama Publishing, 2016), hal 91

⁴ Cecil R Reynold, Ronal B Livingston, Victor Willson, *Measurement and Assessment in Education*, (London : Pearson Education, 2006), hal 108

- b. Jika tes digunakan untuk mengukur prestasi belajar baik secara individu maupun kelompok, maka tingkat reliabilitas yang diperlukan adalah 0,80 keatas.
 - c. Jika tes digunakan untuk mengukur prestasi belajar tingkat kelas , maka reliabilitas yang diperlukan minimal 0,70.
2. Tingkat Kesukaran
Untuk menghitung tingkat kesukaran dalam soal pilihan ganda menggunakan persamaan berikut :

$$p_i = \frac{\sum B}{N}$$

Keterangan :

p_i = Proporsi menjawab benar atau tingkat kesukaran

$\sum B$ = Banyaknya peserta tes yang menjawab benar

N = Jumlah peserta tes

Klasifikasi tingkat kesukaran soal⁵:

Tabel 3.1
Klasifikasi Tingkat Kesukaran Butir Soal

No.	Tingkat Kesukaran (b_i)	Katagori
1.	$b_i > 2.0$	Kurang Baik
2.	$-2.0 \leq b_i \leq 2.0$	Baik
3.	$b_i < -2.0$	Kurang Baik

3. Daya Pembeda Soal
Untuk mencari daya pembeda soal pilihan ganda dapat menggunakan persamaan :

⁵ Kusaeri, *Acuan dan Teknik Penilaian Proses dan Hasil Belajar dalam Kurikulum 2013*, (Yogyakarta : Ar-Ruzz Media, 2014), hal 107

$$D = \frac{BA - BB}{\frac{1}{2} N} \text{ atau } D = \frac{2(BA - BB)}{N}$$

Keterangan :

D = Daya pembeda

BA = Jumlah jawaban benar pada kelompok atas

BB = Jumlah jawaban benar pada kelompok bawah

N = Jumlah siswa yang mengerjakan tes

Adapun klasifikasi yang digunakan adalah⁶ :

Tabel 3. 2
Klasifikasi Daya Beda Butir Soal

No.	Daya Beda (a_i)	Katagori
1.	$a_i > 2.0$	Kurang Baik
2.	$0.0 \leq a_i \leq 2.0$	Baik

4. Peluang Menebak (*Guessing*)

Peluang menebak merupakan ukuran peluang peserta untuk menebak dengan benar. Suatu soal diharapkan mempunyai peluang menebak yang kecil bahkan mendekati nol. Sehingga soal yang baik adalah soal yang memiliki peluang menebak mendekati 0. Nilai peluang menebak (c_i) adalah antara 0 sampai dengan 1. Suatu butir dikatakan baik jika nilai c_i nya tidak lebih dari $1/k$, dengan k adalah banyaknya pilihan jawaban.⁷ Dalam penelitian ini pilihan jawabannya adalah 4, maka nilai dari c_i untuk setiap butir seharusnya tidak lebih dari $1/4$ atau tidak lebih dari 0.25.

⁶Ibid, hal 109

⁷http://www.academia.edu/29987008/ANALISIS_ITEM_DENGAN_PENDEKATAN_IR_T_DENGAN_BANTUAN_APLIKASI_PROGRAM_BILOG-MGhal , diakses pada tanggal 30 Desember 2017

5. Analisis Data dengan Teori Respon Butir

Analisis Berdasarkan Teori Respon Butir adalah berupa analisis setiap butir soal dengan model satu parameter (tingkat kesukaran), model tiga parameter (tingkat kesukaran, daya beda, dan peluang menebak/*guessing*), dan model tiga parameter (tingkat kesukaran, daya beda, dan peluang menebak). Analisis ini juga menggunakan bantuan program (*soft ware*) komputer yakni BILOG MG.

Adapun tahapannya adalah :

- d. Memasukkan data dan menyimpannya dalam format .txt.
- e. Membuat script ke dalam program BILOG MG.
- f. Untuk melihat hasilnya maka *run* program BILOG MG tersebut.

Untuk hasil yang akan diperoleh adalah :

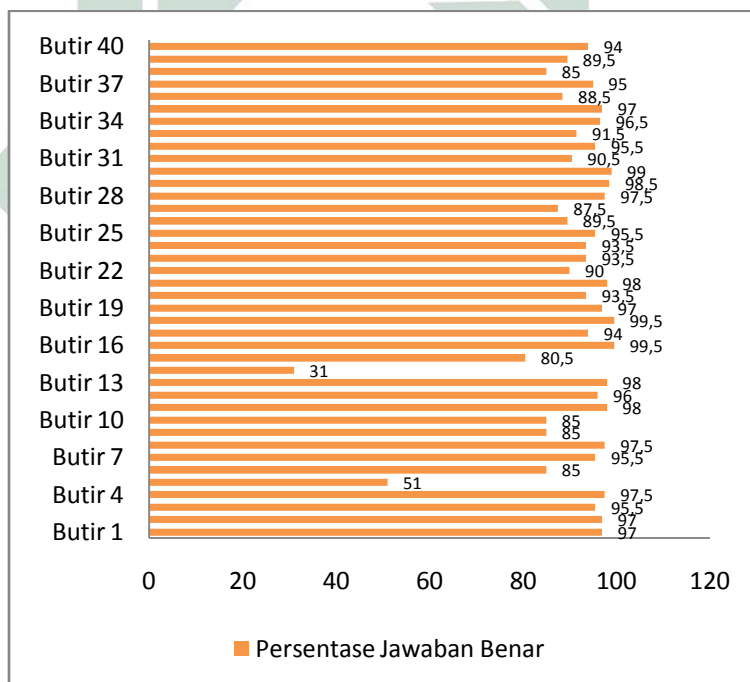
- a. Estimasi 3 PL : Dalam model logistik 3 parameter, estimasi parameter butir dinyatakan dalam bentuk daya beda (a_i), tingkat kesukaran butir (b_i), dan peluang menebak/ *guessing* (c_i).
- b. Fungsi informasi tes : Dalam mencari fungsi informasi tes tergantung pada kemiringan (*slope*), fungsi respon butir dan variansi bersyarat pada masing-masing tingkat kemampuan (θ). Semakin besar nilai *slope* dan semakin kecil nilai variansi, maka semakin banyak informasi yang dihasilkan, dan akan memperkecil kesalahan pengukuran.

BAB IV

HASIL PENELITIAN

A. Deskripsi Data

Data dalam penelitian ini berupa jawaban Ujian Sekolah atau Madrasah (US/M) mata pelajaran matematika tahun ajaran 2016-2017 yang diperoleh dengan cara mengujikan kembali soal tersebut kepada siswa kelas VII SMPN 3 Surabaya dan SMP Islam Al-Amal Surabaya dengan total 200 siswa. Grafik berikut menggambarkan persentase komposisi soal dilihat dari banyaknya siswa yang menjawab benar :



Grafik 4. 1
Persentase Jawaban Benar Setiap Butir

Bagan diatas memberikan informasi bahwa dari 40 butir soal hanya 2 butir soal yang berada dalam katagori soal yang sulit yaitu butir soal nomor 5 yang hanya 51% dari total 200 siswa dapat menjawab dengan benar dan butir soal nomor 14 yang hanya 31% dari total 200 siswa dapat menjawab dengan benar. Selebihnya yaitu 38 butir soal masuk kedalam katagori soal yang mudah. Hal tersebut dikarenakan persentase jawaban benar butir soal berada dalam interval 80% sampai dengan 99%, yang artinya hampir seluruh siswa dari total 200 siswa dapat menjawab benar soal-soal tersebut.

B. Analisis Data

1) Reliabilitas Soal

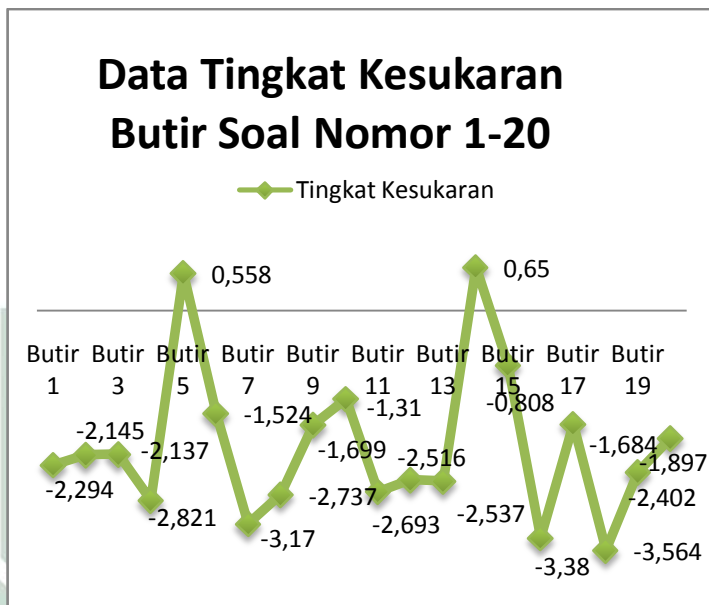
Analisis untuk tingkat reliabilitas soal Ujian Sekolah atau Madrasah (US/M) mata pelajaran matematika tahun ajaran 2016-2017 di Kota Surabaya dapat dilihat dari *output* program BILOG MG pada phase 3 yaitu 0,8213.

2) Analisis Teori Respon Butir

Data yang diperoleh kemudian dianalisis dengan menggunakan bantuan program BILOG MG untuk mendapatkan analisis soal berdasarkan teori respon butir. Teori respon butir memuat dua parameter, yaitu parameter butir dan parameter peserta. Parameter butir dinyatakan melalui model logistik yang digunakan. Dalam penelitian ini model logistik yang digunakan yaitu model logistik 3 parameter. Dalam model logistik 3 parameter, estimasi parameter butir dinyatakan dalam bentuk daya beda (a_i), tingkat kesukaran butir (b_i), dan peluang menebak/ *guessing* (c_i). Hasil parameter tersebut dapat dilihat pada *output* program BILOG fase 2.

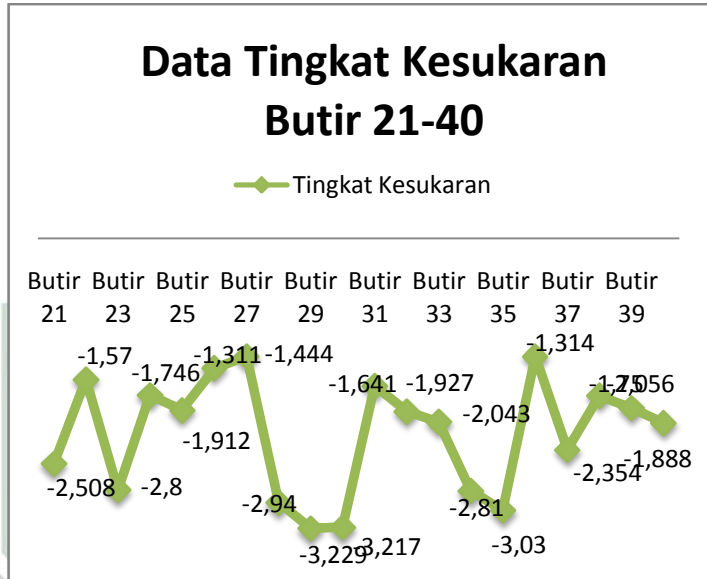
a) Tingkat Kesukaran Butir

Tingkat kesukaran butir dari *output* BILOG MG dapat dilihat pada nilai *threshold* di *output* fase 2. Berdasarkan hasil analisis data diperoleh informasi tingkat kesukaran butir sebagai berikut :



Grafik 4. 2
Analisis Tingkat Kesukaran Butir 1 – 20 dengan
Menggunakan Program BILOG MG

Grafik di atas menunjukkan tingkat kesukaran butir soal nomor 1 sampai dengan butir soal nomor 20. Butir yang memiliki tingkat kesukaran baik adalah yang mempunyai nilai tingkat kesukaran di antara -2 sampai dengan 2. Artinya dari 20 butir soal tersebut hanya 8 soal yang masuk kedalam katagori butir dengan tingkat kesukaran baik yaitu butir soal nomor 5, butir soal nomor 6, butir soal nomor 9, butir soal nomor 10, butir soal nomor 14, butir soal nomor 15, butir soal nomor 17, dan butir soal nomor 20. Selebihnya, yaitu 12 butir soal masuk kedalam katagori butir dengan tingkat kesukaran yang kurang baik.

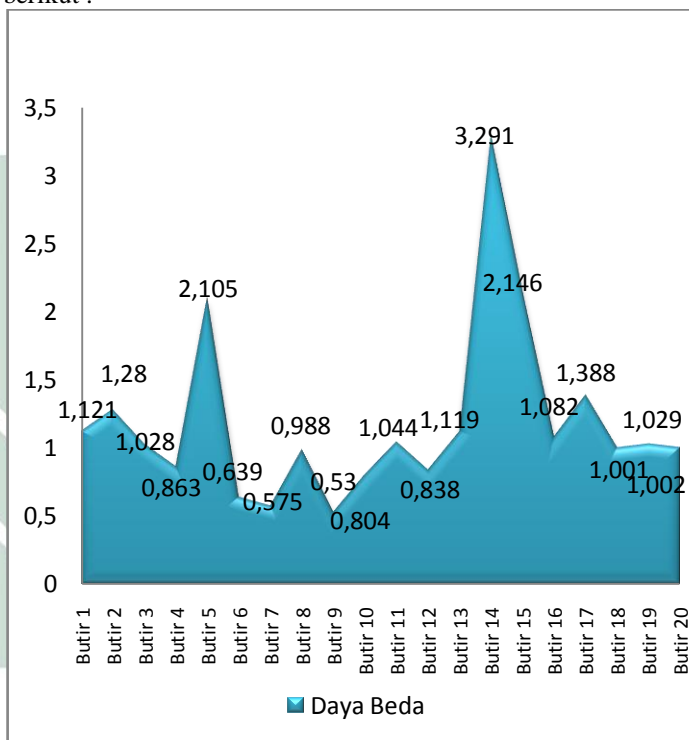


Grafik 4.3
Analisis Tingkat Kesukaran Butir 21-40 dengan
Menggunakan Program BILOG MG

Grafik di atas menunjukkan tingkat kesukaran butir soal nomor 21 sampai dengan butir soal nomor 40. Butir yang memiliki tingkat kesukaran baik adalah yang mempunyai nilai tingkat kesukaran di antara -2 sampai dengan 2. Artinya dari 20 butir soal tersebut terdapat 10 butir soal yang masuk kedalam katagori butir dengan tingkat kesukaran baik yaitu butir soal nomor 22, butir soal nomor 24, butir soal nomor 25, butir soal nomor 26, butir soal nomor 27, butir soal nomor 31, butir soal nomor 32, butir soal nomor 36, butir soal nomor 38 dan butir soal nomor 40. Selebihnya, yaitu 10 butir soal masuk kedalam katagori butir dengan tingkat kesukaran yang kurang baik.

b) Daya Bada Butir

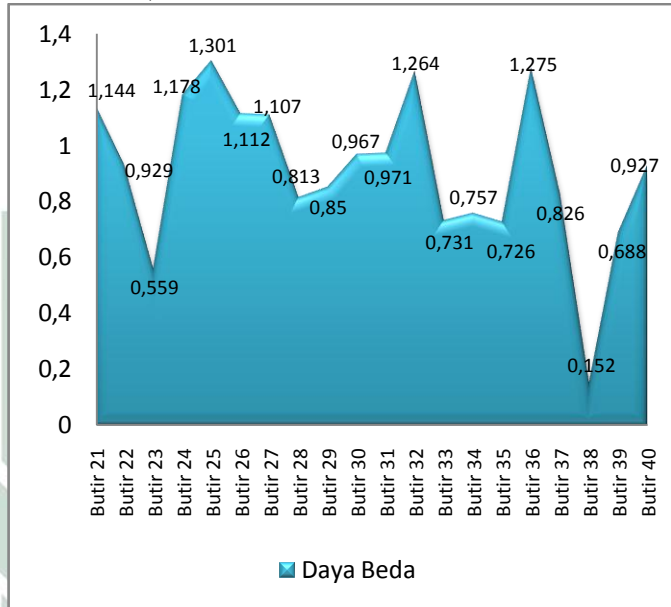
Daya bada butir dari *output* BILOG MG dapat dilihat pada nilai slope yang terdapat di *output* fase 2. Berdasarkan hasil analisis data diperoleh informasi daya bada butir sebagai berikut :



Grafik 4. 4
Analisis Daya Bada Butir 1- 20 dengan Menggunakan Program BILOG MG

Grafik di atas menunjukkan analisis daya bada butir soal nomor 1 sampai dengan butir soal nomor 20. Butir yang memiliki daya bada yang baik adalah yang mempunyai nilai daya bada di antara 0 sampai dengan 2. Artinya dari 20 butir soal tersebut terdapat 17 butir soal yang masuk kedalam

katagori butir dengan daya beda yang baik, dan hanya 3 butir soal yang memiliki daya beda yang kurang baik yaitu butir soal nomor 5, butir soal nomor 14 dan butir soal nomor 15.

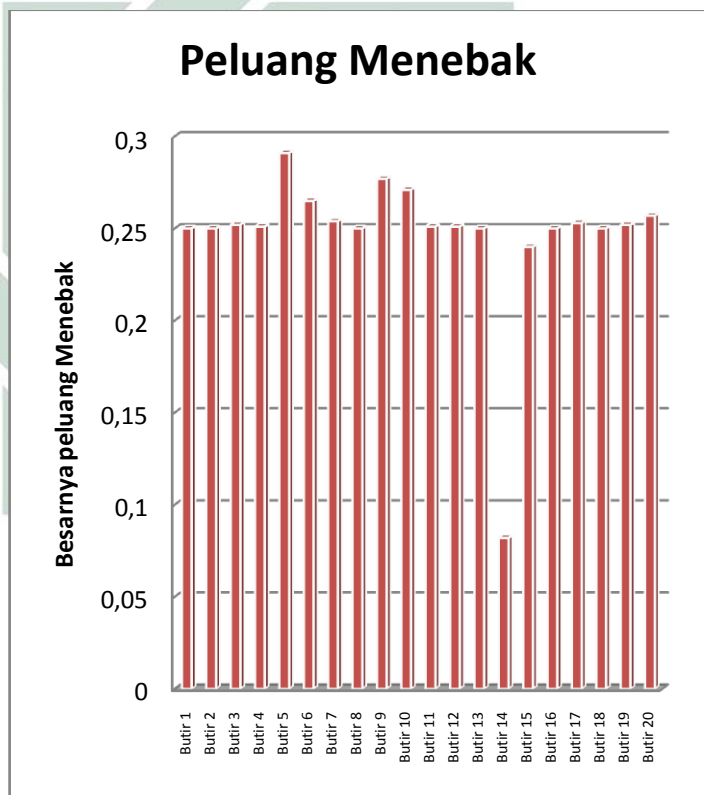


Grafik 4. 5
Analisis Daya Beda Butir 21- 40 dengan Menggunakan
Program BILOG MG

Grafik di atas menunjukkan analisis daya beda butir soal nomor 21 sampai dengan butir soal nomor 40. Butir yang memiliki daya beda yang baik adalah yang mempunyai nilai daya beda di antara 0 sampai dengan 2. Artinya semua butir soal dari nomor 21 sampai dengan butir soal nomor 40 masuk kedalam soal dengan daya beda yang baik, dimana kisaran nilai daya beda dari ke 20 soal tersebut berada dalam interval 0 sampai dengan 2. Adapun secara keseluruhan analisis daya beda butir dapat dilihat pada lampiran.

c) Peluang Menjawab Benar (Peluang Menebak / *Guessing*)

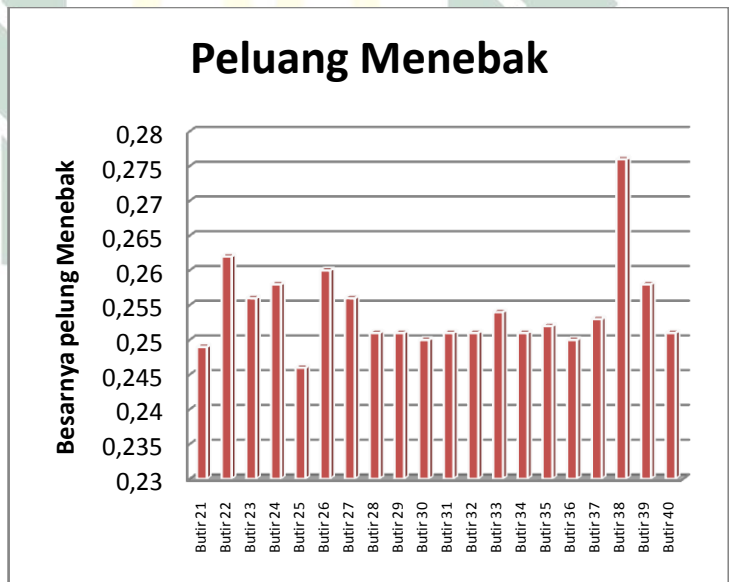
Parameter peluang menebak / *guessing* dari *output* BILOG MG dapat dilihat pada nilai *asymptote* yang terdapat di *output fase 2*. Berdasarkan hasil analisis data diperoleh informasi peluang menebak adalah sebagai berikut:



Grafik 4. 6

Analisis Peluang Menebak / Guessing Butir Soal Nomor 1-20 dengan Menggunakan Program BILOG MG

Grafik di atas menunjukkan besarnya nilai peluang menebak atau *guessing* butir soal nomor 1 sampai dengan butir soal nomor 20. Butir yang memiliki *guessing* yang baik adalah yang mempunyai nilai *guessing* tidak lebih dari 0,250. Artinya dari 20 butir soal tersebut terdapat 15 butir soal yang memiliki nilai *guessing* dengan kategori baik, yaitu butir soal nomor 1, butir soal nomor 2, butir soal nomor 3, butir soal nomor 4, butir soal nomor 7, butir soal nomor 8, butir soal nomor 11, butir soal nomor 12, butir soal nomor 13, butir soal nomor 14, butir soal nomor 14, butir soal nomor 15, butir soal nomor 16, butir soal nomor 17, butir soal nomor 18 dan butir soal nomor 19. Selebihnya, yaitu hanya 5 butir soal masuk kedalam kategori butir dengan *guessing* yang kurang baik



Grafik 4. 7

Analisis Peluang Menebak / Guessing Butir Soal Nomor 1-20 dengan Menggunakan Program BILOG MG

Grafik di atas menunjukkan besarnya nilai peluang menebak atau *guessing* butir soal nomor 21 sampai dengan butir soal nomor 40. Butir yang memiliki *guessing* yang baik adalah yang mempunyai nilai *guessing* tidak lebih dari 0,250. Artinya dari 20 butir soal tersebut terdapat 13 butir soal yang memiliki nilai *guessing* dengan katagori baik, yaitu butir soal nomor 21, butir soal nomor 25, butir soal nomor 28, butir soal nomor 29, butir soal nomor 30, butir soal nomor 31, butir soal nomor 32, butir soal nomor 33, butir soal nomor 34, butir soal nomor 35, butir soal nomor 36, butir soal nomor 37 dan butir soal nomor 40. Sebaliknya, yaitu hanya 7 butir soal masuk kedalam katagori butir dengan *guessing* yang kurang baik. Secara keseluruhan nilai *guessing* butir soal US/M mata pelajaran matematika tahun ajaran 2016-2017 di Kota Surabaya dapat dilihat dalam lampiran.

C. Pembahasan

1) Reliabilitas

Reliabel artinya dapat dipercaya atau dapat diandalkan. Reliabilitas menunjuk pada tingkat keterandalan sesuatu. Reliabilitas mengandung pengertian bahwa suatu tes dapat dipercaya untuk digunakan sebagai pengukur sesuatu. Suatu tes dikatakan reliabel jikates tersebut dalam mengukur sesuatupada waktu yang berlainan,maka akan menunjukkan hasil yang sama.¹

Ada beberapa kriteria dalam menentukan reliabilitas suatu soal:²

- a) Jika tes digunakan untuk sesuatu yang penting yang cenderung dapat mempengaruhi individu secara signifikan, maka tingkat reliabilitas yang diperlukan adalah 0,90 bahkan sampai mencapai 0,95

¹http://file.upi.edu/Direktori/FPBS/JUR._PEND._BAHASA_PERANCIS/19691223199302-2-TRI_INDRI_HARDINI/Uji_Validitas_dan_Reliabilitas.pdf, diakses pada tanggal 3 Januari 2018

² Cecil R Reynold, Ronal B Livingston, Victor Willson, *Measurement and Assessment in Education*, (London : Pearson Education, 2006), hal 108

- b) Jika tes digunakan untuk mengukur prestasi belajar baik secara individu maupun kelompok, maka tingkat reliabilitas yang diperlukan adalah 0,80 keatas.
- c) Jika tes digunakan untuk mengukur prestasi belajar tingkat kelas , maka reliabilitas yang diperlukan minimal 0,70.

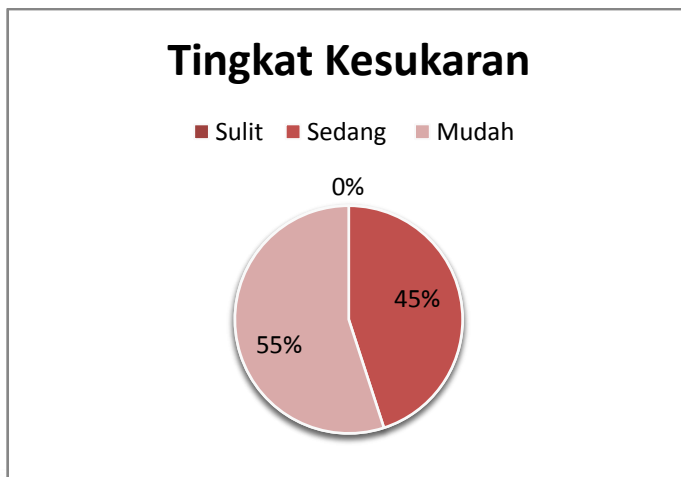
Penelitian ini masuk kedalam katagori yang kedua yaitu tes digunakan untuk mengukur prestasi belajar baik secara individu maupun kelompok, maka soal US/M ini memerlukan tingkat reliabilitas minimal 0,80, dan reliabilitas soal US/M ini adalah 0,82. Artinya soal US/M mata pelajaran matematika tahun ajaran 2016-2017 ini reliabel. Soal yang reliabel adalah soal yang baik. Soal tersebut akan tetap baik kualitasnya walaupun diujikan ke siapapun dan dalam kondisi seperti apapun.

2) Tingkat Kesukaran Butir

Tingkat kesukaran butir dinilai dalam skala $-\infty \leq b \leq \infty$ pada teori respon butir. Butir yang dinyatakan baik adalah butir yang memiliki tingkat kesukaran (b_i) berkisar antara $-2 \leq b \leq 2$.³ Butir yang memiliki tingkat kesukaran mendekati +2 menunjukkan bahwa butir tersebut memiliki tingkat kesukaran yang tinggi. Begitupun dengan sebaliknya, butir yang memiliki tingkat kesukaran mendekati -2 menunjukkan bahwa butir tersebut memiliki tingkat kesukaran yang rendah atau soal tersebut termasuk kedalam katagori mudah.

Adapun Persentase hasil analisis dari tingkat kesukaran soal US/M mata pelajaran matematika tahun ajaran 2016-2017 berdasarkan teori respon butir adalah sebagai berikut :

³ R.K. Hambleton & Swaminathan, Hariharan & Rogers, *Fundamental of Item Response Theory*, (Newbury Park : Sage Publication Inc, 1991), hal 13



Grafik 4. 8
Persentase Analisis Tingkat Kesukaran Butir Soal
US/M Mata Pelajaran Matematika Tahun Ajaran
2016-2017 di Kota Surabaya

Grafik di atas menunjukkan Persentase tingkat kesukaran butir soal Ujian Sekolah atau Madrasah (US/M) mata pelajaran matematika tahun ajaran 2016-2017. Sebanyak 55% dari keseluruhan soal memiliki tingkat kesukaran yang mudah, dan 45% dari keseluruhan butir soal memiliki tingkat kesukaran sedang. Artinya tingkat kesukaran butir soal US/M mata pelajaran matematika tahun ajaran 2016-2017 termasuk kedalam katagori kurang baik, karena soal yang baik adalah soal dengan tingkat kesukaran yang tidak mudah ataupun yang sulit. Dengan kata lain soal yang baik adalah soal yang memiliki tingkat kesukaran yang sedang.

Untuk memperkuat analisis tersebut kita dapat melihat nilai tingkat kesukaran soal US/M secara keseluruhan. Secara keseluruhan nilai tingkat kesukaran soal US/M mata pelajaran matematika tahun ajaran 2016-2017 dapat dilihat dari rata-rata nilai tingkat kesukaran dari ke 40 butir soal US/M. Adapun nilai rata-rata tingkat kesukaran soal US/M adalah -2,089. Nilai tersebut masuk kedalam katagori kurang baik, sehingga dapat

disimpulkan bahwa tingkat kesukaran soal US/M mata pelajaran matematika tahun ajaran 2016-2017 adalah kurang baik.

Faktor yang memungkinkan tingkat kesukaran soal US/M kurang baik adalah konten soal yang kurang bagus. Terdapat beberapa soal yang terlalu sulit dan terlalu mudah. Seperti soal nomor 14 ada sebanyak 137 siswa salah dalam menjawabnya, dan menjadikan soal tersebut masuk kedalam katagori sulit. Faktor lain yang memungkinkan menjadikan soal tersebut masuk dalam katagori sulit adalah kemampuan dari siswa yang menjadi sampel dalam penelitian ini. SMP Islam al-Amal Surabaya merupakan sekolah dengan *great* yang kurang baik dengan kata lain sekolah berakreditasi rendah, sehingga memungkinkan siswa tidak dapat menjawab dengan benar dan menjadikan soal tersebut masuk dalam katagori sulit. Tingkat kesukaran soal biasanya dikaitkan dengan tujuan tes.⁴ Jadi, jika tingkat kesukaran suatu soal kurang baik maka soal tersebut dikatakan tidak mencapai tujuannya.

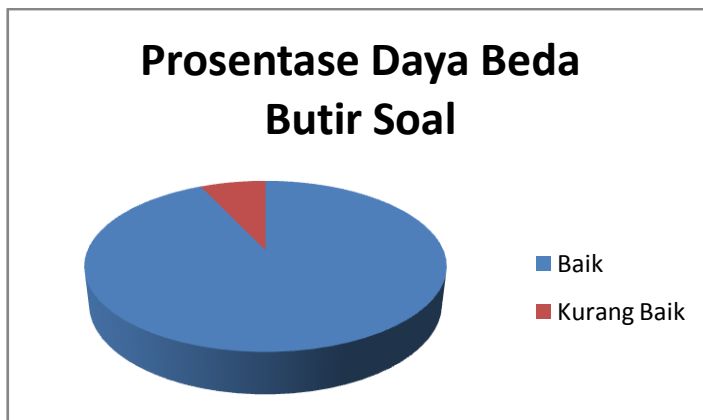
3) Daya Beda Butir

Daya beda butir dalam teori respon butir secara teoritik parameternya didefinisikan pada skala $(-\infty, +\infty)$. Nilai yang negatif menunjukkan bahwa butir tersebut tidak baik. Nilai yang positif menunjukkan bahwa tes tersebut berfungsi dengan baik dalam mengukur kemampuan yang hendak diukur. Semakin kecil koefisien daya beda menunjukkan soal tersebut berdaya beda kurang baik, sedangkan semakin besar koefisien daya beda menunjukkan soal berdaya beda baik. Nilai daya beda yang baik berkisar antara 0 sampai dengan 2.⁵

Adapun Persentase hasil analisis dari daya beda butir soal US/M mata pelajaran matematika tahun ajaran 2016-2017 berdasarkan teori respon butir adalah sebagai berikut :

⁴ Kusaeri, Acuan dan Teknik Penilaian Proses & Hasil Belajar dalam Kurikulum 2013, (Yogtakarta : Ar – Ruzz Media, 2014), hal 107

⁵ Ibid, R.K. Hambelton & Swaminathan, Hariharan & Rogers, hal 15



Grafik 4. 9
Persentase Analisis Daya Beda Butir Soal US/M Mata Pelajaran Matematika Tahun Ajaran 2016-2017 di Kota Surabaya

Grafik diatas menunjukkan Persentase daya beda butir soal Ujian Sekolah atau Madrasah (US/M) mata pelajaran matematika tahun ajaran 2016-2017 di Kota Surabaya. Sebesar 97% dari keseluruhan butir memiliki daya beda yang baik, dan hanya 7% dari keseluruhan butir soal memiliki daya beda yang kurang baik. Artinya daya beda butir soal US/M mata pelajaran matematika tahun ajaran 2016-2017 termasuk kedalam katagori baik.

Untuk memperkuat analisis tersebut kita dapat melihat nilai tingkat daya beda soal US/M secara keseluruhan. Secara keseluruhan daya beda soal US/M mata pelajaran matematika tahun ajaran 2016-2017 dapat dilihat dari rata-rata nilai daya beda dari ke 40 butir soal US/M. Adapun nilai rata-rata daya beda soal US/M adalah 1,063. Nilai tersebut masuk kedalam katagori baik, sehingga dapat disimpulkan bahwa daya beda soal US/M mata pelajaran matematika tahun ajaran 2016-2017 adalah baik. Faktor yang mempengaruhi daya beda soal adalah pengecohnya. Dalam soal ini pengecohnya dapat berfungsi

dengan baik. Soal yang daya bedanya baik memiliki manfaat dapat membedakan kemampuan siswa yang pandai dan kurang.⁶

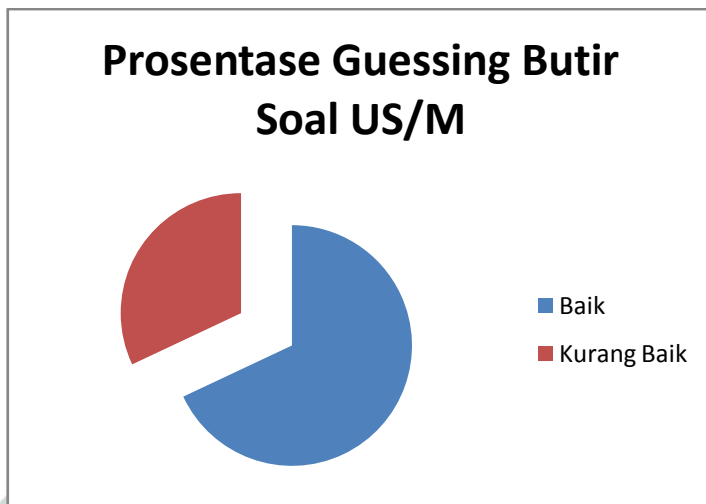
4) Peluang Menebak / *Guessing*

Peluang menebak merupakan ukuran peluang peserta untuk menebak dengan benar. Suatu soal diharapkan mempunyai peluang menebak yang kecil bahkan mendekati nol. Sehingga soal yang baik adalah soal yang memiliki peluang menebak mendekati 0. Nilai peluang menebak (c_i) adalah antara 0 sampai dengan 1. Suatu butir dikatakan baik jika nilai c_i nya tidak lebih dari $1/k$, dengan k adalah banyaknya pilihan jawaban.⁷ Dalam penelitian ini pilihan jawabannya adalah 4, maka nilai dari c_i untuk setiap butir seharusnya tidak lebih dari $1/4$ atau tidak lebih dari 0.25.

Adapun hasil analisis dari peluang menebak atau *guessing* dalam soal US/M mata pelajaran matematika tahun ajaran 2016-2017 adalah sebagai berikut :

⁶ Kusaeri, Acuan dan Teknik Penilaian Proses & Hasil Belajar dalam Kurikulum 2013, (Yogtakarta : Ar – Ruzz Media, 2014), hal 107

⁷http://www.academia.edu/29987008/ANALISIS_ITEM_DENGAN_PENDEKATAN_IR_T_DENGAN_BANTUAN_APLIKASI_PROGRAM_BILOG-MGhal , diakses pada tanggal 30 Desember 2017



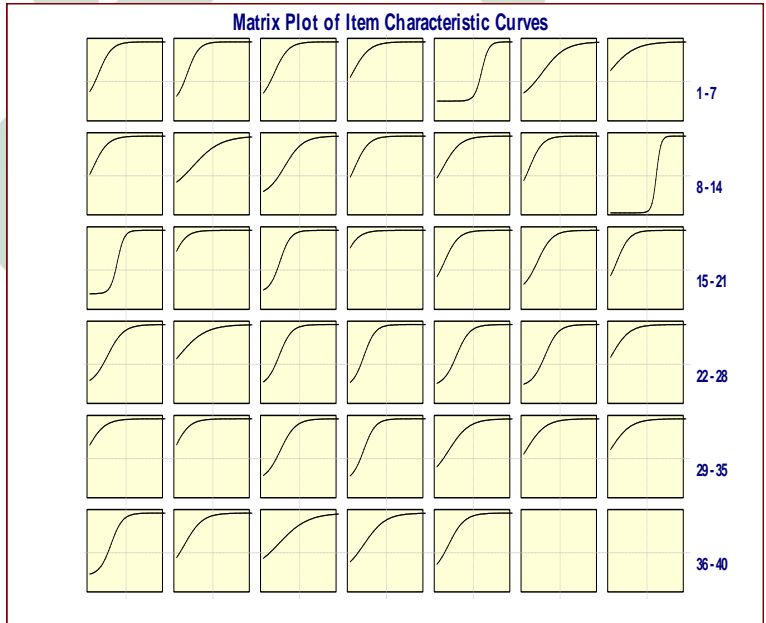
Grafik 4. 10
Persentase Analisis Guessing Butir Soal US/M Mata
Pelajaran Matematika Tahun Ajaran 2016-2017 di Kota
Surabaya

Grafik diatas merupakan Persentase nilai peluang menebak atau *guessing* dari butir soal Ujian Sekolah atau Madrasah mata pelajaran matematika tahun ajaran 206-2017 di Kota Surabaya. 68% dari keseluruhan butir soal memiliki nilai peluang menebak atau *guessing* yang baik dengan indeks berkisar $c_i \leq 0,25$. Sedangkan 32% dari keseluruhan butir soal memiliki nilai peluang menebak atau *guessing* yang kurang baik dengan indeks kesukaran diatas 0,25. Secara klesuruhan nilai peluang menebak atau *guessing* dalam soal US/M ini dapat dilihat dari rata-rata nilai *guessing* dari ke 40 butir soal. Adapun nilai rata-ratanya adalah 0,251 dan masuk dalam katagori sedang. Artinya pilihan jawaban berfungsi dengan baik. Adapun faktor yang mempengaruhi nilai peluang menebak atau *guessing* dari suatu soal adalah daya beda soal tersebut. Semakin bagus daya bedanya maka nilai peluang menebak atau *guessing* soal tersebut akan

semakin kecil dengan kata lain tidak banyak siswa yang menebak dalam menjawab soal tersebut.

5) Kurva Karakteristik Butir

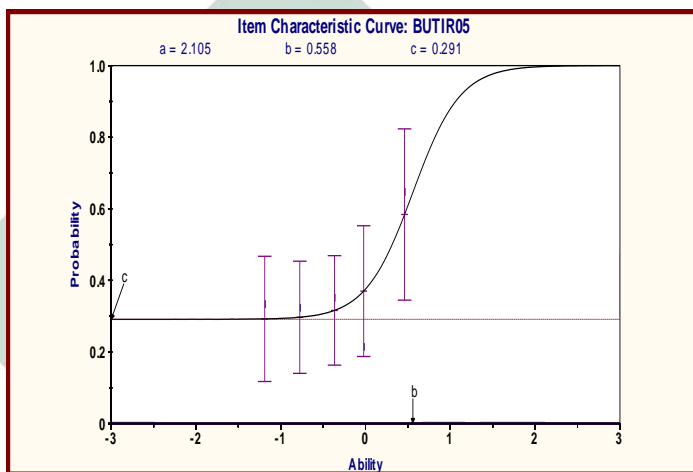
Pada dasarnya untuk analisis berdasarkan teori respon butir model logistik 3 parameter, butir soal yang baik ataupun yang kurang baik dapat dilihat dari bentuk kurva karakteristik dan nilai ketiga parameter yang mendukung yaitu, tingkat kesukaran butir, daya beda butir dan peluang menebak. Berikut merupakan kurva karakteristik butir soal dengan model logistik 3 parameter.



Gambar 4.1
Kurva Karakteristik Butir Soal US/M Nomor 1 sampai
dengan Butir Soal Nomor 40

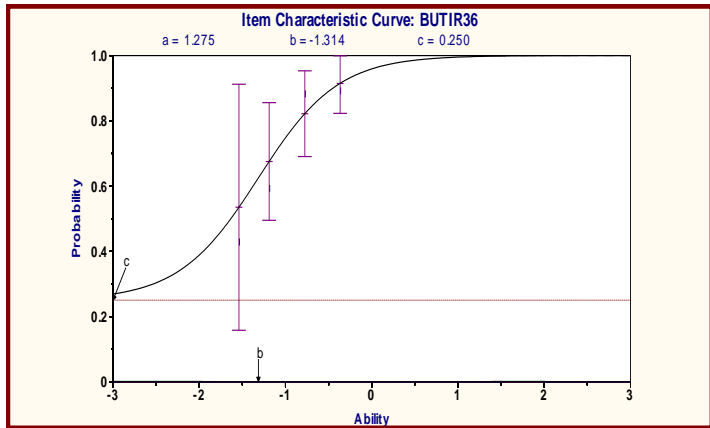
Kurva karakteristik 40 butir soal pada gambardiatas, merupakan bentuk kurva karakteristik butir soal berdasarkan model logistik 3 parameter. Butir soal yang ideal adalah butir soal dengan kurva karakteristik berbentuk huruf S yang cukup landai.⁸

Dalam mengamati kurva karakteristik butir berikut ini dapat dilihat contoh estimasi kurva karakteristik dua butir soal yang berbeda seperti gambar dibawah ini :



Gambar 4. 2
Kurva Karakteristik Butir Soal US/M Nomor 5
Berdasarkan Teori Respon Butir

⁸ Retnawati Heri, *Validitas, Reliabilitas & Karakteristik Butir*, (Yogyakarta : Parama Publishing, 2016), hal 137



Gambar 4. 3
Kurva Karakteristik Butir Soal US/M Nomor 36
Berdasarkan Teori Respon Butir

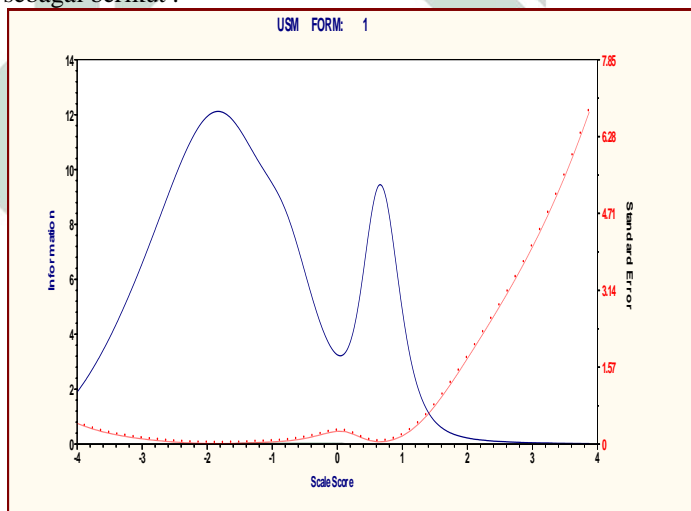
Kurva karakteristik butir soal pada gambar diatas menggambarkan peluang menjawab butir soal berdasarkan daya beda, tingkat kesukaran dan peluang menebak butir soal. Dari kedua gambar diatas, butir soal nomor 36 yang paling ideal karena berbentuk huruf S dan cukup landai. Kurva seperti ini menunjukkan butir soal yang baik. Dimana semakin besar kemampuan peserta maka peluang menjawab soal juga semakin besar.

Sedangkan pada gambar pertama dimana gambar tersebut merupakan gambar kurva karakteristik soal nomor 5 adalah termasuk katagori kurang baik, karena soal ini mempunyai peluang menebak paling tinggi yaitu 0,291. Artinya peserta dengan kemampuan rendahpun mempunyai peluang menjawab benar soal ini. Karakteristik daya bedanya juga kurang baik.

6) Fungsi Informasi Tes

Nilai- nilai indeks parameter butir merupakan hasil estimasi. Oleh karena itu kebenarannya bersifat probabilitas dan tidak terlepas dengan kesalahan pengukuran. Dalam teori respon butir, kesalahan penaksiran standart (*Standart Error of Measurement, SEM*) berkaitan erat dengan fungsi informasi. Fungsi informasi dengan SEM mempunyai hubungan yang berbanding terbalik. Semakin besar fungsi informasi, maka SEM semakin kecil dan sebaliknya semakin kecil fungsi informasi maka akan semakin besar nilai SEM.⁹

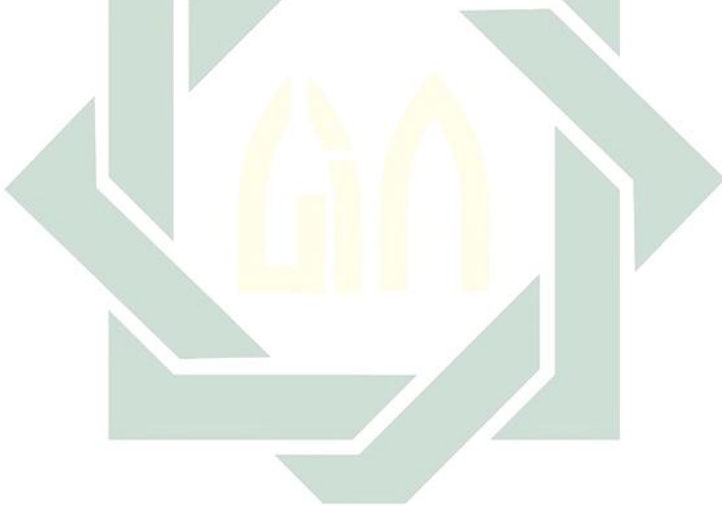
Adapun grafik fungsi informasi tes dari keseluruhan butir soal Ujian Sekolah atau Madrasah (US/M) mata pelajaran matematika tahun ajaran 2016-2017 di Kota Surabaya adalah sebagai berikut :



Gambar 4. 4
Kurva Fungsi Informasi Butir Soal US/M
Berdasarkan Teori Respon Butir

⁹ Ibid, hal 138

Berdasarkan grafik di atas maka nilai fungsi informasi tes maksimal tercapai pada kemampuan siswa sebesar -2. Sedangkan garis merah menunjukkan kesalahan pengukuran. Karena semakin besar nilai fungsi informasi maka akan semakin kecil kesalahan pengukurannya, maka kesalahan pengukuran dalam soal US/M ini adalah besar. Sehingga soal US/M mata pelajaran matematika tahun ajaran 2016-2017 di Kota Surabaya jika dianalisis dengan menggunakan teori respon butir tidak dapat mengukur kemampuan siswa sesuai yang diharapkan. Hal tersebut terjadi karena ketidakstabilan ketiga parameter di atas yaitu, tingkat kesukaran yang kurang baik, daya bedanya yang baik dan nilai peluang menebak atau *guessing* yang juga baik.



BAB V

PENUTUP

A. Simpulan

Kualitas soal Ujian Sekolah atau Madrasah (US/M) mata pelajaran matematika tahun ajaran 2016-2017 di Kota Surabaya berdasarkan teori respon butir adalah baik. Hal tersebut dapat dilihat dari beberapa indikator berikut: (1) nilai reliabilitasnya adalah 0,8213 (reliabel), (2) nilai tingkat kesukaran adalah -2,089 (kurang baik), (3) nilai daya beda adalah 1,063 (baik), dan (4) nilai peluang menebak atau *guessing* adalah 0,251 (baik).

B. Saran

1. Untuk penelitian selanjutnya diharapkan meneliti pada tingkat SMP atau MTS.
2. Untuk penelitian selanjutnya diharapkan meneliti pada tingkat SMA atau MA.
3. Untuk penelitian selanjutnya diharapkan bisa memperoleh data secara langsung.

DAFTAR PUSTAKA

- Adek, Purnama.2014. Skripsi: *Analisis Soal Ujian Nasional Matematika SMP/MTS yang Didasarkan pada Tingkat Pemahaman Konsep Penalaran dan Pemecahan Masalah*.Jakarta : Universitas Pendidikan Indonesia
- Allen, M. J & W. M Yen.1979. *Introduction to measurement theory*. California: Brookd/Cole Publishing Company, 1979.
- Arifin, Zaenal.2009. *Metodologi Penelitian Pendidika Filosofi, Teori & Aplikasinya*. Surabaya: Lentera Cendikia
- Azwar, S.2015. *Dasar-dasar Psikometri*.Yogyakarta: Pustaka Pelajar.
- Baker, Frank B. 2001.*The Basic OfItem Response Theori. America :ERIC Clearinghouse On Assesment and Evaluation*.
- Chabib Thoha, M. 1996. *Teknik Evaluasi Pendidikan*. Jakarta: PT. RajaGrafindo Persada.
- Departemen Pendidikan Nasional. 2002. *Kamus Besar Bahasa Indonesia*. Jakarta: Balai Pustaka.
- Dwi Kurniawan, Devi. 2015. *Anilisis Kualitas Soal Ujian Akhir Semester Matematika Berdasarkan Teori Respon Butir*, Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS.
- Hambelton, R.K. & Swaminathan, Hariharan & Rogers.1991. *Fundamental of Item Response Theory*. Newbury Park : Sage Publication Inc
- Hanifah , Nanang Dan Cucu Suhana. 2009. *Konsep Strategi Pembelajaran*. Bandung : Rafika Aditama.
- Herwelis. 2013. Thesis : *Analisis Kualitas Soal Ujian Akhir Sekolah (Uas) Mata Pelajaran Matematika Tingkat Smk Kabupaten Takalar Tahun Pelajaran 2012/2013*. Makasar : Program Pascasarjana Universitas Negeri Makasar.
- Hidayanto, N. 2011.*Karakteristik Butir Soal Ulangan Kenaikan Kelas Persiapan Bank Soal Bahasa Inggris*. Yogyakarta : Jurnal Penelitian dan Evaluasi Pendidikan, No. 1.
- Kemendikbud. 2016. *Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor23 Tahun 2016. Tentang Standar Penilaian Pendidikan Bab II Pasal 2*. Jakarta : Kemendikbud.
- Kemendikbud. 2017. *Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 102 Tahun 2013 Tentang Penyelenggaraan Ujian Sekolah/Madrasah Pada Sekolah*

Dasar/Madrasah Ibtidaiyah, Sekolah Dasar Luar Biasa, Dan Program Paket A/Ula. Jakarta : Kemendikbud.

- Khoiriyah, Nur. 2010. Thesis : *Kualitas Tes Uasbn Mata Pelajaran Matematika Sd/Mi Di Kabupaten Jombang Tahun 2009.* Surabaya: UIN Sunan Ampel Surabaya
- Kusaeri, K, (2014). *Acuan dan Teknik Penilaian Proses dan Hasil Belajar dalam Kurikulum 2013.* Yogyakarta : Ar-Ruzz Media.
- Kurniawan, Devi Dwi. 2015. *Anilisis Kualitas Soal Ujian Akhir Semester Matematika Berdasarkan Teori Respon Butir.* Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS.
- Malau, R. A& D. J. Ratnaningsih. 2002. *Penerapan Teori Respon Butir dalam Menentukan Karakteristik Butir Soal.* Jakarta : Lembaga Penelitian Universitas Terbuka Jakarta.
- Mardapi, Djemari. 2008. *Teknik penyusunan instrument tes dan nontes.* Yogyakarta: Mitra Cendekia Press.
- Mendikbud. 2017. *Peraturan Kepala Badan Penelitian Dan Pengembangan Kementerian Pendidikan Dan Kebudayaan Nomor: 007 /H/Ep/2017 Tentang Prosedur Operasional Standar Ujian Sekolah/Madrasah Pada Sekolah Dasar/Madrasah Ibtidaiyah, Sekolah Dasar Luar Biasa, Dan Penyelenggara Program Paket A/Ula Tahun Pelajaran 2016/2017.* Jakarta : Mendikbud.
- Mulyadi. 2010. *Evaluasi Pendidikan.* Malang: UIN Maliki Press.
- Nurung, Muh. 2008. Thesis: *Kualitas Tes Ujian Akhir Sekolah Berstandar Nasional (Uasbn) Ipa Sd Tahun Pelajaran 2007/2008 Di Kota Kendari.* Yogyakarta: Graduate School, State University of Yogyakarta.
- Purwanti, Ani dan Irni Wulandari. *Studi Kualitas UASBN Mata Pelajaran Matematika Wilayah Jakarta Timur Tahun Ajaran 2007/2008,* www.lpmpdki.web.id/pdf/ani%20-%20irni.pdf
Diakses pada tanggal 2 September 2017.
- Ratnaningsih, Dewi Julia. 2013. *Analisis Butir Tes Objektif Ujian Akhir Semester Mahasiswa Universitas Terbuka Berdasarkan Teori Tes Modern.* Jurnal Pendidikan Terbuka dan Jarak Jauh Volume:14. September 2013.
- R, Arrynda. 2012. Skripsi : *Analisis Kualitas Soal-Soal Try Out Ujian Nasional Mata Pelajaran Ipa Smp Di Kabupaten Banjarnegara.* Malang : Universitas Negeri Malang.

- R Reynold , Cecil, Ronal B Livingston, Victor Willson.2006. *Measurement and Assessment in Education*.London : Pearson Education.
- Retnawati Heri.2014. *Teori Respons Butir dan Penerapannya*. Yogyakarta : Parama Publishing
- Retnawati Heri.2016. *Validitas Reliabilitas & Karakteristik Butir*. Yogyakarta : Parama Publishing
- Sudijono, Anas. 2011. *Pengantar Evaluasi Pendidikan*.Jakarta : Raja Grafindo Persada.
- Sudaryono. 2012. *Dasar-Dasar Evaluasi Pembelajaran*.Yogyakarta: Graha Ilmu.
- Sudaryono.2011.*Implementasi Teori Responsi Butir (Item Response Theory) pada Penilaian Hasil Belajar Akhir di Sekolah*, Jurnal Pendidikan dan Kebudayaan Volume 17 Nomor 6 Tahun 2011.
- Surapranata, Sumarna.2005. *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes Implementasi Kurikulum 2004*.Bandung: Remaja Rosdakarya.
- Sulistiawan,C. Heri. *Kualitas Soal Ujian Siswa Sekolah Matematika Program IPA dan Kontribusinya Terhadap Hasil Ujian Nasional*.Jurnal Penelitian dan Evaluasi Pendidikan. Volume:20.Juni 2006.
- Suwarto. *Teori Tes Klasik dan Teori Tes Modern*. 2011. Jurnal Widyatama Volume:20 No.1. Sukoharjo : Universitas Veteran Bangun Nusantara.
- Syainudin, Syamsir dan Muhammad Ilyas. 2014. *Karakteristik Butir Tes Matematika pada Tes Buatan MGMP Matematika Kota Palopo Berdasarkan Teori Klasik*. Jurnal Pedagogy. Volume:1 No.1. Palopo : Universitas Cokroaminoto.
- Tim Dosen Administrasi Pendidikan UPI. 2009.*Manajemen Pendidikan*. Bandung: Alfabeta.
- https://id.wikipedia.org/wiki/Ujian_Nasional , diakses pada tanggal 9 Agustus 2017
- <http://id.wikipedia.org/wiki/Reliabilitas>, diakses pada tanggal 4 September 2017.
- <https://mepsychology.wordpress.com/2014/12/01/validitas-dan-reliabilitas/>, diakses pada tanggal 24 Juli 2017

- http://www.academia.edu/29987008/ANALISIS_ITEM_DENGAN_PENDEKATAN_IRT_DENGAN_BANTUAN_APLIKASI_PROGRAM_BILOG-MGhal , diakses pada tanggal 30 Desember 2017
- www.gurupembaharu.com/home/download/panduan-analisis-butir-oal.pdf, diakses pada tanggal 14 Juni 2017
- www.file.upi.edu/Direktori/.../TEORI_TES_KLASIK_%5BCompatibility_Mode%5D.pdf , diakses pada tanggal 2 Juni 2017
- <http://www.semestapsikometrika.com/2017/07/sekilas-tentang-teori-tes-lasik.html> , diakses pada tanggal 30 Oktober 2017
- http://file.upi.edu/Direktori/FPBS/JUR._PEND._BAHASA_PERANCIS/196912231993022-TRI_INDRI_HARDINI/Uji_Validitas_dan_Reliabilitas.pdf, diakses pada tanggal 3 Januari 2018
- toswari.staff.gunadarma.ac.id/Downloads/files/8165/Uji+Validitas+dan+Reliabilitas.pdf, diakses pada tanggal 4 September 2017

