

**PENGEMBANGAN
INSTRUMEN PENILAIAN *HIGHER ORDER THINKING*
UNTUK PELAJARAN MATEMATIKA
DENGAN MEMASUKKAN KONTEKS ISLAM**

SKRIPSI

Oleh:

MUHAMMAD IKMAL FAZIEN

NIM. D04213020



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

**PENGEMBANGAN
INSTRUMEN PENILAIAN *HIGHER ORDER THINKING*
UNTUK PELAJARAN MATEMATIKA
DENGAN MEMASUKKAN KONTEKS ISLAM**

SKRIPSI

Diajukan kepada Universitas Islam Negeri Sunan Ampel Surabaya
untuk memenuhi salah satu persyaratan dalam menyelesaikan Program
Sarjana Pendidikan (S.Pd)

Oleh:

MUHAMMAD IKMAL FAZIEN

NIM. D04213020

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

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Ikmal Faizien
NIM : D04213020
Jurusan/Program Studi : PMIPA/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, 22 Januari 2018
Yang membuat pernyataan

METERAI
TEMPEL

913B2AEF858638163

6000
ENAM RIBU RUPIAH

Muhammad Ikmal Faizien
NIM. D04213020

PENGESAHAN TIM PENGUJI SKRIPSI
 Skripsi oleh Muhammad Haniq Fauzan ini telah dipertahankan di depan
 Tim Penguji Skripsi
 tanggal, 6 Februari 2019
 di hadapan Tim Penguji Skripsi
 dan telah diterima oleh Tim Penguji Skripsi



[Signature]
 Dr. Nurul Huda, M.Pd
 NIP. 196708051983031003

Tim Penguji
 Pengaji I,
[Signature]
 Dr. Sidiq M. H.
 NIP. 197701032004123006

Pengaji II,
[Signature]
 Abdulloh Lubis, M. Si
 NIP. 198111152007121003

Pengaji III,
[Signature]
 Dr. Rizkiyati, M. Pd
 NIP. 197206071997031001

Pengaji IV,
[Signature]
 Lailatul Firdausy, M. Pd
 NIP. 199007200808141002

PERSetujuan PEMBIMBING SKRIPSI

Skripsi oleh:

Nama : Muhammad Ikmal Fauzan

Nim : D04213020

Judul : PENGEMBANGAN INSTRUMEN PENILAIAN *HIGHER ORDER THINKING* UNTUK PELAJARAN MATEMATIKA DENGAN MEMASUKKAN KONTEKS ISLAM

ini telah diperiksa dan disetujui untuk diajukan.

Pembimbing I.

Dr. Kusari, M.Pd

NIP. 197206071997011001

Sumbawa, 22 Januari 2018

Pembimbing II.

Lutfan Urah Santosa, S.Si, M.Pd

NIP. 198309262000042002



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60217 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpustakaan@uin-sby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

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

Nama : MUHAMMAD IKMAL FAZIEH
NIM : 004213020
Fakultas/Jurusan : FTK / PENDIDIKAN MATEMATIKA
E-mail address : ikmolfaizien@gmail.com

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

☒ Skripsi ☐ Tesis ☐ Disertasi ☐ Lain-lain (.....)

yang berjudul :

PENGEMBANGAN INSTRUMEN PENILAIAN HIGHER ORDER
THINKING UNTUK PELAJARAN MATEMATIKA DEHAAN MEMAJUKAN
KONTEKSI ISLAM

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengolah-media/format-lain, 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.

Penulis

(M. IKMAL FAZIEH)
nama terang dan tanda tangan

Pengembangan Instrumen Penilaian *Higher Order Thinking* Untuk Pelajaran Matematika Dengan Memasukkan Konteks Islam

Oleh:

Muhammad Ikmal Faizien

ABSTRAK

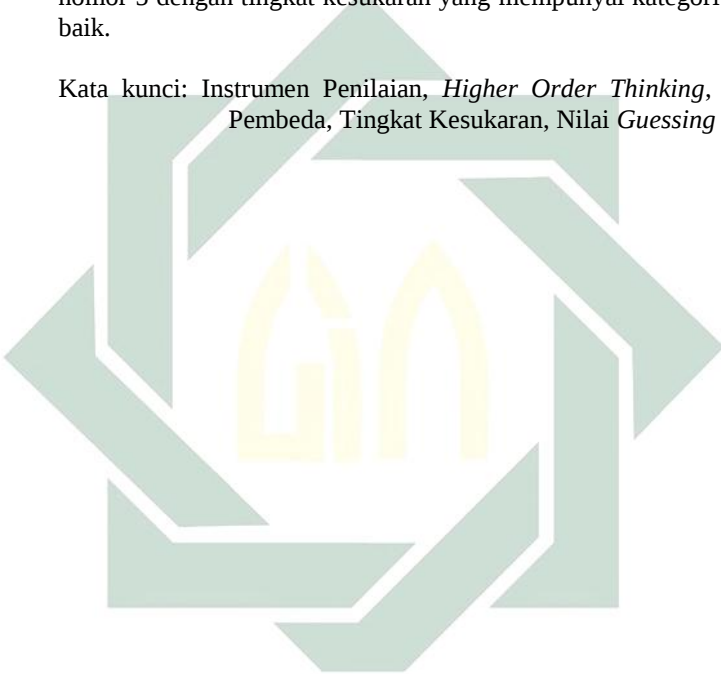
Tujuan pembelajaran matematika pada kurikulum 2013 agar siswa memiliki kemampuan dalam memecahkan masalah. Untuk melatih kemampuan pemecahan masalah, diperlukan latihan soal yang berkualitas. Soal yang dapat melatih kemampuan menganalisis, mengevaluasi dan mengkreasi siswa merupakan soal *Higher Order Thinking* (HOT). Penelitian ini bertujuan untuk mendeskripsikan proses pengembangan instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam dan untuk mengetahui karakteristik instrumen penilaian terkait daya pembeda, tingkat kesukaran dan nilai *guessing*.

Jenis penelitian ini merupakan penelitian pengembangan. Penelitian pengembangan ini mengadaptasi model pengembangan Mardapi yang terdiri dari lima langkah yaitu (1) menyusun spesifikasi instrumen soal tes, (2) menulis butir soal tes, (3) menelaah soal tes, (4) melakukan uji coba, dan (5) menganalisis butir soal tes. Uji coba instrumen dilakukan di kelas IX pada empat sekolah, yaitu MTsN Gresik, MTsN Sidoarjo, MTsN 2 Surabaya dan MTsN 3 Surabaya. Jumlah keseluruhan subjek uji coba sebanyak 165 siswa. Data yang diperoleh dalam penelitian ini berupa hasil tes instrumen. Analisis data dilakukan dengan menggunakan program BILOG-MG.

Hasil penelitian ini menunjukkan: (1) Terdapat temuan menarik pada saat proses pengembangan yakni pada langkah menelaah soal tes dan menganalisis butir soal tes. Pada langkah menelaah soal tes, awalnya jumlah instrumen penilaian yang dikembangkan sejumlah 10 butir soal. Melalui tahap telaah butir soal tes, terdapat 8 butir soal yang dikatakan valid dengan alokasi waktu selama 2×40 menit dan layak untuk di uji coba. Berikutnya pada langkah menganalisis butir soal tes. Terdapat dua temuan topik yang menarik pada langkah ini. Pertama, butir soal nomor 3 dikeluarkan dalam tahap analisis karakteristik instrumen dikarenakan korelasi biserial yang bernilai

negatif. Kedua, hasil dari estimasi perhitungan karakteristik instrumen penilaian didapatkan 3 karakteristik butir soal yang baik yaitu butir soal nomor 3,4 dan 6 yang lolos uji kecocokan model. (2) karakteristik instrumen penilaian 7 butir soal yang meliputi daya pembeda, tingkat kesukaran, dan nilai *guessing* termasuk kategori baik, hanya butir soal nomor 1 dengan daya pembeda dan butir soal nomor 5 dengan tingkat kesukaran yang mempunyai kategori tidak baik.

Kata kunci: Instrumen Penilaian, *Higher Order Thinking*, Daya Pembeda, Tingkat Kesukaran, Nilai *Guessing*



DAFTAR ISI

SAMPUL DALAM.....	i
PERSETUJUAN PEMBIMBING.....	ii
PENGESAHAN TIM PENGUJI	iii
PERNYATAAN KEASLIHAN TULISAN	iv
PERSEMBAHAN.....	v
ABSTRAK	vii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR.....	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN.....	1
A. Latar Belakang.....	1
B. Rumusan Masalah Penelitian.....	4
C. Tujuan Penelitian dan Pengembangan.....	4
D. Spesifikasi Produk yang Dikembangkan	5
E. Manfaat Pengembangan.....	5
F. Asumsi dan Keterbatasan Penelitian.....	5
G. Definisi Operasional	6
BAB II KAJIAN PUSTAKA	7
A. <i>Higher Order Thinking</i>	7
1. Pengertian <i>Higher Order Thinking</i>	7
2. Soal <i>Higher Order Thinking</i>	8

3. Karakteristik <i>Higher Order Thinking</i>	13
B. Karakteristik Soal <i>Higher Order Thinking</i>	14
C. Matematika dalam Konteks Islam	18
BAB III METODE PENELITIAN	27
A. Metode Penelitian dan Pengembangan	27
B. Prosedur Penelitian dan Pengembangan	27
1. Menyusun Spesifikasi Instrumen Soal Tes	27
2. Menulis Butir Soal Tes	28
3. Menelaah Soal Tes	28
4. Melakukan Uji Coba	29
5. Menganalisis Butir Soal Tes	29
C. Uji Coba Produk	32
1. Desain Uji Coba	32
2. Subjek Uji Coba	32
3. Jenis Data	33
4. Instrumen Pengumpulan Data	33
5. Teknik Analisis Data	34
BAB IV HASIL PENELITIAN	37
A. Penyajian dan Analisis Data Pengembangan Instrumen Penilaian <i>Higher Order Thinking</i> untuk Pelajaran Matematika dengan Memasukkan Konteks Islam	37
1. Data Spesifikasi Instrumen Soal Tes	37
2. Menulis Butir Soal Tes	38
3. Menelaah Soal Tes	38
4. Data Uji Coba	54
5. Menganalisis Butir Soal Tes	54
B. Penyajian dan Analisis Data Karakteristik Instrumen Penilaian <i>Higher Order Thinking</i> untuk Pelajaran Matematika dengan Memasukkan Konteks Islam	55
1. Analisis Instrumen Penilaian Butir Soal	55
2. Analisis Karakteristik Instrumen Penilaian	57
3. Analisis Karakteristik Instrumen Penilaian Butir Soal	59
C. Kajian Produk Akhir	71

1. Temuan Terkait Proses Pengembangan Instrumen Penilaian <i>Higher Order Thinking</i> untuk Pelajaran Matematika dengan Memasukkan Konteks Islam	71
2. Temuan Terkait Karakteristik Instrumen Penilaian <i>Higher Order Thinking</i> untuk Pelajaran Matematika dengan Memasukkan Konteks Islam.....	74
BAB V PENUTUP	92
A. Simpulan	92
B. Saran	93
DAFTAR PUSTAKA	94
LAMPIRAN	

BAB I

PENDAHULUAN

A. Latar Belakang

Tujuan utama pembelajaran pada abad 21 adalah untuk mengembangkan dan meningkatkan berpikir tingkat tinggi siswa¹. Hal ini sesuai dengan kompetensi-kompetensi inti pada Standar Isi kurikulum 2013, utamanya kompetensi inti (KI) domain kognitif. Untuk setiap mata pelajaran, peserta didik harus dibekali dengan pengetahuan faktual, konseptual, dan prosedural berdasarkan rasa ingin tahu siswa tentang ilmu pengetahuan, teknologi, seni budaya terkait fenomena dan kejadian tampak mata². Berdasarkan Standar Isi tersebut, matematika sebagai salah satu mata pelajaran wajib diharapkan tidak hanya membekali siswa dengan kemampuan menggunakan perhitungan atau rumus dalam mengerjakan soal tes saja. Akan tetapi juga mampu melibatkan kemampuan bernalar dan analisisnya dalam memecahkan masalah³.

Salah satu tujuan dalam mata pelajaran matematika pada kurikulum 2013 adalah agar peserta didik memiliki kemampuan dalam memecahkan masalah yang meliputi kemampuan memahami masalah, merancang model matematika, menyelesaikan model dan memberikan solusi yang diperoleh. Dengan demikian, dalam pembelajaran matematika penting untuk mengembangkan kemampuan siswa dalam memecahkan masalah⁴.

Pemecahan masalah merupakan proses berpikir menggunakan pengetahuan dan keterampilan yang sudah ada

¹Zaenal Arifin - Heri Retnawati., “Analisis Instrumen Pengukur Higher Order Thinking Skills (HOTS) Matematika Siswa SMA”. (Paper presented at Seminar Nasional Matematika Dan Pendidikan Matematika, Universitas Negeri Yogyakarta, 2015), 784.

²Rosalia Hera Novita Sari, “Literasi Matematika: Apa, Mengapa dan Bagaimana?”. (Paper presented at Seminar Nasional Matematika dan Pendidikan Matematika, Universitas Negeri Yogyakarta, 2015), 713.

³Ibid, halaman 713.

⁴Distariana Haniffah - Janet Trineke Manoy., “Identifikasi Tipe Berpikir Dengan Soal Higher Order Thinking (Hot) Ditinjau Berdasarkan Kemampuan Matematika”, *Mathedunesa Jurnal Ilmiah Pendidikan Matematika*, 3:3, (2014), 40.

untuk menjawab pertanyaan yang belum terjawab⁵. Dalam pemecahan masalah, Polya mengemukakan empat langkah yang dapat dilakukan yaitu memahami masalah, merencanakan penyelesaian, menyelesaikan masalah sesuai rencana dan diakhiri dengan melakukan pengecekan kembali terhadap semua langkah yang telah dikerjakan. Pada setiap langkah yang dilakukan, kemampuan berpikir tingkat tinggi siswa akan optimal dan berkembang. Hal ini terlihat dari uraian langkah pemecahan masalah yang benar-benar menuntut siswa untuk dapat menguraikan permasalahan, menemukan ide-ide dalam memecahkan masalah, menemukan alternatif lain yang mungkin kemudian memilih salah satu alternatif yang paling dapat diterima oleh logika normal untuk diterapkan dalam penyelesaian masalah⁶.

Penting bagi siswa memiliki kemampuan berpikir tingkat tinggi. Seorang siswa yang berkembang kemampuan berpikir tingkat tingginya akan memiliki ketajaman analisis, memiliki kemampuan evaluasi yang baik, dan memiliki kemampuan menciptakan atau menghasilkan sesuatu⁷. Untuk melatih siswa berpikir tingkat tinggi diperlukan latihan soal yang berkualitas, soal yang tidak hanya menguji kemampuan mengingat, memahami serta aplikasi tetapi mencakup pula analisis, evaluasi, dan kreasi. Soal yang melatih kemampuan menganalisis, mengevaluasi dan mengkreasi adalah soal *Higher Order Thinking* (HOT).

Banyak peneliti yang telah melakukan penelitian mengenai HOT. Misalkan penelitian yang dilakukan oleh Agus Budiman - Jailani, Samritin, dan Zaenal Arifin - Heri Retnawati. Setiap peneliti mempunyai fokus tersendiri dalam setiap penelitiannya, sehingga penelitian mengenai *Higher Order Thinking* sangatlah beragam.

⁵Emi Rofiah, et.al., "Penyusunan Instrumen Tes Kemampuan Berpikir Tingkat Tinggi Fisika Pada Siswa SMP", *Jurnal Pendidikan Fisika*, 1: 2, (September, 2013), 18.

⁶Maya Kusumaningrum - Abdul Aziz Saefudin., "Mengoptimalkan Kemampuan Berpikir Matematika Melalui Pemecahan Masalah Matematika" (Paper presented at Seminar Nasional Matematika dan Pendidikan Matematika, Yogyakarta, 2012), 579.

⁷Samritin, Disertasi Doktor: "Pengembangan Instrumen Penilaian Kemampuan *Higher Order Thinking* Siswa SMP dalam Mata Pelajaran Matematika". (Yogyakarta: Universitas Negeri Yogyakarta, 2014), 2.

Agus Budiman dan Jailani melakukan penelitian pengembangan instrumen penilaian HOT pada pelajaran matematika. Dalam penelitian ini dihasilkan sebuah produk instrumen penilaian HOT berupa soal yang terdiri dari 24 butir soal pilihan ganda dan 19 soal uraian⁸.

Samritin melakukan penelitian pengembangan instrumen penilaian HOT pada mata pelajaran matematika tingkat SMP. Instrumen penilaian yang dihasilkan berupa soal tes HOT yang terdiri atas 12 butir soal dan semua butir soal berbentuk uraian. Untuk menghasilkan instrumen yang valid, dilakukan forum *Focus Group Discussion* (FGD) dan teknik Delphi⁹.

Zaenal Arifin dan Heri Retnawati melakukan penelitian pengembangan instrumen pengukur HOT pada pelajaran matematika. Instrumen pengukur HOT yang digunakan berupa tes tertulis, soal uraian, soal pilihan ganda dan soal jawaban singkat. Instrumen tersebut terbagi menjadi dua paket soal. Soal paket A sebanyak 15 butir soal dan soal paket B sebanyak 16 butir soal¹⁰.

Dari beberapa penelitian yang telah dilakukan, soal matematika HOT konteksnya masih terkait dengan kehidupan sehari-hari. Belum satupun penelitian yang mengembangkan instrumen penilaian HOT dengan memasukkan konteks Islam. Secara tidak langsung, konteks Islam dapat dimasukkan pada konten soal. Kajian terkait dengan soal matematika dengan memasukkan konteks Islam, sudah banyak dilakukan. Akan tetapi, belum ada kajian terkait dengan soal matematika HOT dengan memasukkan konteks Islam. Berdasarkan hal tersebut, akan diadakan penelitian dengan judul “Pengembangan Instrumen Penilaian *Higher Order Thinking* untuk Pelajaran Matematika dengan Memasukkan Konteks Islam”.

⁸Agus Budiman – Jailani., “Pengembangan Instrumen Asesmen *Higher Order Thinking Skill* (HOTS) Pada Mata Pelajaran Matematika SMP Kelas VIII Semester 1”, *Jurnal Riset Pendidikan Matematika*, 1: 2, (November, 2014).

⁹Samritin, Disertasi Doktor: “Pengembangan Instrumen Penilaian Kemampuan *Higher Order Thinking* Siswa SMP dalam Mata Pelajaran Matematika”. (Yogyakarta: Universitas Negeri Yogyakarta, 2014).

¹⁰Zaenal Arifin – Heri Retnawati., “*Analisis Instrumen Pengukur Higher Order Thinking Skills (HOTS) Matematika Siswa SMA*” (Paper presented at Seminar Nasional Matematika dan Pendidikan Matematika UNY, Yogyakarta, 2015).

B. Rumusan Masalah Penelitian

Berdasarkan latar belakang yang telah diuraikan di atas, permasalahan penelitian dapat dirumuskan sebagai berikut:

1. Bagaimana proses pengembangan instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam?
2. Bagaimana karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam?

Karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan dapat diketahui dari pertanyaan berikut ini:

- a. Bagaimana daya pembeda instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan?
- b. Bagaimana tingkat kesukaran instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan?
- c. Bagaimana nilai *guessing* instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan?

C. Tujuan Penelitian dan Pengembangan

Berdasarkan rumusan masalah di atas, tujuan penelitian penelitian ini adalah sebagai berikut:

1. Untuk mendeskripsikan proses pengembangan instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam.
2. Untuk mendeskripsikan karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang meliputi:
 - a. Untuk mendeskripsikan daya pembeda instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan.
 - b. Untuk mendeskripsikan tingkat kesukaran instrumen penilaian *Higher Order Thinking* untuk pelajaran

matematika dengan memasukkan konteks Islam yang dikembangkan.

- c. Untuk mendeskripsikan nilai guessing instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan.

D. Spesifikasi Produk yang Dikembangkan

Produk yang dikembangkan dalam penelitian ini berupa *prototype* instrumen yang memiliki karakteristik untuk mengukur *Higher Order Thinking* siswa Madrasah Tsanawiyah dalam mata pelajaran matematika. Karakteristik instrumen yang dimaksud adalah daya pembeda, tingkat kesukaran dan nilai guessing. Butir soal tes yang dikembangkan mengacu pada Kompetensi Dasar 3 (KD 3, pengetahuan) yang ada pada Kurikulum 2013.

E. Manfaat Pengembangan

Berdasarkan tujuan penelitian, adapun beberapa manfaat dari penelitian ini sebagai berikut:

1. Hasil pengembangan ini dapat dijadikan acuan dalam pengembangan instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam untuk konteks yang lebih luas.
2. Secara teoritis, hasil penelitian ini diharapkan dapat menambah wawasan tentang pengembangan instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam dalam mata pelajaran matematika.
3. Secara praktis, instrumen yang dihasilkan dalam penelitian dapat digunakan oleh guru yang akan menilai kemampuan *Higher Order Thinking* siswa dalam mata pelajaran matematika.

F. Asumsi dan Keterbatasan Pengembangan

1. Asumsi Pengembangan

Siswa yang menjadi subjek penelitian ini harus menjawab soal-soal dengan baik dan sungguh-sungguh,

sehingga respon yang diberikan benar-benar menggambarkan kemampuan sesungguhnya.

2. Keterbatasan Pengembangan

Pengembangan instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dihasilkan dalam penelitian ini masih terbatas pada siswa tingkat Madrasah Tsanawiyah. Uji coba dilakukan di tiga kota yaitu Gresik, Sidoarjo, dan Surabaya.

G. Definisi Operasional

Untuk menghindari terjadinya perbedaan penafsiran terhadap istilah yang digunakan dalam penelitian ini, maka didefinisikan beberapa istilah berikut.

1. Instrumen penilaian adalah alat ukur yang digunakan dalam penelitian berupa soal pilihan ganda untuk menilai hasil belajar siswa.
2. *Higher Order Thinking* (HOT) adalah proses berpikir yang tidak sekedar menghafal dan menyampaikan kembali pengetahuan yang diketahui akan tetapi suatu pemikiran dengan tingkat kesulitan yang lebih tinggi kepada siswa untuk melatih kemampuan menganalisis, mengevaluasi dan mencipta.
3. Daya pembeda adalah alat ukur yang digunakan untuk mengetahui kesanggupan soal dalam membedakan siswa yang berkemampuan rendah dan siswa yang berkemampuan tinggi.
4. Tingkat kesukaran adalah alat ukur yang digunakan untuk mengukur kemampuan siswa menjawab menjawab soal yang benar.
5. Nilai *guessing* adalah alat ukur yang digunakan untuk penyebaran *option* pada setiap butir soal yang berfungsi secara merata atau tidak, untuk mengukur kualitas pengecoh.

BAB II KAJIAN PUSTAKA

A. Higher Order Thinking

1. Pengertian *Higher Order Thinking*

Dalam Bahasa Inggris *Higher Order Thinking* (HOT) diartikan sebagai berpikir tingkat tinggi. Baker berpendapat bahwa HOT dapat terbentuk dari keterampilan metakognisi, seperti perencanaan dan *self-checking*, pemecahan masalah, serta proses intelektual¹. Definisi yang sama dikemukakan oleh Wang & Wang yang menyatakan bahwa HOT sebagai proses berpikir yang lebih dari sekedar hafalan dan pemahaman saja, proses berpikir melibatkan berbagai proses kognitif seperti menyimpulkan, mengidentifikasi, melakukan eksplorasi, menilai hasil dan sebagainya².

King et al menyatakan bahwa “*Higher order thinking skills include critical, logical, reflective, metacognitive, and creative thinking*”. Artinya, kemampuan berpikir tingkat tinggi memuat berpikir kritis, logis, reflektif, metakognitif, dan kreatif³. Sejalan dengan paham King, McMahon mendefinisikan HOT sebagai proses mental yang memungkinkan siswa untuk mengembangkan pengetahuan faktual prosedural, konseptual dan metakognisi dalam domain berpikir kritis dan kreatif⁴.

Wardana mengemukakan bahwa kemampuan berpikir tingkat tinggi adalah proses berpikir yang melibatkan aktivitas mental dalam usaha mengeksplorasi pengalaman yang kompleks, reflektif dan kreatif yang dilakukan secara sadar untuk mencapai tujuan yaitu memperoleh pengetahuan yang meliputi tingkat berpikir analisis,

¹Samritin, Disertasi Doktor: “*Pengembangan Instrumen Penilaian Kemampuan Higher Order Thinking Siswa SMP dalam Mata Pelajaran Matematika*”. (Yogyakarta: Universitas Negeri Yogyakarta, 2014), 2.

²Shouhong Wang - Hai Wang., “Supporting Higher-Order Thinking In E-Learning Environment”, *The College of Information Sciences and Technology*, (2010).

³Fj King, et.al., “Higher Order Thinking Skills”, *Center for Advancement of Learning and Assessment*, (1998).

⁴McMahon G, “Critical Thinking and ICT Integration in a Western Australian Secondary School”, *Educational Technology & Society*, 12:4 (2009).

sintesis, dan evaluasi. Menurut Corebima keterampilan berpikir tingkat tinggi juga dapat diketahui dari kemampuan kognitif siswa pada tingkat analisis, sintesis, dan evaluasi⁵.

Newman menyatakan bahwa “*Higher order thinking is challenges the student to interpret, analyze, or manipulate information*”⁶. Maksud yang dikemukakan oleh Newman bahwa HOT adalah suatu pemikiran yang menantang siswa untuk menginterpretasi, mengevaluasi, atau memanipulasi suatu informasi⁷. Secara tersirat, Newman menyatakan bahwa HOT membutuhkan suatu pemikiran yang kompleks dan mengevaluasi karena dalam pernyataannya terdapat poin menganalisis, interpretasi, atau memanipulasi informasi.

Berdasarkan beberapa pendapat di atas, dapat disimpulkan bahwa *Higher Order Thinking* (HOT) merupakan proses berpikir yang tidak sekedar menghafal dan menyampaikan kembali informasi yang diketahui akan tetapi suatu pemikiran yang menantang kepada siswa untuk menganalisis, mengevaluasi dan mencipta suatu informasi.

2. Soal *Higher Order Thinking*

Tes merupakan salah satu upaya pengukuran terencana yang digunakan oleh guru untuk mencoba menciptakan kesempatan bagi siswa dalam memperlihatkan prestasi mereka yang berkaitan dengan tujuan yang telah ditentukan. Tes terdiri atas sejumlah soal yang harus dikerjakan siswa. Setiap soal dalam tes menghadapkan siswa pada suatu tugas dan menyediakan kondisi bagi siswa untuk menanggapi tugas atau soal tersebut⁸.

⁵Femmy Kawuwung, “Profil Guru, Pemahaman Kooperatif NHT, dan Kemampuan Berpikir Tingkat Tinggi Di Smp Kabupaten Minahasa Utara”, *El-Hayah*, 1: 4, (Maret, 2011), 158.

⁶Yousef Abosalem, “Assessment Techniques And Students’ Higher-Order Thinking Skills”, *International Journal of Secondary Education*, 4: 1, (Maret, 2016).

⁷Guntur Suhandoyo - Pradnyo Wijayanti., “Profil Kemampuan Berpikir Kreatif Siswa Dalam Menyelesaikan Soal Higher Order Thinking Ditinjau Dari Adversity Quotient (AQ)”, *Jurnal Ilmiah Pendidikan Matematika*, 3: 5, (2016), 158.

⁸Ana Ratna Wulan, “Pengertian Dan Esensi Konsep Evaluasi, Asesmen, Tes, Dan Pengukuran”, On line at http://file.upi.edu/direktori/fpmipa/jurpend_biologi/anaratnawulan/pengertianasesmen.Pdf, pada tanggal 4 Juni 2017.

Terdapat hal yang penting dalam penulisan soal. Dalam penulisan soal tidak hanya sekedar membuat soal tetapi harus memperhatikan kaidah penulisan soal. Hal ini harus diperhatikan agar soal yang dibuat sesuai dengan tujuan pembuatannya yaitu sebagai alat untuk mengevaluasi keberhasilan pembelajaran dan mengembangkan kemampuan berpikir siswa⁹. Guru seringkali memberikan soal pada setiap akhir pembelajaran sebagai alat evaluasi pembelajaran yang telah dilakukan.

Komarudin melakukan wawancara dengan beberapa guru dalam penelitiannya. Hasil wawancara tersebut menunjukkan, ketika siswa di berikan soal yang memuat suatu permasalahan oleh guru, cara berpikir siswa cenderung sama dengan contoh - contoh yang telah diberikan oleh guru. Namun pada saat siswa diberikan soal yang sedikit berbeda dari contoh yang telah diberikan, maka siswa akan merasa kesulitan untuk menyelesaikan soal tersebut. Pada kondisi yang demikian biasanya siswa hanya dituntut untuk menerima sesuatu yang dianggap penting oleh guru dan menghafalnya saja. Sedangkan cara berpikir siswa menjadi terhambat dan tidak dapat berkembang secara optimal. Selain itu kemampuan dari siswa dalam memecahkan masalah tergolong masih rendah¹⁰.

Rendahnya kemampuan siswa dalam memecahkan masalah perlu diperbaiki dengan memberikan latihan soal-soal yang berbeda dengan contoh yang telah diberikan oleh guru. Salah satunya dengan menggunakan soal *Higher Order Thinking* (HOT)¹¹. Soal objektif yang digunakan untuk mengukur kemampuan siswa dalam menyelesaikan soal HOT adalah soal pilihan ganda. Gronlund & Kinn juga menyatakan bahwa soal objektif tipe pilihan ganda (*multiple*

⁹Indah Hesti Pratiwi, Disertasi Doktor: “Kemampuan Guru Mata Pelajaran Ipa Dalam Pembuatan Soal HOT (Higher Order Thinking) Dan Kesesuaian Penulisan Soal Di Smp Negeri 1 Kragan Rembang”. (Surakarta: Universitas Muhammadiyah Surakarta, 2015).

¹⁰Komarudin A, “*Analisis Tipe Berpikir Dengan Soal Higher Order Thinking Ditinjau Berdasarkan Kemampuan Matematika Siswa*” (Paper presented at Seminar Nasional Matematika Dan Pendidikan Matematika Universitas Yogyakarta, Yogyakarta, 2015), 986.

¹¹Distariana Haniffah - Janet Trineke Manoy., “Identifikasi Tipe Berpikir Dengan Soal *Higher Order Thinking* (HOT) Ditinjau Berdasarkan Kemampuan Matematika”, *Jurnal Ilmiah Pendidikan Matematika*, 3: 3, (2014), 41.

choice) diakui sebagai soal yang paling luas digunakan dan diaplikasikan¹².

Soal objektif pilihan ganda merupakan jenis soal yang meminta siswa memilih jawaban yang benar dari beberapa pilihan jawaban, atau menjawab pertanyaan dengan satu kata atau satu istilah¹³. Soal pilihan ganda terdiri dari pokok soal (*stem*) dan pilihan jawaban (*option*). Pilihan jawaban terdiri atas kunci jawaban dan pengecoh¹⁴. Dalam penulisan soal HOT bentuk pilihan ganda, terbilang cukup sulit. Dalam penyusunannya sendiri, soal pilihan ganda haruslah dapat mengukur keterampilan berpikir kritis siswa dan soal yang diberikan harus kompleks. Terdapat kelebihan dalam penulisan soal HOT bentuk pilihan ganda, diantaranya¹⁵.

- a. Bentuk soal pilihan ganda merupakan bentuk soal yang paling luas penggunaannya, sehingga guru yang mengajar dengan jumlah peserta didik yang banyak dapat menggunakannya.
- b. Soal pilihan ganda memiliki tingkat objektivitas dan validitas yang tinggi, sehingga dapat digunakan sebagai soal yang standar.
- c. Dalam pengujiannya, soal pilihan ganda terbilang lebih cepat dan lebih efisien serta dalam analisis penskorannya lebih mudah.
- d. Penyusunan butir soal pilihan ganda untuk mengukur keterampilan berpikir kritis tidak mudah, memerlukan banyak waktu dan sumber referensi, sehingga diperlukan perhatian khusus untuk menyusunnya

Ada beberapa kaidah tersendiri dalam penulisan soal HOT bentuk pilihan ganda. Menurut Kubiszyn & Borich, soal pilihan ganda umumnya ditulis dengan tujuan pada

¹²Hartini – Sukardjo., “Pengembangan Higher Order Thinking Multiple Choice Test Untuk Mengukur Keterampilan Berpikir Kritis Ipa Kelas VII SMP/Mts”, *Jurnal Inovasi Pendidikan IPA*, 1: 1, (April, 2015), 90.

¹³Kementerian Pendidikan Dan Kebudayaan, *Modul Teknik Penulisan Soal Berpikir Orde Lebih Tinggi/Higher Order Thinking (HOT)*.

¹⁴Kusaeri, *Acuan & Teknik Penilaian Proses & Hasil Belajar dalam Kurikulum 2013* (Yogyakarta: AR-RUZZ MEDIA 2014).

¹⁵Hartini – Sukardjo., “Pengembangan Higher Order Thinking Multiple Choice Test Untuk Mengukur Keterampilan Berpikir Kritis Ipa Kelas VII SMP/Mts”, *Jurnal Inovasi Pendidikan IPA*, 1: 1, (April, 2015), 89.

level pengetahuan. Langkah pertama dalam menulis soal pilihan ganda untuk mengukur keterampilan tingkat tinggi adalah menentukan tujuan yang mengukur memahami, mengaplikasi, menganalisis, mengevaluasi, dan mengkreasi atau tingkat berpikir yang lebih tinggi dari mengetahui. Pendekatan yang disarankan untuk mengukur berpikir tingkat tinggi, yaitu: (1) menggunakan gambar, grafik tabel; dan (2) menggunakan situasi atau ide baru¹⁶.

Nitko & Brookhart menjelaskan cara yang sama seperti yang dikemukakan oleh Kubiszyn & Borich, yaitu butir soal berpikir tingkat tinggi harus menggunakan bahan atau materi baru. Nitko & Brookhart menjelaskan cara tersebut dengan lebih rinci. Cara yang disarankan oleh Nitko & Brookhart untuk membuat materi baru dalam membuat butir soal berpikir tingkat tinggi, yaitu dengan menggunakan *context dependent item sets* (perlengkapan butir yang bergantung pada konteks). *Context dependent item sets* terdiri atas *the introductory materials* (materi pengantar) yang diikuti oleh beberapa item. Peserta didik harus berpikir mengenai informasi pada materi pengantar untuk menjawab pertanyaan untuk memecahkan masalah¹⁷.

Context dependent item sets ini biasanya disebut dengan latihan menafsirkan. Materi pengantar untuk membuat butir soal berpikir tingkat tinggi di antaranya, yaitu: (a) mengutip dari materi bacaan, (b) gambar foto, (c) grafik, (d) gambar, (e) ayat, (f) formula, (g) tabel data, (h) daftar kata dan simbol, (i) contoh, (j) peta, (k) film, dan (l) suara rekaman. Penggunaan materi pengantar dapat membuat butir soal menjadi butir soal berpikir tingkat tinggi¹⁸.

Pembuatan butir soal dengan materi pengantar merupakan langkah yang kompleks. Materi pengantar harus

¹⁶Kubiszyn – Borich, *Educational testing and measurement classroom application and practice (seventh edition.)*, (New York: John Wiley & Sons, Inc, 2002).

¹⁷Nitko – Brookhart, *Educational assessment of student 6th*, (Boston: Pearson Education, Inc, 2011).

¹⁸Hartini, Disertasi Doktor: “Pengembangan Higher Order Thinking Multiple Choice Test Untuk Mengukur Keterampilan Berpikir Kritis Ipa Kelas VII SMP/Mts” (Yogyakarta: Universitas Negeri Yogyakarta, 2013).

sesuai dengan fakta dan relevan dengan konsep materi pada butir soal tersebut¹⁹. Dapat ditarik kesimpulan bahwa soal HOT bentuk pilihan ganda merupakan soal yang kompleks dan menuntut siswa untuk berpikir kritis. Menurut Krathworl indikator untuk mengukur HOT meliputi:²⁰

a. Menganalisis

- 1) Menganalisis informasi yang masuk dan membagi-bagi atau menstrukturkan informasi ke dalam bagian yang lebih kecil untuk mengenali pola atau hubungannya,
- 2) Mampu mengenali serta membedakan faktor penyebab dan akibat dari sebuah skenario yang rumit,
- 3) Mengidentifikasi/merumuskan pertanyaan.

b. Mengevaluasi

- 1) Memberikan penilaian terhadap solusi, gagasan, dan metodologi dengan menggunakan kriteria yang cocok atau standar yang ada untuk memastikan nilai efektivitas atau manfaatnya,
- 2) Membuat hipotesis, mengkritik dan melakukan pengujian,
- 3) Menerima atau menolak suatu pernyataan berdasarkan kriteria yang telah ditetapkan.

c. Mencipta

- 1) Membuat generalisasi suatu ide atau cara pandang terhadap sesuatu,
- 2) Merancang suatu cara untuk menyelesaikan masalah,
- 3) Mengorganisasikan unsur-unsur atau bagian-bagian menjadi struktur baru yang belum pernah ada sebelumnya.

Untuk mencapai kemampuan berpikir tingkat tinggi, siswa harus dibiasakan memecahkan permasalahan yang membutuhkan pemikiran untuk menganalisis, menilai, dan

¹⁹Hartini – Sukardjo., “Pengembangan Higher Order Thinking Multiple Choice Test Untuk Mengukur Keterampilan Berpikir Kritis Ipa Kelas VII SMP/Mts”, *Jurnal Inovasi Pendidikan IPA*, 1: 1, (April, 2015), 95.

²⁰David R. Krathwohl, “A Revision Of Bloom's Taxonomy: An Overview”, *Theory Into Practice*, 41: 4, (2002).

mencipta. Butkowski dalam tesisnya juga menyimpulkan bahwa kemampuan berpikir tingkat tinggi matematika seseorang tidak dapat diperoleh secara instan, namun harus dilatihkan oleh guru dalam pembelajaran matematika²¹.

3. Karakteristik *Higher Order Thinking*

Higher Order Thinking (HOT) mempunyai berbagai karakteristik baik dari segi bentuk soalnya maupun konten materi. Menurut Resnick karakteristik HOT diantaranya adalah non algoritmik, bersifat kompleks, *multiple solutions* (banyak solusi), melibatkan variasi pengambilan keputusan dan interpretasi, penerapan *multiple criteria* (banyak kriteria), dan bersifat *effortful* (membutuhkan banyak usaha)²².

Conklin menyatakan bahwa “*characteristics of higher-order thinking skills: higher-order thinking skills encompass both critical thinking and creative thinking*”. Maksudnya, karakteristik kemampuan berpikir tingkat tinggi mencakup berpikir kritis dan berpikir kreatif. Berpikir kritis dan kreatif merupakan dua kemampuan manusia yang sangat mendasar karena berpikir kritis dan berpikir kreatif dapat mendorong seseorang untuk senantiasa memandang setiap permasalahan yang dihadapi secara kritis, dan mencoba mencari penyelesaiannya secara kreatif, sehingga diperoleh suatu hal baru yang lebih baik dan bermanfaat bagi kehidupannya²³.

Senk et al menjelaskan karakteristik berpikir tingkat tinggi sebagai “*solving tasks where no algorithm has been taught, where justification or explanation are required, and where more than one solution may be possible*”²⁴. Artinya, karakteristik HOT sebagai kemampuan untuk

²¹Siti Sofiyah, et.al., “Pengembangan Paket Tes Kemampuan Berpikir Tingkat Tinggi Matematika Berdasarkan Revisi Taksonomi Bloom Pada Siswa Kelas V SD”, *Artikel Ilmiah Mahasiswa*, 1: 1, (2015), 2.

²²Tony Thompson, “An Analysis of Higher-Order Thinking on Algebra I End-of Course Tests”, *International Journal for Mathematics Teaching and Learning*, (2011).

²³Agus Budiman – Jailani., “Pengembangan Instrumen Asesmen Higher Order Thinking Skill (HOTS) Pada Mata Pelajaran Matematika Smp Kelas Viii Semester 1”, *Jurnal Riset Pendidikan Matematika*, 1: 2 (November, 2014), 141.

²⁴Tony Thompson, “An Analysis of Higher-Order Thinking on Algebra I End-of Course Tests”, *International Journal for Mathematics Teaching and Learning*, (2011).

menyelesaikan tugas-tugas dimana tidak ada algoritma yang telah diajarkan, yang membutuhkan justifikasi atau penjelasan dan mungkin mempunyai lebih dari satu solusi yang mungkin.

B. Karakteristik Soal *Higher Order Thinking*

Kemampuan siswa didasarkan atas hasil analisis terhadap respon atau jawaban yang diberikan siswa terhadap tes yang diberikan. Secara garis besar, terdapat dua teori yang digunakan dalam analisis hasil tes, yaitu teori tes klasik (*Classical Test Theory*, CTT) dan teori respon butir (*Item Response Theory*, IRT)²⁵. Menurut O'Connor et al, CTT merupakan suatu model pengukuran berdasarkan informasi yang didapatkan pada level skor tes²⁶. Sejalan dengan pemahaman O'Connor, Hambleton dan Jones mendefinisikan CTT sebagai teori mengenai skor tes yang mengenalkan tiga konsep yaitu *test score/observed score*, *true score* dan *error score*²⁷. Lain halnya dengan CTT, IRT merupakan pengembangan baru dari teori CTT. Hambleton et al mendefinisikan IRT sebagai teori yang menyatakan bahwa hasil tes dapat diprediksikan atau dijelaskan melalui serangkaian faktor yang disebut dengan sifat atau karakter (*trait*), karakter terpendam (*latent trait*) atau kemampuan (*abilities*) dan hubungan antara jawaban peserta tes dengan kemampuannya dapat dijelaskan dengan grafik karakteristik butir atau *Item Characteristic Curve* (ICC).

Classical Test Theory telah berkembang secara luas dan menjadi aliran utama di kalangan ahli psikologi dan pendidikan, serta bidang kajian perilaku (*behavioral*) yang lain, selama 20 dekade. Namun kemudian ditemukan bahwa CTT memiliki kelemahan karena bersifat *examinee sample dependent* dan *item sample dependent*. Kelemahan tersebut kemudian memicu teori baru yang lebih memadai, yaitu IRT. Jika CTT memfokuskan

²⁵Budi Manfaat – Zara Zahra Anasha., “Analisis Kemampuan Berpikir Kritis Matematik Siswa Dengan Menggunakan Graded Response Models (GRM)”, (Paper Presented at Seminar Nasional Matematika dan Pendidikan Matematika, Yogyakarta, 2013), 120.

²⁶Nahadi et.al., “Pengembangan Dan Analisis Soal Ulangan Kenaikan Kelas Kimia SMA Kelas X Berdasarkan Classical Test Theory Dan Item Response Theory”, *Jurnal Pengajaran MIPA*, 16: 2, (Oktober 2011), 110.

²⁷Tbid, halaman 110

pada informasi pada level tes, IRT terutama memfokuskan pada informasi pada level butir sehingga diharapkan dapat menutupi kekurangan yang terdapat pada CTT²⁸.

Item Response Theory (IRT) dapat dikatakan sebagai analisis butir soal secara modern. Teori ini merupakan suatu teori yang menggunakan fungsi matematika untuk menghubungkan antara peluang menjawab benar suatu skala dengan kemampuan siswa. Nama lain IRT adalah *Latent Trait Theory* (LTT), atau *Characteristics Curve Theory* (ICC). Asal mula IRT adalah kombinasi suatu versi hukum phi-gamma dengan suatu analisis faktor butir soal (*item factor analisis*) kemudian bernama *Teori Trait Latent* (*Latent Trait Theory*), sekarang secara umum dikenal menjadi teori jawaban butir soal (*Item Response Theory*)²⁹.

Hambleton dan Swaminathan menyatakan bahwa tujuan utama IRT adalah memberikan kesamaan antara statistik soal dan estimasi kemampuan. Ada tiga keuntungan IRT adalah: (1) asumsi banyak soal yang diukur pada *trait* yang sama, perkiraan tingkat kemampuan peserta didik adalah independen; (2) asumsi pada populasi tingkat kesukaran, daya pembeda merupakan independen sampel yang menggambarkan untuk tujuan kalibrasi soal; (3) statistik yang digunakan untuk menghitung tingkat kemampuan siswa diperkirakan dapat terlaksana³⁰. Dapat disimpulkan bahwa IRT merupakan hubungan antara probabilitas jawaban suatu butir soal yang benar dan kemampuan siswa atau tingkatan/level prestasi siswa.

Dalam penelitian ini, peneliti ingin melihat karakteristik soal *Higher Order Thinking* dengan menggunakan *Item Response Theory* (IRT). Model IRT yang digunakan adalah model tiga parameter logistik. Model tiga parameter logistik merupakan analisis karakteristik soal dengan menitikberatkan pada tiga parameter. Tiga parameter yang dimaksud adalah

²⁸Budi Manfaat – Zara Zahra Anasha., “Analisis Kemampuan Berpikir Kritis Matematik Siswa Dengan Menggunakan Graded Response Models (GRM)”, (Paper Presented at Seminar Nasional Matematika dan Pendidikan Matematika, Yogyakarta, 2013), 120.

²⁹Aslam Fatkhudin, et.al., “Item Response Theory Model Empat Parameter Logistik Pada Computerized Adaptive Test”, *Jurnal Sistem Informasi Bisnis*, 2 (Maret, 2014), 122.

³⁰Hambleton, R.K. & Swaminathan, H., *Item Response Theory: Principles and Applications* (Boston: Kluwer. Nijhoff Publishing 1985).

tingkat kesukaran, daya pembeda dan *guessing* (pengecoh) soal. IRT merupakan perkembangan dari CTT, sehingga tiga parameter yang digunakan berawal dari parameter CTT,

1. Daya Pembeda

Pada model CTT, daya pembeda merupakan pengukuran sejauh mana suatu butir soal mampu membedakan peserta didik yang sudah menguasai kompetensi dengan peserta didik yang belum/kurang menguasai kompetensi berdasarkan kriteria tertentu. Semakin tinggi koefisien daya pembeda suatu butir soal, semakin mampu butir soal tersebut membedakan antara peserta didik yang menguasai kompetensi dengan peserta didik yang kurang menguasai kompetensi³¹.

Pada model IRT, parameter daya pembeda dikenal sebagai parameter a . Parameter a menunjukkan seberapa baik sebuah butir soal dapat membedakan peserta tes dengan tingkat kemampuan yang berbeda-beda. Tingkat kemampuan diklasifikasikan menjadi dua kelompok. Peserta tes pada kelompok atas akan memiliki pilihan jawaban benar lebih banyak daripada peserta tes pada kelompok bawah. Butir soal yang baik biasanya memiliki rentang nilai daya pembeda dari 0,5 sampai 2. Hal ini digambarkan dengan plot grafik ICC. Semakin tinggi kemiringan yang dihasilkan grafik ICC, semakin tinggi pula daya pembeda suatu butir soal. Daya pembeda yang tinggi menunjukkan bahwa peserta tes yang memiliki skor tinggi cenderung menjawab butir soal dengan benar, sedangkan peserta tes dengan skor rendah cenderung memilih pilihan jawaban yang salah³².

2. Tingkat Kesukaran

Dalam konteks CTT, perhitungan tingkat kesukaran soal adalah pengukuran seberapa besar derajat kesukaran suatu soal.³³ Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan bahwa soal

³¹Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

³²Nahadi et.al., “Pengembangan Dan Analisis Soal Ulangan Kenaikan Kelas Kimia SMA Kelas X Berdasarkan Classical Test Theory Dan Item Response Theory”, *Jurnal Pengajaran MIPA*, 16: 2, (Oktober 2011), 109.

³³Ibid, halaman 110.

tersebut baik. Suatu soal tes hendaknya tidak terlalu sukar dan tidak pula terlalu mudah³⁴. Menurut Arikunto, soal yang terlalu mudah tidak merangsang peserta didik untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan peserta didik menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya³⁵.

Dalam konteks IRT, tingkat kesukaran sebuah butir soal dikenal sebagai parameter ***b***. Parameter ***b*** adalah titik dimana kurva memiliki kemiringan paling tinggi. Semakin sukar suatu butir soal, semakin tinggi kemampuan yang diperlukan dari peserta tes untuk menjawab butir soal tersebut dengan benar. Suatu butir soal dengan nilai ***b*** yang tinggi adalah soal sukar, dimana peserta tes kelompok bawah tidak dapat menjawab dengan benar. Butir soal dengan nilai ***b*** rendah adalah soal mudah, dimana sebagian besar peserta tes, termasuk peserta tes kelompok bawah, akan memiliki peluang minimal setengah untuk menjawab soal tersebut dengan benar³⁶.

3. Nilai *Guessing*

Soal pilihan ganda sangat rentan dengan menebak (*guessing*) jawaban benar. Secara umum, ada dua bentuk menebak yaitu “*blind guessing*” dan “*informed guessing*”³⁷. *Blind guessing* (menebak buta) terjadi ketika peserta tes tidak tahu jawaban yang benar dan memilih secara acak, sedangkan *informed guessing* (menebak informasi) terjadi ketika peserta tes memilih jawaban atas pengetahuan yang pernah didapat sebelumnya³⁸.

Dalam model IRT, parameter faktor tebakan (*guessing*) dikenal sebagai parameter ***c***. Parameter ***c*** menunjukkan

³⁴Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

³⁵Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan* (Jakarta: Bumi Aksara, 2010).

³⁶Nahadi et.al., “Pengembangan Dan Analisis Soal Ulangan Kenaikan Kelas Kimia SMA Kelas X Berdasarkan Classical Test Theory Dan Item Response Theory”, *Jurnal Pengajaran MIPA*, 16: 2, (Oktober 2011), 110.

³⁷Song Gao, Doctoral Dissertation: “*The Exploration of The Relationship Between Guessing and Latent Ability in IRT Models*” (Carbondale: Southern Illinois University, 2011).

³⁸Nur Aida, Skripsi: “Pengembangan Instrumen Penilaian Hasil Belajar Kognitif Matematika Berdasarkan Model Pisa Pada Konten Quantity” (Surabaya: Universitas Islam Negeri Sunan Ampel Surabaya, 2017).

kecenderungan peserta tes kelompok bawah dapat menebak jawaban yang benar dari sebuah butir soal sehingga memiliki peluang lebih dari nol dalam menjawab soal dengan benar. Sebagai contoh, peserta tes yang memilih jawaban secara acak dari sebuah butir soal yang memiliki empat pilihan respon jawaban dapat menjawab butir soal dengan benar satu dari empat kali kesempatan, artinya peluang menjawab benar dengan menebak adalah 0,25. Gambar grafik ICC masih berbentuk-S, tetapi nilai terendah dari kurva adalah lebih dari nol³⁹.

C. Matematika dalam Konteks Islam

Al-Qur'an merupakan kitab Allah yang diturunkan kepada Nabi Muhammad SAW untuk dijadikan sebagai pedoman dan petunjuk bagi manusia dalam berbagai dimensi kehidupan. Umat Islam telah sepakat bahwa Al-Qur'an adalah kitabullah yang kekal, tidak terbatas pada dimensi ruang dan waktu, dan tidak ada sedikitpun keraguan. Al-Qur'an juga merupakan sumber ajaran Islam, termasuk di dalamnya sebagai dasar ilmu Islam⁴⁰.

Pendefinisian ilmu Islam sangatlah luas. Ilmu Islam dikatakan sebagai ilmu dengan segala ajaran yang bersumber pada Al-Qur'an, Sunnah Rasulullah (As-Sunnah) dan *Ra'yu* (hasil pikir manusia).⁴¹ Mufid mendefinisikan ilmu Islam sebagai ilmu yang mempelajari atau mengkaji tentang segala sesuatu yang berhubungan dan bersumber dari Islam atau yang berasal dari non Islam yang telah dijustifikasi dengan al-Qur'an sebagai konsep dasar⁴². Dapat ditarik kesimpulan bahwa ilmu Islam merupakan ilmu yang mempelajari tentang segala sesuatu yang bersumber pada Al-Qur'an, As-Sunnah dan *Ra'yu*.

Ilmu Islam mempunyai kajian yang sangat luas. Dalam bukunya yang terkenal *Muqaddimah*, Ibn Khaldun membagi

³⁹Nahadi et.al., "Pengembangan Dan Analisis Soal Ulangan Kenaikan Kelas Kimia SMA Kelas X Berdasarkan Classical Test Theory Dan Item Response Theory", *Jurnal Pengajaran MIPA*, 16: 2, (Oktober 2011), 110.

⁴⁰Fathul Mufid, "Posisi Al-Qur'an Dalam Struktur Dan Sumber Ilmu Islam", *Jurnal ADDIN Media Dialektika Ilmu Islam*, 2: 2, (Juli-Desember, 2010), 33.

⁴¹Ibid, halaman 34.

⁴²Fathul Mufid, "Integrasi Ilmu-Ilmu Islam", *Equilibrium*, 1: 1, (Juni, 2013).

ilmu yang berkembang di dunia Islam ke dalam dua kategori utama yaitu ilmu agama (*naqliyah*), dan ilmu rasional (*aqliyah*).

1. Ilmu *Naqliyah*

Ilmu *naqliyah* merupakan ilmu yang bersumber dari ajaran syari'at yaitu Al-Qur'an dan As-Sunnah yang telah ditetapkan bagi kita dari Allah SWT dan Rasul-Nya. Kategori ilmu *naqliyah* adalah ilmu tafsir, ilmu hadist, ilmu ushul fiqih dan lain sebagainya⁴³.

2. Ilmu *Aqliyah*

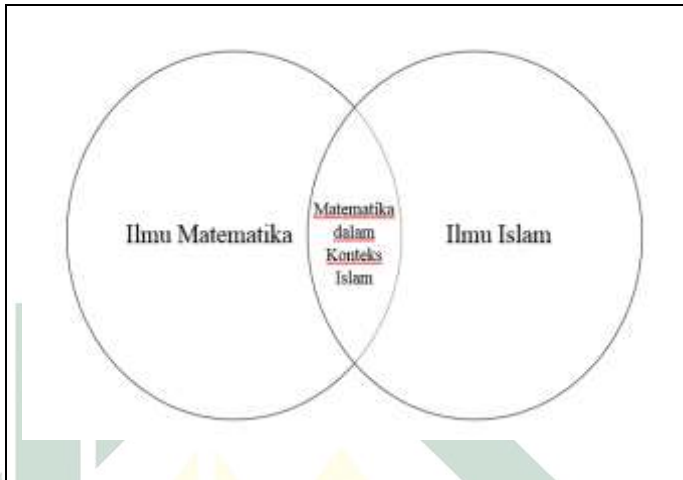
Ilmu *aqliyah* merupakan ilmu pengetahuan yang bisa didapat manusia karena alam berfikirnya dan dengan indera kemanusiaannya, segi-segi demonstrasinya dan aspek-aspek pengajarannya, sehingga penelitian dan penyelidikan menyampaikannya kepada mana yang benar dan yang salah, sesuai dengan kedudukannya sebagai manusia berfikir. Kategori ilmu *aqliyah* adalah ilmu mantiq, ilmu metafisika, ilmu matematika dan lain sebagainya⁴⁴.

Menurut Kamus Besar Bahasa Indonesia, matematika didefinisikan sebagai ilmu tentang bilangan, hubungan antara bilangan dan prosedur operasional yang digunakan dalam penyelesaian masalah mengenai bilangan. Dalam ilmu matematika banyak sekali bidang ilmu yang dipelajari. Ilmu matematika yang pada umumnya dipelajari di sekolah meliputi Aljabar, Geometri, Aritmetika, dan Analisis.

Integrasi mempunyai arti perpaduan. Jika ilmu Islam diintegrasikan dengan Ilmu Matematika akan membentuk suatu pemahaman baru yaitu matematika dalam konteks Islam. Perhatikan gambar berikut:

⁴³Fathul Mufid, "Posisi Al-Qur'an Dalam Struktur Dan Sumber Ilmu Islam", *Jurnal ADDIN Media Dialektika Ilmu Islam*, 2: 2, (Juli-Desember, 2010), 34.

⁴⁴Ibid, halaman 33



Gambar 2.1
Integrasi Ilmu Islam dengan Ilmu Matematika

Gambar di atas menunjukkan, jika kedua ilmu di integrasikan maka hasil dari irisan dari kedua ilmu tersebut membentuk suatu pemahaman baru yaitu matematika dalam konteks Islam. Pengintegrasian ini dapat diperkuat dengan ilmu *aqliyah* yang membahas tentang matematika⁴⁵. Adapun ilmu Islam yang dapat dikaitkan dengan soal matematika, diantaranya:

1. Ilmu Muamalah

Muamalah adalah bagian dari hukum Islam yang mengatur hubungan antara dua pihak atau lebih dalam suatu transaksi⁴⁶. Sedangkan menurut Mahmud, muamalah adalah ketentuan-ketentuan hukum mengenai hubungan perekonomian yang dilakukan anggota masyarakat, dan bertendensikan kepentingan material yang saling menguntungkan satu sama lain⁴⁷. Dapat diambil kesimpulan bahwa ilmu muamalah adalah ilmu yang mempelajari

⁴⁵Ibid, halaman 37.

⁴⁶AIN Sunan Ampel Press, *Konsep Dasar Fiqih Muamalah* (Sidoarjo: CV Cahaya Intan, 2014),

⁴⁷Dede Rosyada, *Hukum Islam dan Pranata Sosial* (Jakarta: RajaGrafindo Persada, 1993), 70-71.

tentang pengetahuan mengenai ketentuan hukum tentang usaha-usaha untuk memperoleh dan mengembangkan harta, jual-beli, hutang piutang dan lain sebagainya. Pembahasan ilmu muamalah berkaitan dengan jual-beli, sewa-menyewa, pinjam-meminjam, upah-mengupah dan lain sebagainya.

Dalam kehidupan sehari-hari, ilmu muamalah sering digunakan. Ilmu muamalah digunakan pada saat proses transaksi antara dua pihak atau lebih. Secara tidak langsung, proses transaksi yang dilakukan sebenarnya menggunakan ilmu matematika sebagai proses perhitungannya. Ilmu matematika yang berkaitan dengan ilmu muamalah adalah aritmetika sosial. Alasan peneliti menggunakan ilmu muamalah sebagai konteks matematika Islam karena ilmu muamalah berkaitan dengan aritmatika sosial. Dalam aritmetika sosial, semua materi yang ada telah berkaitan dengan ilmu muamalah.

2. Ilmu Faraid

Faraid adalah bagian yang telah ditentukan oleh syariat kepada yang berhak menerimanya⁴⁸. Menurut al-Qalyubi dan al-Umairah, faraid adalah ilmu tentang masalah bagian kewarisan⁴⁹. Ahmad mendefinisikan faraid sebagai ilmu tentang bagaimana membagi harta peninggalan seseorang setelah ia meninggal⁵⁰. Dapat diambil kesimpulan bahwa ilmu faraid adalah ilmu yang berkaitan dengan pembagian harta warisan, pengetahuan cara menghitung warisan dan pengetahuan pembagian warisan dari seseorang yang telah meninggal. Pada dasarnya ilmu faraid hanya berkaitan dengan berapa nilai masing-masing harta yang diterima oleh para ahli waris.

Ilmu faraid digunakan pada saat pembagian harta warisan dari seseorang yang telah meninggal kepada ahli waris. Dalam pembagian harta warisan, ada proses perhitungan untuk menentukan berapa bagian yang

⁴⁸Mardani, *Hukum Kewarisan Islam Di Indonesia* (Jakarta: PT RajaGrafindo Persada, 2014),

⁴⁹al-Qalyubi - al-Umairah, *Hasyitani' ala Minhajati Thalibin* (Beirut: Dar al-Fikr, juz 3, tth), 134.

⁵⁰Ahmad Jamil, et. All. *Sarana Penunjang Aktif Belajar Al Fath*. (Gresik: CV. Putra Kembar Jaya, 2007), 2.

diterima. Proses perhitungannya sendiri, menggunakan ilmu matematika. Ilmu matematika yang berkaitan dengan ilmu faraid adalah bilangan pecahan. Bilangan pecahan merupakan konsep dasar yang digunakan untuk melakukan perhitungan dalam ilmu faraid. Perhitungannya sendiri digunakan untuk menentukan berapa bagian yang diterima oleh ahli waris.

3. Perhitungan Zakat

Zakat secara bahasa berarti berkah, tumbuh, bersih, dan baik. Sedangkan secara istilah berarti sejumlah harta tertentu yang diwajibkan Allah diserahkan kepada sejumlah orang-orang yang berhak. Zakat ada dua macam yaitu, zakat fitrah dan zakat mal. Zakat fitrah merupakan zakat berupa makanan pokok yang wajib ditunaikan untuk setiap jiwa satu tahun sekali. Sedangkan zakat mal merupakan zakat yang berupa harta yang wajib ditunaikan (dikeluarkan) bagi pemilik harta setiap setahun sekali. Zakat mal wajib dikeluarkan apabila sudah mencapai nishab (batas minimal wajib zakat). Sebagai hukum Islam, pengeluaran dan perhitungan zakat sudah diatur dalam Al-Qur'an. Sebagai contoh perhitungan zakat mal. Jika seseorang sudah mencapai nishabnya, maka wajib untuk mengeluarkan zakat. Nishab emas adalah 93,6 gram dan zakatnya adalah 2,5 % atau sekitar 2,34 gram. Sedangkan nishabnya perak adalah 624 gram dan zakatnya adalah 2,5 % atau sekitar 15,6 gram⁵¹. Jika seseorang mempunyai emas atau perak yang lebih banyak, maka zakat yang dikeluarkan juga akan lebih banyak. Dilakukanlah perhitungan zakat untuk menentukan jumlah zakat yang akan dikeluarkan.

Perhitungan zakat dilakukan dengan menggunakan konsep matematika. Konsep matematika yang digunakan adalah bilangan pecahan. Bilangan pecahan sebagai dasar perhitungan untuk mengeluarkan zakat dari harta yang dimiliki. Secara tidak langsung, matematika menjadi peran penting dalam perhitungan zakat.

⁵¹Dewi Nurani dan Tri Wahyuni, Matematika Konsep dan Aplikasinya 1, (Jakarta: Pusat Perbukuan Dep Pendidikan Nasional, 2008), hal 74-76.

Banyak sekali ilmu Islam yang dapat dikaitkan dengan ilmu matematika. Dari penjelasan diatas, berikut merupakan contoh soal yang mana soal tersebut menggambarkan soal matematika dalam konteks Islam berdasarkan kajian di atas:

1. Andi wafat dengan meninggalkan harta sebanyak *Rp. 60.000.000,00*. Sebelumnya, Andi berwasiat bahwa harta yang dimiliki disumbangkan kepada panti asuhan sebanyak *Rp. 5.000.000,00*, masjid sebesar *Rp. 5.000.000,00* dan sisanya diberikan kepada kedua anaknya Ahmad dan Ulfa. Beberapa hari setelah pak Andi wafat, Budi tetangga pak Andi datang dengan memberitahukan bahwa pak Andi memiliki hutang sebesar *6.000.000*. Pak Andi merupakan sahabat terbaik dari Budi, sehingga Budi hanya menagih separuh dari hutang yang dimiliki pak Andi kepada pihak keluarga. Suatu hari, adik pak Andi datang dengan memberikan uang sebesar *10.000.000* sebagai pengganti uang yang dulu pernah dipinjam semasa almarhum masih hidup. Sisa dari harta yang dimiliki diwariskan kepada kedua anaknya. Jika Ahmad mendapatkan 2 bagian lebih banyak dari pada Ulfa, berapa jumlah warisan yang didapatkan oleh Ulfa?
 - a. *Rp. 19.000.000,00*
 - b. *Rp. 57.000.000,00*
 - c. *Rp. 38.000.000,00*
 - d. *Rp. 8.000.000,00*

Pembahasan:

- Harta warisan sebanyak *Rp. 60.000.000,00*.
- Disumbangkan ke masjid

$$\begin{aligned} &Rp. 60.000.000,00 - Rp. 5.000.000,00 \\ &= Rp. 55.000.000,00 \end{aligned}$$
- Disumbangkan ke panti asuhan

$$\begin{aligned} &Rp. 55.000.000,00 - Rp. 5.000.000,00 \\ &= Rp. 50.000.000,00 \end{aligned}$$
- Hutang Pak Andi

$$\begin{aligned} &Rp. 50.000.000,00 - Rp. 3.000.000,00 \\ &= Rp. 47.000.000 \end{aligned}$$
- Tambahan uang yang diterima

$$\begin{aligned} &Rp. 47.000.000,00 + Rp. 10.000.000,00 \\ &= Rp. 57.000.000,00 \end{aligned}$$

- Ahmad mendapat 2 bagian lebih banyak dari pada Ulfa, maka Ulfa mendapatkan $\frac{1}{3} \times Rp. 57.000.000,00 = Rp. 19.000.000,00$ (a)
 - 2. Ahmad, Alawi dan Ulfa masing-masing mempunyai uang sebanyak Rp. 600.000,00, Rp. 650.000,00 dan Rp. 700.000,00. Pada suatu hari, ketiga sahabat tersebut menginvestasikan uangnya kepada perusahaan ternama dikotanya. Komisi yang diberikan oleh perusahaan sebanyak 10% per bulan. Ketiga sahabat bernadzar, uang hasil komisi yang didapatkan dalam satu tahun akan disumbangkan kepada masjid didesanya. Ahmad menyumbang sebanyak 5%, Alawi sebanyak 15%, dan Ulfa sebanyak 10% dari komisi yang didapatkannya selama 1 tahun. Maka jawaban yang benar, uang yang diperoleh dari hasil investasi adalah . . .
- a. Ahmad < Alawi < Ulfa
 - b. Alawi < Ulfa < Ahmad
 - c. Ulfa < Ahmad < Alawi
 - d. Alawi < Ahmad < Ulfa

Pembahasan:

- Uang yang diperoleh Ahmad
Komisi per bulan Ahmad

$$Rp. 600.000,00 \times \frac{10}{100} = Rp. 60.000$$

Komisi Ahmad setahun

$$Rp. 60.000,00 \times 12 = Rp. 720.000,00$$

Uang yang disumbangkan

$$Rp. 720.000,00 \times \frac{5}{100} = Rp. 36.000,00$$

Uang yang diperoleh

$$Rp. 720.000,00 - Rp. 36.000,00 = Rp. 684.000,00$$

- Uang yang diperoleh Alawi
Komisi per bulan Alawi

$$Rp. 650.000,00 \times \frac{10}{100} = Rp. 65.000$$

Komisi Alawi setahun

$$Rp. 65.000,00 \times 12 = Rp. 780.000,00$$

Uang yang disumbangkan

$$Rp. 780.000,00 \times \frac{15}{100} = Rp. 117.000,00$$

Uang yang diperoleh

$$Rp. 780.000,00 - Rp. 117.000,00$$

$$= Rp. 663.000,00$$

- Uang yang diperoleh Ulfa
Komisi per bulan Ulfa

$$Rp. 700.000,00 \times \frac{10}{100} = Rp. 70.000$$

Komisi Ahmad setahun

$$Rp. 70.000,00 \times 12 = Rp. 840.000,00$$

Uang yang disumbangkan

$$Rp. 840.000,00 \times \frac{10}{100} = Rp. 84.000,00$$

Uang yang diperoleh

$$Rp. 840.000,00 - Rp. 84.000,00 = Rp. 756.000,00$$

Maka uang yang diterima adalah

$$Rp. 663.000,00 < Rp. 684.000,00 < Rp. 756.000,00$$

Alawi < Ahmad < Ulfa (d)

3. Seorang saudagar bernama Fulan menyimpan 200 gram emas dan 600 gram perak pada bulan Syawal 1436 H. Si Fulan adalah orang yang rajin membayar zakat dari harta yang dimilikinya. Pada bulan Dzulhijah 1437 H si Fulan membayar zakatnya. Jika x_1, x_2 adalah akar-akar persamaan $ax^2 - bx + 3 = 0$, dimana a adalah zakat mas dan b adalah zakat perak yang harus dikeluarkan si Fulan, maka tentukan nilai dari $x_1^2 + x_2^2$ (Adaptasi Soal Olmat UINSA '16)

a. 3

b. $\frac{39}{5}$

c. $\frac{3}{5}$

d. $\frac{51}{5}$

Pembahasan

$$\text{Zakat emas} = 2,5\% \times 200 = 5$$

$$\text{Zakat perak} = 2,5\% \times 600 = 15$$

$$\text{Maka persamaannya adalah } 5x^2 - 15x + 3 = 0$$

Akar-akarnya adalah x_1, x_2

$$x_1 + x_2 = -\frac{b}{a} = -\frac{-15}{5} = 3$$

$$\begin{aligned}
 x_1 \cdot x_2 &= \frac{c}{a} = \frac{3}{5} \\
 x_1^2 + x_2^2 &= (x_1 + x_2)^2 - 2x_1x_2 \\
 x_1^2 + x_2^2 &= (3)^2 - 2 \cdot \frac{3}{5} \\
 x_1^2 + x_2^2 &= 9 - \frac{6}{5} = \frac{39}{5} \text{ (b)}
 \end{aligned}$$

Tiga soal di atas merupakan contoh soal matematika dengan memasukkan konteks Islam. Masih banyak lagi soal matematika yang dapat dikaitkan dengan konteks Islam. Dalam penelitian ini, peneliti membuat soal matematika HOT yang berbeda dengan peneliti-peneliti sebelumnya. Jika banyaknya peneliti yang membuat soal matematika HOT hanya dikaitkan dengan materi pelajaran di sekolah saja, maka kali ini peneliti ingin memberikan gambaran yang baru dengan membuat soal matematika HOT dengan memasukkan konteks Islam. Keterbatasan pengetahuan siswa terkait materi yang diajarkan di sekolah, peneliti membatasi konteks Islam hanya pada ilmu muamalah. Materi terkait dengan ilmu muamalah adalah aritmatika sosial. Untuk pengaplikasian soal sendiri, siswa akan lebih mudah memahaminya karena soal yang diberikan sudah didapatkan sebelumnya. Konteks Islam yang disajikan pada soal, bisa juga pada gambaran pengetahuan terkait dengan keislaman yang sudah didapatkan oleh siswa.

BAB III

METODE PENELITIAN

A. Metode Penelitian dan Pengembangan

Jenis penelitian ini adalah penelitian pengembangan. Penelitian pengembangan ini mengadaptasi model pengembangan Mardapi yang terdiri dari sembilan langkah, yaitu (1) menyusun spesifikasi instrumen soal tes, (2) menulis butir soal tes, (3) menelaah soal tes, (4) melakukan uji coba, (5) menganalisis butir soal tes, (6) memperbaiki soal tes, (7) merakit soal tes, (8) melakukan tes, dan (9) menafsirkan hasil tes¹. Namun pada penelitian ini, peneliti membatasi model pengembangan hanya sampai dengan langkah kelima yaitu menganalisis butir soal tes. Dengan kata lain, peneliti tidak melakukan memperbaiki soal tes, merakit soal tes, melakukan tes dan menafsirkan hasil tes. Hal demikian dikarenakan, produk yang dihasilkan dalam penelitian ini hanya berupa *prototype*. Harapannya, penelitian ini dapat dikembangkan kembali dalam konteks yang lebih luas. Dalam penelitian ini, peneliti bertujuan untuk menghasilkan produk berupa *prototype* instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam. Produk *prototype* instrumen penilaian yang dihasilkan adalah soal pilihan ganda.

B. Prosedur Penelitian dan Pengembangan

Penelitian ini mengadaptasi model pengembangan Mardapi yang terdiri dari sembilan langkah. Namun pada penelitian ini, peneliti hanya sampai dengan langkah kelima, yaitu (1) menyusun spesifikasi instrumen soal tes, (2) menulis butir soal tes, (3) menelaah soal tes, (4) melakukan uji coba, dan (5) menganalisis butir soal tes. Berikut adalah penjelasan dari lima langkah pengembangan yang dilakukan oleh peneliti:

1. Menyusun Spesifikasi Instrumen Soal Tes

Langkah pertama dalam mengembangkan suatu instrumen tes adalah menetapkan spesifikasi instrumen soal tes. Spesifikasi instrumen soal tes berisi tentang uraian yang

¹Djemari Mardapi, *Teknik Penyusunan Instrumen Tes dan Non Tes* (Yogyakarta: Mitra Cendika, 2008), 88.

menggambarkan keseluruhan karakteristik yang harus dimiliki suatu tes. Disamping itu, spesifikasi instrumen soal tes yang dibuat haruslah jelas, agar mempermudah dalam penulisan soal. Berikut adalah langkah-langkah yang dilakukan dalam penyusunan spesifikasi instrumen soal tes, a. menentukan tujuan tes, b. menyusun kisi-kisi tes, dan c. memilih bentuk tes.

2. Menulis Butir Soal Tes

Menulis butir soal tes dilakukan setelah menyusun spesifikasi instrumen soal tes. Penulisan soal dilakukan dengan menjabarkan indikator menjadi pertanyaan-pertanyaan yang disesuaikan dengan spesifikasi instrumen soal tes yang telah dibuat. Kualitas tes secara keseluruhan harus memuat kriteria *Higher Order Thinking* (HOT). Soal yang nanti dibuat memuat tiga karakteristik soal yaitu tingkat kesukaran, daya pembeda dan nilai *guessing* yang disajikan dalam soal pilihan ganda.

3. Menelaah Soal Tes

Soal yang telah dibuat tidak langsung diuji cobakan. Perlunya menelaah soal tes untuk memperbaiki soal jika didapati kekurangan atau kesalahan dalam pembuatannya. Walaupun soal yang dibuat telah dipersiapkan dengan baik, kekurangan dan kesalahan pembuatan soal mungkin saja terjadi selama dalam proses pembuatan berlangsung. Proses ini disebut sebagai proses validasi. Proses validasi dilakukan untuk menghasilkan kualitas soal yang baik. Telaah soal dilakukan oleh pakar di bidangnya yang disebut sebagai validasi ahli. Hal yang diperhatikan dalam penelaahan soal adalah kesesuaian soal dengan indikator, kesesuaian pilihan jawaban dengan atribut serta efektivitas dan daya tarik pengecoh masing-masing soal². Tidak hanya dalam kualitas soal yang akan divalidasi, dalam segi bahasa juga akan ditelaah.

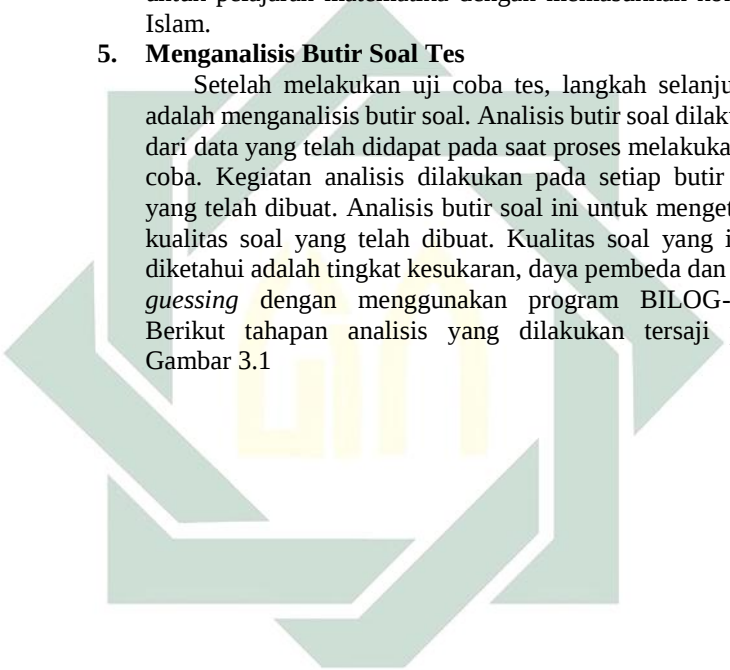
²Kusaeri, K. (2012). *Pengembangan Tes Diagnostik Dengan Menggunakan Model Dina Untuk Mendapatkan Informasi Salah Konsepsi Dalam Aljabar* (Doctoral dissertation, Universitas Negeri Yogyakarta), 112.

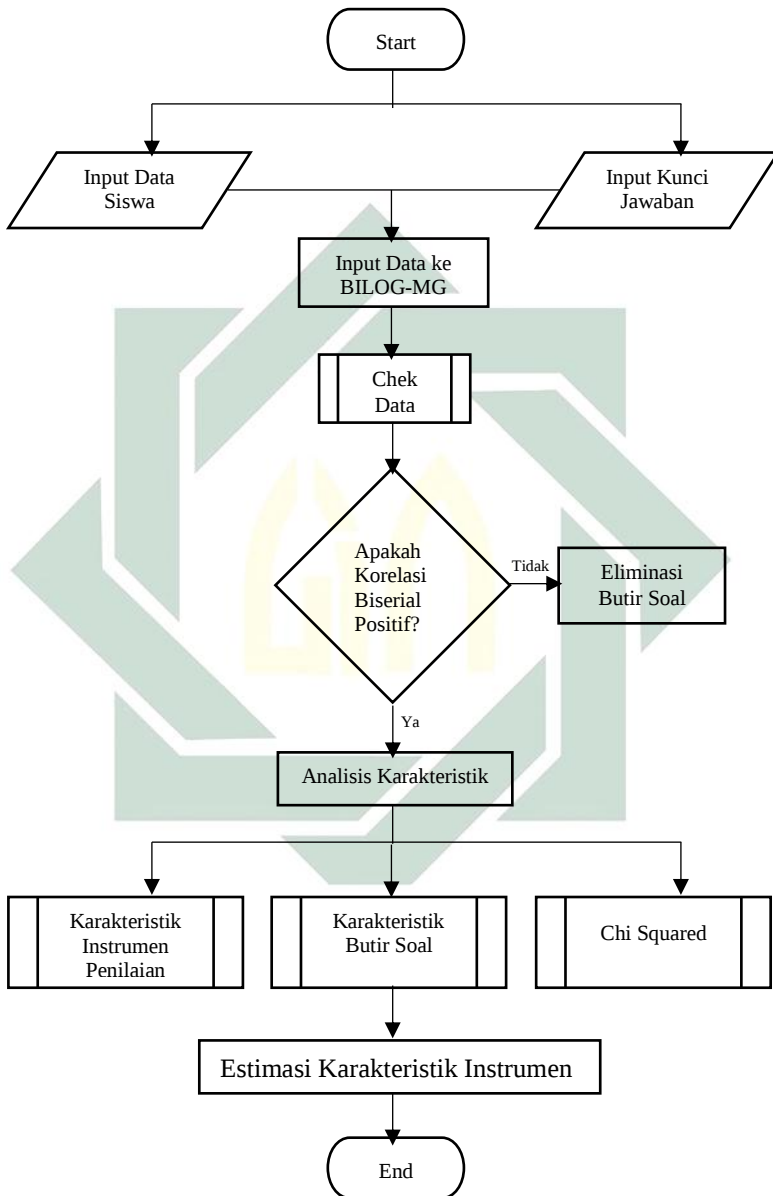
4. Melakukan Uji Coba

Uji coba sangat penting dilakukan sebelum soal digunakan dalam tes yang sesungguhnya. Uji coba digunakan sebagai sarana memperoleh data terkait dengan tujuan dari penelitian. Tujuannya adalah untuk mengetahui karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam.

5. Menganalisis Butir Soal Tes

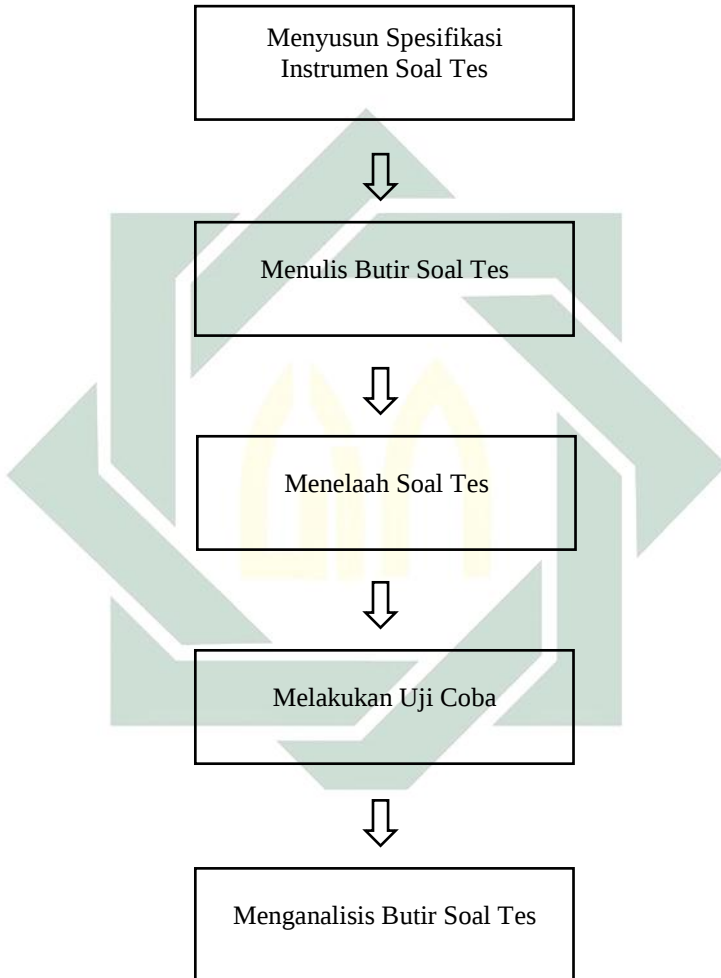
Setelah melakukan uji coba tes, langkah selanjutnya adalah menganalisis butir soal. Analisis butir soal dilakukan dari data yang telah didapat pada saat proses melakukan uji coba. Kegiatan analisis dilakukan pada setiap butir soal yang telah dibuat. Analisis butir soal ini untuk mengetahui kualitas soal yang telah dibuat. Kualitas soal yang ingin diketahui adalah tingkat kesukaran, daya pembeda dan nilai *guessing* dengan menggunakan program BILOG-MG. Berikut tahapan analisis yang dilakukan tersaji pada Gambar 3.1





Gambar 3.1
Tahapan Analisis Karakteristik Instrumen Penilaian

Lebih ringkasnya, tahapan-tahapan model pengembangan tersebut disajikan dalam Gambar 3.2



Gambar 3.2
Model Pengembangan Mardapi

C. Uji Coba Produk

Produk yang dihasilkan dalam penelitian ini adalah *prototype* instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam. Instrumen penilaian yang digunakan adalah soal pilihan ganda. Kegiatan uji coba produk sangat penting dilakukan. Tujuan pelaksanaan uji coba adalah untuk memperoleh data empirik. Data empirik digunakan untuk menguji apakah produk yang dikembangkan sudah memenuhi karakteristik butir soal yang baik.

1. Desain Uji Coba

Desain uji coba dilakukan melalui beberapa tahapan. Pada tahapan menelaah instrumen, rumusan dimensi dan indikator direview oleh pembimbing dan validator ahli dalam bidangnya. Peneliti meminta pendapat dari para ahli untuk melihat kesesuaian soal dengan indikator yang telah dibuat hingga menjadi butir soal. Setelah divalidasi oleh para ahli dalam bidangnya, rumusan indikator disusun dalam draft instrumen kemudian diuji cobakan pada subjek yang telah ditentukan.

2. Subjek Uji Coba

Subjek uji coba dalam penelitian ini adalah siswa tingkat Sekolah Menengah Pertama kelas IX. Pemilihan subjek uji coba, berdasarkan pada kebutuhan pengembangan. Sekolah yang digunakan sebagai subjek uji coba adalah sekolah Madrasah Tsanawiyah Negeri (MTsN). Ada 4 sekolah yang akan digunakan dalam penelitian ini, meliputi MTsN 1 Gresik, MTsN 1 Sidoarjo, MTsN 2 Surabaya, dan MTsN 3 Surabaya. secara detail dapat dilihat pada Tabel 3.1.

Tabel 3.1
Subjek Uji Coba Instrumen HOT

No.	Sekolah	Kelas	Banyaknya Siswa
1.	MTsN Gresik	9C	31
		9A	33
2.	MTsN 1 Sidoarjo	9A	29
3.	MTsN 2 Surabaya	9A	36
4.	MTsN 3 Surabaya	9A	36

3. Jenis Data

Jenis data yang dimaksud dalam penelitian ini adalah data lembar validasi dan hasil tes instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan untuk mengetahui daya pembeda, tingkat kesukaran, dan nilai *guessing*.

4. Instrumen Pengumpulan Data

Instrumen pengumpulan data adalah alat yang digunakan untuk memperoleh data. Data yang diperlukan dalam penelitian ini adalah data lembar validasi dan hasil tes instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam.

Lembar validasi ditunjukkan untuk menggali kualitas instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan dari validator. Instrumen divalidasi oleh pembimbing dan para ahli dalam bidangnya. Instrumen divalidasi oleh empat validator, berikut nama-nama validator instrumen yang tersaji pada Tabel 3.2.

Tabel 3.2

**Daftar Validator Instrumen Penilaian
Higher Order Thinking untuk Pelajaran Matematika
dengan Memasukkan Konteks Islam
yang Dikembangkan**

Validator	Jabatan
Tatik Dwi Utami, S. Pd.	Guru Matematika SMP Negeri 3 Sidoarjo
Ika Diah Yustifa, S. Pd.	Guru Matematika MTsN Tlasih
Dara Nur Indah	Guru Matematika SMP Negeri 5 Sidoarjo
Yuli Suhandono, M. Pd.	Dosen Pendidikan Matematika UIN Sunan Ampel Surabaya

Sedangkan hasil tes instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan

konteks Islam digunakan untuk mengetahui dan mengukur karakteristik instrumen penilaian berupa daya pembeda, tingkat kesukaran, dan nilai *guessing*. Data hasil tes didapatkan pada saat uji coba instrumen penilaian.

5. Teknik Analisis Data

Teknik analisis data dalam penelitian ini yaitu dengan menganalisis hasil tes instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan. Uraian singkat tentang teknik analisis beserta kriteria yang menjadi acuan hasil analisis masing-masing jenis data sebagai berikut:

a. Analisis Hasil Tes

Analisis hasil tes yang dilihat adalah analisis karakteristik instrumen penilaian. Analisis karakteristik instrumen penilaian meliputi daya pembeda, tingkat kesukaran, dan nilai *guessing* berdasarkan teori respon butir model tiga parameter logistik (3PL) dilakukan dengan menggunakan bantuan program BILOG-MG. Berikut persamaan yang diberikan terkait dengan model tiga parameter logistik:³

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

Keterangan:

- | | | |
|---------------|---|---|
| $P_i(\theta)$ | = | Peluang bahwa peserta tes dengan kemampuan menjawab butir soal ke-i |
| a_i | = | Parameter daya pembeda soal butir ke-i |
| b_i | = | Parameter tingkat kesukaran, yaitu satu titik pada skala <i>ability</i> dimana kemungkinan untuk menjawab benar sebesar 0,5 |
| c_i | = | Peluang tebakan benar butir ke-i |
| θ | = | Parameter kemampuan peserta tes |
| D | = | Faktor penskalaan yang diikutkan untuk menjadikan fungsi logistic |

³Hambleton, et.al., *Fundamentals of Item Response Theory* (California: Sage Publications, The International Profesional Publishers, 1991).

serupa mungkin dengan fungsi ogive norma (D=1,702)

Pengaplikasian dari teori respon butir disebut sebagai pendugaan parameter⁴. Ada beberapa metode yang digunakan untuk menduga parameter pada model teori respon butir dengan salah satunya adalah metode kemungkinan maksimum (*Maximum Likelihood Estimation*, MLE). Prinsip dasar metode MLE adalah jika terdapat contoh acak $X_1, X_2, \dots, X_n$ dari sebuah sebaran yang memiliki suatu fungsi kepekatan peluang $f(x; \Theta)$, $\Theta \in \Omega$. Fungsi kepekatan peluang bersama $X_1, X_2, \dots, X_n$ adalah $f(x_1; \Theta), f(x_2; \Theta), \dots, f(x_n; \Theta)$. Fungsi kepekatan peluang bersama ini dipandang sebagai fungsi dari θ . Dengan definisi yang telah dijelaskan maka fungsi nisbah kemungkinan (L) dapat dinotasikan sebagai berikut:⁵

$$L(\theta; x_1, x_2, \dots, x_n) = f(x_1; \theta), f(x_2; \theta), \dots, f(x_n; \theta)$$

Hambelton et al mendefinisikan fungsi likelihood untuk model teori respon butir sebagai berikut:⁶

$$L(x_1, x_2, \dots, x_n | \theta) = \prod_{j=1}^N \prod_{i=1}^n P_{ij}^{x_{ij}} Q_{ij}^{1-x_{ij}}$$

$$l(X | \theta) = \ln L(x_1, x_2, \dots, x_n | \theta)$$

$$l(X | \theta) = \ln L \left(\prod_{j=1}^N \prod_{i=1}^n P_{ij}^{x_{ij}} Q_{ij}^{1-x_{ij}} \right)$$

$$l(X | \theta) = \sum_{j=1}^N \sum_{i=1}^n (x_{ij} \ln P_{ij} + (1 - x_{ij}) \ln Q_{ij})$$

$$\text{dengan } \frac{\partial l}{\partial \theta} = 0, \theta = (\theta; a, b, c)$$

⁴Ibid, halaman

⁵Hong, et.al., *Introduction to Mathematical Statistics 4th Ed* (New York: Macmillan Publishing, 1978).

⁶Hambleton, et.al., *Fundamentals of Item Response Theory* (California: Sage Publications, The International Profesional Publishers, 1991).

Dimana a , b , dan c adalah parameter butir soal, dan θ adalah parameter kemampuan (model IRT 3PL). Berikut merupakan estimasi dari setiap parameter butir soal:

1) Daya Pembeda

Daya pembeda disimbolkan dengan parameter “ a ” atau nama lainnya adalah *discrimination parameters*. Secara teoritis daya pembeda soal terlatak pada skala $-\infty \leq a \leq \infty$. Dalam prakteknya, Hambleton memberikan kisaran antara $0 \leq a \leq 2^7$.

2) Tingkat Kesukaran

Tingkat kesukaran disimbolkan dengan parameter “ b ” atau nama lainnya adalah *difficulty parameters*. Secara teoritis tingkat kesukaran soal terlatak pada skala $-\infty \leq b \leq \infty$. Dalam prakteknya, Hambleton memberikan kisaran antara $-2 \leq b \leq 2^8$.

3) Nilai *Guessing*

Guessing disimbolkan dengan parameter “ c ” atau nama lainnya adalah *pseudo-guessing*. *Guessing* mempunyai kisaran antara $0 \leq c \leq 1$. Suatu butir soal dikatakan baik jika nilai c tidak lebih dari $1/k$, dengan k banyaknya pilihan jawaban⁹.

⁷Ibid, halaman 15

⁸Ibid, halaman 13

⁹Frank B. Baker, *The Basics of Item Response Theory* (United States of Amerika: ERIC Clearinghouse on Assessment and Evaluation, 2001), 28.

BAB IV

HASIL PENELITIAN

A. Penyajian dan Analisis Data Pengembangan Instrumen Penilaian *Higher Order Thinking* untuk Pelajaran Matematika dengan Memasukkan Konteks Islam

Penelitian ini merupakan penelitian pengembangan yang bertujuan untuk menghasilkan suatu produk berupa *prototype* instrumen penilaian *higher order thinking* untuk pelajaran matematika dengan memasukkan konteks Islam. Dalam mencapai tujuan tersebut, penelitian ini mengadopsi pada model pengembangan Mardapi dengan membatasi sampai langkah ke lima yaitu (1) menyusun spesifikasi instrumen soal tes, (2) menulis butir soal tes, (3) menelaah soal tes, (4) melakukan uji coba, dan (5) menganalisis butir soal tes. Deskripsi dari setiap langkah pengembangan instrumen penilaian, dijelaskan sebagai berikut:

1. Data Spesifikasi Instrumen Soal Tes

Langkah awal yang dilakukan dalam penelitian ini merupakan penyusunan spesifikasi instrumen soal tes. Langkah ini dilakukan untuk mengetahui kebutuhan yang diperlukan dalam proses pengembangan instrumen penilaian *higher order thinking* untuk pelajaran matematika dengan memasukkan konteks Islam. Menyusun spesifikasi instrumen soal tes meliputi:

a. Menentukan Tujuan Tes

Tes ini bertujuan untuk mengukur kemampuan HOT siswa pada pelajaran matematika. Tingkat keberhasilan dapat dinyatakan oleh hasil nilai yang diperoleh siswa setelah melakukan tes tersebut.

b. Menyusun Kisi-kisi Tes

Menyusun kisi-kisi tes menjadi langkah yang sangat penting dalam proses pengembangan. Hal yang utama diperhatikan adalah menyusun kisi-kisi tes berdasarkan Kompetensi Dasar (KD) Kurikulum 2013 Revisi 2017 dengan indikator dari kajian mengenai *Higher Order Thinking* (HOT).

c. Memilih Bentuk Tes

Bentuk tes yang digunakan dalam penelitian pengembangan ini merupakan pilihan ganda. Pemilihan bentuk tes pilihan ganda berdasarkan kajian teori pada bab kedua.

Dari proses penyusunan diatas, terbentuklah spesifikasi penyusunan instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang disajikan pada Lampiran 1.

2. Menulis Butir Soal Tes

Peneliti membuat dan mengembangkan instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam. Instrumen yang dibuat secara keseluruhan mencakup mata pelajaran matematika jenjang Sekolah Menengah Pertama kelas VII dan VIII. Tidak hanya itu, instrumen dibuat berdasarkan Kompetensi Dasar (KD) Kurikulum 2013 Revisi 2017. Berbantu pada kisi-kisi yang telah dibuat, sebanyak 10 butir soal pilihan ganda telah dihasilkan yang tercantum pada Lampiran 2 beserta pembahasan pada Lampiran 3.

3. Menelaah Soal Tes

Instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang telah dibuat, perlu ditelaah. Telaah ini disebut proses validasi yang dilakukan oleh para ahli. Terdapat 4 validator dengan rincian 3 guru matematika dan 1 dosen yang menjadi penelaah dari instrumen yang telah dibuat. Berikut nama validator yang terpilih dalam penelitian ini:

Tabel 4.1
Daftar Nama Validator Instrumen Penilaian

No.	Nama Validator	Keterangan	Tanggal Validasi
1.	Tatik Dwi Utami, S. Pd.	Guru Matematika SMP Negeri 3 Sidoarjo	02 November 2017
2.	Ika Diah Yustifa, S. Pd.	Guru Matematika MTsN Tlasi	06 November 2017
3.	Dara Nur Indah	Guru Matematika SMP Negeri 5 Sidoarjo	06 November 2017
4.	Yuli Suhandono, M. Pd.	Dosen Pendidikan Matematika UIN Sunan Ampel Surabaya	17 November 2017

Hasil dari telaah yang dilakukan oleh para validator, menunjukkan bahwa instrumen yang telah dibuat peneliti masih belum baik. Berikut merupakan hasil revisi dari para validator:

a. Validator 1

Penelaahan instrumen yang pertama, dilakukan oleh ahli/pakar pada pelajaran matematika. Ahli/pakar pada pelajaran matematika yang dimaksudkan disini adalah guru berpengalaman yang sudah lama berada di lingkungan sekolah dan mengetahui karakteristik dari siswanya. Tatik Dwi Utami merupakan guru matematika di SMP Negeri 3 Sidoarjo.

Terkait dengan aspek penilaian yang diberikan validator 1 terhadap instrumen penilaian yang dikembangkan, memiliki penilaian berbeda dari setiap butir soal. Merujuk pada Lampiran 4, aspek penilaian yang diberikan pada butir soal nomor 1 sampai 4 bernilai 95. Dengan demikian butir soal nomor 1 sampai 4 dalam kategori A, artinya soal dapat digunakan tanpa revisi. Untuk butir soal nomor 5 sampai 7 memiliki nilai 81

dalam kategori B, artinya soal dapat digunakan dengan sedikit revisi. Sedangkan butir soal nomor 8 sampai 10, memiliki nilai 85 dalam kategori B yang artinya soal dapat digunakan dengan sedikit revisi. Dari hasil keseluruhan, instrumen penilaian yang dikembangkan memiliki total nilai 87.8 dalam kategori A, artinya instrumen penilaian yang dikembangkan dapat digunakan tanpa revisi.

Setelah dilakukan penilaian terhadap instrumen penilaian yang dikembangkan, terdapat masukan dan beberapa saran yang disampaikan oleh validator 1. Masukan dan saran yang diberikan meliputi: 1) Sebaiknya soal diberi waktu beberapa menit untuk mengerjakan, dan 2) Indikator soal sebaiknya memancing ke soal jangan terlalu luas. Dari masukan dan saran yang diberikan, terdapat revisi pada bagian spesifikasi pengembangan instrumen. Pada bagian spesifikasi pengembangan instrumen, diberikan informasi terkait waktu pengerjaan soal. Waktu yang diberikan untuk mengerjakan adalah 2×40 menit.

b. Validator 2

Penelaahan instrumen yang kedua, dilakukan oleh ahli/pakar pada pelajaran matematika. Ahli/pakar pada pelajaran matematika yang dimaksudkan disini adalah guru berpengalaman yang sudah lama berada di lingkungan sekolah dan mengetahui karakteristik dari siswanya. Berbeda dari validator pertama, pada validator kedua lebih ditekankan pengetahuan terkait keislaman yang diajarkan pada jenjang MTs yang disesuaikan dengan soal matematika. Dari hal tersebut, penelaah yang kedua dilakukan disalah satu sekolah Madrasah Tsanawiyah. MTs Negeri Tlasi menjadi tempat validasi instrumen HOT dengan validatornya adalah Ika Diah Yusfita yang merupakan guru matematika.

Terkait dengan aspek penilaian yang diberikan validator 2 terhadap instrumen penilaian yang dikembangkan, memiliki penilaian yang sama untuk setiap butir soal. Merujuk pada Lampiran 4, aspek penilaian yang diberikan pada butir soal nomor 1 sampai 10 dalam kategori B dengan nilai untuk butir soal nomor 6 adalah 72, butir soal nomor 5 adalah 78.7, butir soal nomor 3 dan 4 adalah 74.7, butir soal nomor 1 adalah 77.3, butir soal nomor 2, 5, dan 9 adalah 78.7, butir soal nomor 8 dan 10 adalah 80. Artinya, soal dapat digunakan dengan sedikit

revisi. Dari hasil keseluruhan, instrumen penilaian yang dikembangkan memiliki total nilai 76.8 dalam kategori B, artinya instrumen penilaian yang dikembangkan dapat digunakan dengan sedikit revisi.

Setelah dilakukan penilaian terhadap instrumen penilaian yang dikembangkan, terdapat masukan dan beberapa saran yang disampaikan oleh validator 2. Masukan dan saran yang diberikan meliputi: 1) Setiap kompetensi dasar, dibuat beberapa indikator soal, dan 2) Membuat indikator soal sebaiknya mengarah ke pertanyaan soal. Dari masukan dan saran yang diberikan, terdapat revisi pada bagian spesifikasi pengembangan instrumen. Pada bagian spesifikasi pengembangan instrumen, dipetakan untuk setiap kompetensi dasar dari beberapa indikator soal yang dibuat. Awalnya dari peneliti, jika terdapat indikator *Higher Order Thinking* (HOT) yang sama maka peneliti membuatnya dalam satu kolom. Dari masukan dan saran yang diberikan, untuk setiap indikator HOT dibuat secara terpisah.

c. Validator 3

Penelaahan instrumen yang ketiga, sedikit berbeda dengan sebelumnya. Validator ketiga melengkapi dari validator sebelumnya. Peninjauan instrumen yang dilakukan oleh validator ketiga dilihat dari penyesuaian karakteristik soal dengan siswa dan pengetahuan terkait keislaman yang diajarkan pada jenjang MTs yang disesuaikan dengan soal matematika. Dara Nur Indah merupakan guru matematika di SMP Negeri 5 Sidoarjo yang menjadi validator ketiga.

Terkait dengan aspek penilaian yang diberikan validator 3 terhadap instrumen penilaian yang dikembangkan, memiliki penilaian yang sama untuk setiap butir soal. Merujuk pada Lampiran 4, aspek penilaian yang diberikan pada butir soal nomor 1 sampai 10 dalam kategori B dengan nilai untuk butir soal nomor 5 adalah 72, butir soal nomor 2 dan 7 adalah 73, butir soal nomor 10 adalah 75, butir soal nomor 1 dan 3 adalah 76, butir soal nomor 6 dan 9 adalah 77, butir soal nomor 4 adalah 79, butir soal nomor 8 adalah 81. Artinya, soal dapat digunakan dengan sedikit revisi. Dari hasil keseluruhan, instrumen penilaian yang dikembangkan memiliki total nilai 76 dalam kategori B, artinya instrumen penilaian yang dikembangkan dapat digunakan dengan sedikit revisi.

Setelah dilakukan penilaian terhadap instrumen penilaian yang dikembangkan, terdapat masukan dan beberapa saran yang disampaikan oleh validator 3. Masukan dan saran yang diberikan meliputi: 1) Kompetensi Dasar sebaiknya menggunakan KD. 3 (pengetahuan, karena alat ukurnya (soal) merupakan pilihan ganda, 2) Soal no. 1 tidak ada jawaban yang benar. Jika C menjadi pilihan jawaban benar maka 215 diganti 235, 3) Soal nomor 2 option jawaban A dan C diperbaiki, sehingga tidak ada jawaban ganda, 4) Soal no 3, supaya tercapai apa yang dimaksud bisa diperbaiki pada kecepatan yang digunakan untuk alat transportasi Rahmad, misal kurang dari 400 km/jam, dan 5) Jika waktu yang diminta 2×40 menit, maka soal yang diberikan sebanyak 8 menyesuaikan dengan waktu yang diberikan.

Dari masukan dan saran yang diberikan, terdapat revisi pada bagian spesifikasi pengembangan instrumen. Pada bagian spesifikasi pengembangan instrumen, Kompetensi Dasar yang dipakai peneliti adalah aspek keterampilan (KD.4). Akan tetapi, soal yang digunakan dalam penelitian berupa soal pilihan ganda. Dari masukan yang diberikan, peneliti mengganti dari Kompetensi Dasar aspek keterampilan (KD.4) dengan Kompetensi Dasar aspek pengetahuan (KD.3). Terkait dengan instrumen penilaian yang dikembangkan, dari 10 butir soal yang dibuat adanya eliminasi 2 butir soal agar dapat lebih efisien dalam waktu pengerjaan yang diberikan.

Dari masukan dan saran yang diberikan, butir soal nomor 2 dan 7 dikeluarkan dari instrumen penilaian. Di lain pihak pada option jawaban yang diberikan, peneliti menggunakan huruf kecil untuk setiap option jawaban. Saran dan masukan yang diberikan diganti dengan menggunakan huruf kapital. Adapun beberapa masukan dan saran yang diberikan untuk setiap butir soalnya, yang tersaji dengan hasil revisi:

1) Butir soal nomor 1

Redaksi awal jawaban butir soal nomor 1

“c. Pada saat pemberangkatan, ternyata di setiap gerbong tidak terisi penuh. Setiap gerbong terdapat kekosongan penumpang. Kekosongan di setiap gerbongnya membentuk pola kelipatan 4. Saat tiba distasiun Kota Sidoarjo, hampir setiap gerbong terisi penuh dan hanya ada satu gerbong yang berisi 40 dari kapasitas yang ada. Setibanya di Kota

malang, sebanyak 45 jumlah pemudik yang turun. Tanpa ada penambahan jumlah penumpang di Kota Malang, sebanyak 215 jumlah pemudik yang akan turun di stasiun terakhir”

Redaksi jawaban butir soal nomor 1 direvisi menjadi

“C. Pada saat pemberangkatan, ternyata di setiap gerbong tidak terisi penuh. Setiap gerbong terdapat kekosongan penumpang. Kekosongan di setiap gerbongnya membentuk pola kelipatan 4. Saat tiba distasiun Kota Sidoarjo, hampir setiap gerbong terisi penuh dan hanya ada satu gerbong yang berisi 40 dari kapasitas yang ada. Setibanya di Kota malang, sebanyak 45 jumlah pemudik yang turun. Tanpa ada penambahan jumlah penumpang di Kota Malang, sebanyak 235 jumlah pemudik yang akan turun di stasiun terakhir”

2) Butir Soal Nomor 3

Redaksi awal butir soal nomor 3

“Rahmad, Ali dan Aisah bersahabat sejak kecil. Setelah lulus SMA, mereka berpisah karena mereka masuk perguruan tinggi yang berbeda-beda. Pada saat liburan, mereka merencanakan bertemu. Rahmad berangkat pukul 17.00, Ali berangkat selisih 15 menit dari Rahmad dan Aisah berangkat selisih 15 menit dari Ali. Mereka janjian dikota D pukul 19.30. Akan tetapi, hanya Rahmad yang tiba lebih awal pukul 18.15. Transportasi yang mereka gunakan berbeda-beda. Rahmad menggunakan transportasi yang mempunyai kecepatan 400km/jam, Ali menggunakan transportasi yang mempunyai kecepatan 80km/jam dan Aisah menggunakan transportasi yang mempunyai kecepatan 40km/jam. Pernyataan yang benar berkaitan dengan diatas adalah . . .”

Redaksi butir soal nomor 3 direvisi menjadi

“Rahmad, Ali dan Aisah bersahabat sejak kecil. Setelah lulus SMA, mereka berpisah karena mereka masuk perguruan tinggi yang berbeda-beda. Pada saat liburan, mereka merencanakan bertemu. Rahmad berangkat pukul 17.00, Ali berangkat selisih 15 menit dari Rahmad dan Aisah berangkat selisih 15 menit

dari Ali. Mereka janjian dikota D pukul 19.30. Akan tetapi, hanya Rahmad yang tiba lebih awal pukul 19.00. Transportasi yang mereka gunakan berbeda-beda. Rahmad menggunakan transportasi yang mempunyai kecepatan 300km/jam, Ali menggunakan transportasi yang mempunyai kecepatan 80km/jam dan Aisah menggunakan transportasi yang mempunyai kecepatan 40km/jam. Jika waktu sholat Isya' memasuki pukul 18.45, maka pernyataan yang benar berkaitan dengan kondiai diatas adalah . . .”

3) Butir Soal Nomor 4

Redaksi Awal butir soal nomor 4

“Adinda, Fitri, Siti dan Aisah merupakan santriwati di ponpes Darul ‘Ulum Jombang. Mereka juga sekolah disalah satu Madrasah Tsanawiyah dipondok tersebut. Untuk menyelesaikan studi mereka, mereka harus dapat menghafal juz 30. Tes hafalan dilaksanakan satu minggu setelah Ujian Nasional. Tiga minggu sebelum tes hafalan berlangsung, jumlah surat yang Adinda hafal 5 kali banyaknya surat yang dihafal fitri. Jumlah surat yang dihafal fitri 2 kali lebih banyak dari Siti. Sedangkan jumlah surat yang dihafaal Siti tiga lebihnya dari jumlah surat Aisah. Model matematika yang dapat dibuat dari pernyataan-pernyataan diatas adalah . . .”

Redaksi butir soal nomor 4 direvisi menjadi

“Adinda, Fitri, Siti dan Aisah merupakan santriwati di ponpes Darul ‘Ulum Jombang. Mereka juga sekolah disalah satu Madrasah Tsanawiyah dipondok tersebut. Untuk menyelesaikan studi mereka, mereka harus dapat menghafal juz 30. Tes hafalan dilaksanakan satu minggu setelah Ujian Nasional. Tiga minggu sebelum tes hafalan berlangsung, jumlah surat yang Adinda hafal 5 kali banyaknya surat yang dihafal fitri. Jumlah surat yang dihafal fitri 2 kali lebih banyak dari Siti. Sedangkan jumlah surat yang dihafaal Siti tiga lebihnya dari jumlah surat Aisah. Pernyataan yang benar berkaitan dengan kondisi diatas adalah. . .”

4) Butir Soal Nomor 5

Redaksi Awal butir soal nomor 5

- “ (i) Sudut yang dibentuk oleh pertemuan pangkal dua buah garis lurus
yang saling tegak lurus merupakan gerakan ruku' dalam sholat.
- (ii) Sudut yang kaki-kaki sudutnya tidak saling tegak lurus dan jarak kaki sudutnya lebih kecil dari jarak kaki-kaki sudut pada sudut siku-siku merupakan gerakan sujud dalam sholat.
- (iii) Jika dalam satu rakaat gerakan shalat Ruku', I'tidal dan sujud dijumlahkan, maka sudut yang dihasilkan kurang dari 360^0 .
- (iv) Tidak ada kaitannya gerakan sholat dengan materi sudut dalam matematika.”

Redaksi butir soal nomor 5 direvisi menjadi

- “ (i) Sudut yang dibentuk oleh pertemuan pangkal dua buah garis lurus yang saling tegak lurus merupakan gerakan Ruku' dalam sholat.
- (ii) Sudut 90^0 yang membagi dua sama besar merupakan gerakan Sujud dalam sholat.
- (iii) Jika dalam satu rakaat gerakan shalat Ruku', I'tidal dan Sujud dijumlahkan, maka sudut yang dihasilkan kurang dari 360^0 .
- (iv) Sudut yang dihasilkan dalam satu rakaat gerakan shalat Sujud merupakan sudut yang sama dalam gerakan shalat I'tidal.”

5) Butir Soal Nomor 6

Redaksi Awal butir soal nomor 6

“Hari ini Bu Nisa' membagikan hasil tes Hafalan Al-Qur'an Juz 30 kepada siswanya. Ahmad mendapatkan nilai 2 kali lebih banyak dari pada Kevin. Sedangkan Aini mendapatkan nilai satu lebih banyak dari Ahmad. Fitri sahabat Ani, mendapatkan dua lebih sedikit dari sahabatnya. Ali tidak mengikuti tes

hafalan ulang, karena nilainya sama seperti nilai standart kelulusan Ujian Pondok Pesantren. Faizien hampir mendapatkan nilai sempurna, akan tetapi terdapat hafalan yang kurang lancar pada awal surat juz 30. Sehingga Faizien mendapatkan satu lebih sedikit dari nilai sempurna. Jika nilai yang didapatkan Kevin pada saat tes hafalan Al-Qur'an bernilai 4 dan nilai standart kelulusan Pondok Pesantren adalah adalah tiga lebih banyak dari nilai Kevin, tentukan modus dari nilai hasil tes hafalan Al-Qur'an Juz 30 yang dibagikan Bu Nisa' kepada siswanya adalah. ..."

Redaksi butir soal nomor 6 direvisi menjadi

"Hari ini Bu Nisa' membagikan hasil tes Hafalan Al-Qur'an Juz 30 kepada siswanya yang berjumlah 6 orang. Ahmad mendapatkan nilai 2 kali lebih banyak dari pada Kevin. Sedangkan Aini mendapatkan nilai satu lebih banyak dari Ahmad. Fitri sahabat Aini, mendapatkan dua kurang dari sahabatnya. Ali tidak mengikuti tes hafalan ulang, karena nilainya sama seperti nilai standart kelulusan Ujian Pondok Pesantren. Faizien hampir mendapatkan nilai sempurna, akan tetapi terdapat hafalan yang kurang lancar pada awal surat juz 30. Sehingga Faizien mendapatkan satu lebih sedikit dari nilai sempurna. Nilai yang didapatkan Kevin pada saat tes hafalan Al-Qur'an bernilai 4 dan nilai standart kelulusan Pondok Pesantren adalah adalah tiga lebih banyak dari nilai Kevin. Modus dari nilai hasil tes hafalan Al-Qur'an Juz 30 yang dibagikan Bu Nisa' kepada siswanya adalah. ..."

6) Butir Soal Nomor 8

Redaksi Awal butir soal nomor 8

"Ahmad, Alawi dan Ulfa masing-masing mempunyai uang sebanyak Rp. 600.000,00, Rp. 650.000,00 dan Rp. 700.000,00. Pada suatu hari, ketiga sahabat tersebut melakukan musahammah (investasi) kepada perusahaan ternama diketanya. Komisi yang diberikan oleh perusahaan sebanyak 10% per bulan. Ketiga sahabat bernadzar, uang hasil nisbah (komisi) yang didapatkan dalam satu tahun akan disumbangkan kepada masjid di desanya. Ahmad menyumbang sebanyak 5%, alawi sebanyak 15%, dan ulfa sebanyak 10% dari komisi yang

didapatkannya selama 1 tahun. Pernyataan yang benar berkaitan dengan diatas adalah. . .”

Redaksi butir soal nomor 8 direvisi menjadi

“Ahmad, Alawi dan Ulfa masing-masing mempunyai uang sebanyak Rp.600.000,00, Rp.650.000,00 dan Rp.700.000,00. Pada suatu hari, ketiga sahabat tersebut melakukan musahammah (investasi) kepada perusahaan ternama dikotanya. Komisi yang diberikan oleh perusahaan sebanyak 10% per bulan. Ketiga sahabat bernadzar, uang hasil nisbah (komisi) yang didapatkan dalam satu tahun akan disumbangkan kepada masjid di desanya. Ahmad menyumbang sebanyak 5%, alawi sebanyak 15%, dan ulfa sebanyak 10% dari komisi yang didapatkannya selama 1 tahun. Pernyataan yang benar berkaitan dengan kondisi diatas adalah. . .”

Redaksi Awal jawaban butir soal nomor 8

“c. Sisa uang Alawi lebih sedikit daripada Ulda dan lebih banyak daripada Ahmad”

Redaksi jawaban butir soal nomor 8 direvisi menjadi

“C. Sisa uang Alawi lebih sedikit daripada Ulfa dan lebih banyak daripada Ahmad”

7) Butir Soal Nomor 9

Redaksi Awal butir soal nomor 9

“Lilah mendapatkan kesempatan untuk menyempurnakan rukun islamnya dengan menuaikan ibadah Haji. Kesempatan ini didapatkannya setelah dia menabung selama bertahun-tahun. Diapun langsung mendaftarkan dirinya ke Kementrian Agama. Setelah melakukan pendaftaran dan adiministrasi, petugas pendaftaran mengatakan kepada Lilah bahwa “Dalam dua puluh tahun kedepan, peluang keberangkatan Haji adalah lima per enam.” Manakah di bawah ini yang paling mencerminkan maksud dari petugas pendaftaran . . .”

Redaksi butir soal nomor 9 direvisi menjadi

“Lilah mendapatkan kesempatan untuk menyempurnakan rukun islamnya dengan menuaikan ibadah Haji. Kesempatan ini

didapatkannya setelah dia menabung selama bertahun-tahun. Diapun langsung mendaftarkan dirinya ke Kementerian Agama. Setelah melakukan pendaftaran dan administrasi, petugas pendaftaran mengatakan kepada Lilah bahwa “Dalam dua puluh tahun kedepan, peluang keberangkatan Haji adalah lima per enam.” Pernyataan di bawah ini yang paling mencerminkan maksud dari petugas pendaftaran adalah . . .”

8) Butir Soal Nomor 10

Redaksi Awal butir soal nomor 10

“Masjid Al-Akbar Surabaya mengadakan kegiatan buka bersama dengan yatim piatu dan dhuafa. Acara tersebut diikuti lebih dari 1000 peserta. Panitia kegiatan mengatur tempat duduk sedemikianrupa diamana barisan pertama terdapat 2 kursi. Barisan kedua dan ketiga terdiri dari 6 dan 12 kursi. Barisan keempat terdiri dari 20 kursi. Ahmad, Imron dan Firman merupakan salah satu peserta kegiatan buka bersama. Ahmad berada di barisan ke sepuluh. Imron berada di barisan ke dua belas. Sedangkan Firman berada di barisan ke empat belas. Berapakah jumlah kursi dari barisan ke sepuluh, dua belas dan empat belas pada kegiatan tersebut...”

Redaksi butir soal nomor 10 direvisi menjadi

“Masjid Al-Akbar Surabaya mengadakan kegiatan buka bersama dengan yatim piatu dan dhuafa. Acara tersebut diikuti lebih dari 1000 peserta. Panitia kegiatan mengatur tempat duduk sedemikianrupa diamana tempat duduk yang diatur membentuk suatu pola barisan. Barisan pertama terdapat 2 kursi. Barisan kedua dan ketiga terdiri dari 6 dan 12 kursi. Barisan keempat terdiri dari 20 kursi. Ahmad, Imron dan Firman merupakan peserta kegiatan buka bersama. Ahmad berada di barisan ke sepuluh. Imron berada di barisan ke dua belas. Sedangkan Firman berada di barisan ke empat belas. Banyaknya kursi pada barisan ke sepuluh, dua belas dan empat belas pada kegiatan buka bersama adalah...”

d. Validator 4

Penelaahan instrumen yang keempat, dilakukan setelah hasil pertimbangan dari para validator sebelumnya. Diperlukan

penelaah, ahli/pakar pada pelajaran matematika yang pengalamannya lebih dari seorang guru. Validator ahli yang dimaksudkan adalah dosen Pendidikan Matematika Universitas Islam Negeri Sunan Ampel (UINSA) Surabaya. Yuli Suhandono, M. Pd. merupakan dosen dari Prodi Pendidikan Matematika UIN Sunan Ampel Surabaya menjadi validator keempat.

Terkait dengan aspek penilaian yang diberikan validator 4 terhadap instrumen penilaian yang dikembangkan, memiliki penilaian berbeda dari setiap butir soal. Merujuk pada Lampiran 4, aspek penilaian yang diberikan pada butir soal nomor 1 dan 7 dalam kategori B dengan nilai butir soal nomor 1 adalah 84 dan butir soal nomor 7 adalah 85. Artinya, soal dapat digunakan dengan sedikit revisi. Sedangkan butir soal nomor 2, 3, 4, 5, 6, dan 8 dalam kategori A dengan nilai butir soal nomor 2 sampai 6 adalah 87 dan butir soal nomor 8 adalah 88. Artinya, soal dapat digunakan tanpa revisi. Dari hasil keseluruhan, instrumen penilaian yang dikembangkan memiliki total nilai 86.3 dalam kategori A, artinya instrumen penilaian yang dikembangkan dapat digunakan tanpa revisi.

Setelah dilakukan penilaian terhadap instrumen penilaian yang dikembangkan, terdapat masukan dan beberapa saran yang disampaikan oleh validator 4. Masukan dan saran yang diberikan meliputi: 1) Soal dapat digunakan dengan sedikit revisi pada pertanyaan. Berikut merupakan beberapa masukan dan saran yang diberikan untuk setiap butir soalnya, yang tersaji dengan hasil revisi:

1) Butir soal nomor 1

Redaksi awal jawaban butir soal nomor 1

“C. Pada saat pemberangkatan, ternyata di setiap gerbong tidak terisi penuh. Setiap gerbong terdapat kekosongan penumpang. Kekosongan di setiap gerbongnya membentuk pola kelipatan 4. Saat tiba distasiun Kota Sidoarjo, hampir setiap gerbong terisi penuh dan hanya ada satu gerbong yang berisi 40 dari kapasitas yang ada. Setibanya di Kota Malang, sebanyak 45 jumlah pemudik yang turun. Tanpa ada penambahan jumlah penumpang di Kota Malang, sebanyak 235 jumlah pemudik yang akan turun di stasiun terakhir”

Redaksi jawaban butir soal nomor 1 direvisi menjadi

“C. Pada saat pemberangkatan, gerbong terisi penuh. Saat tiba distasiun Kota Sidoarjo, hampir setiap gerbong terisi penuh dan hanya ada satu gerbong yang berisi 40 dari kapasitas yang ada. Setibanya di Kota Malang, sebanyak 45 jumlah pemudik yang turun. Tanpa ada penambahan jumlah penumpang di Kota Malang, sebanyak 235 jumlah pemudik yang akan turun di stasiun terakhir”

2) Buti Soal Nomor 2

Redaksi awal butir soal nomor 2

“Rahmad, Ali dan Aisah bersahabat sejak kecil. Setelah lulus SMA, mereka berpisah karena mereka masuk perguruan tinggi yang berbeda-beda. Pada saat liburan, mereka merencanakan bertemu. Rahmad berangkat pukul 17.00, Ali berangkat selisih 15 menit dari Rahmad dan Aisah berangkat selisih 15 menit dari Ali. Mereka janji dikota D pukul 19.30. Akan tetapi, hanya Rahmad yang tiba lebih awal pukul 19.00. Transportasi yang mereka gunakan berbeda-beda. Rahmad menggunakan transportasi yang mempunyai kecepatan 300km/jam, Ali menggunakan transportasi yang mempunyai kecepatan 80km/jam dan Aisah menggunakan transportasi yang mempunyai kecepatan 40km/jam. Jika waktu sholat Isya’ memasuki pukul 18.45, maka pernyataan yang benar berkaitan dengan kondia di atas adalah . . .”

Redaksi butir soal nomor 2 direvisi menjadi

“Rahmad, Ali dan Aisah bersahabat sejak kecil. Setelah lulus SMA, mereka berpisah karena mereka masuk perguruan tinggi yang berbeda-beda. Pada saat liburan, mereka merencanakan bertemu. Rahmad berangkat pukul 17.00, Ali berangkat selisih 15 menit dari Rahmad dan Aisah berangkat selisih 15 menit dari Ali. Mereka janji dikota D pukul 19.30. Akan tetapi, hanya Rahmad yang tiba lebih awal pukul 19.00. Transportasi yang mereka gunakan berbeda-beda. Rahmad menggunakan transportasi yang mempunyai kecepatan 300km/jam, Ali menggunakan transportasi yang mempunyai kecepatan 80km/jam dan Aisah menggunakan transportasi yang

mempunyai kecepatan 40km/jam. Jika waktu sholat Isya' memasuki pukul 18.45, maka pernyataan yang benar berkaitan dengan syarat manjamak sholat adalah . . .”

3) Butir Soal Nomor 3

Redaksi awal butir soal nomor 3

“Adinda, Fitri, Siti dan Aisah merupakan santriwati di ponpes Darul ‘Ulum Jombang. Mereka juga sekolah disalah satu Madrasah Tsanawiyah dipondok tersebut. Untuk menyelesaikan studi mereka, mereka harus dapat menghafal juz 30. Tes hafalan dilaksanakan satu minggu setelah Ujian Nasional. Tiga minggu sebelum tes hafalan berlangsung, jumlah surat yang Adinda hafal 5 kali banyaknya surat yang dihafal fitri. Jumlah surat yang dihafal fitri 2 kali lebih banyak dari Siti. Sedangkan jumlah surat yang dihafal Siti tiga lebihnya dari jumlah surat Aisah. Model matematika yang dapat dibuat dari pernyataan-pernyataan diatas adalah . . .”

Redaksi butir soal nomor 3 direvisi menjadi

“Adinda, Fitri, Siti dan Aisah merupakan santriwati di ponpes Darul ‘Ulum Jombang. Mereka juga sekolah disalah satu Madrasah Tsanawiyah dipondok tersebut. Untuk menyelesaikan studi mereka, mereka harus dapat menghafal juz 30. Tes hafalan dilaksanakan satu minggu setelah Ujian Nasional. Tiga minggu sebelum tes hafalan berlangsung, jumlah surat yang Adinda hafal 5 kali banyaknya surat yang dihafal fitri. Jumlah surat yang dihafal fitri 2 kali lebih banyak dari Siti. Sedangkan jumlah surat yang dihafal Siti tiga lebihnya dari jumlah surat Aisah. Pernyataan yang benar berkaitan dengan kondisi diatas adalah. . .”

4) Butir Soal Nomor 4

Redaksi Awal butir soal nomor 4

- “ (i) Sudut yang dibentuk oleh pertemuan pangkal dua buah garis lurus
yang saling tegak lurus merupakan gerakan ruku' dalam sholat.*
- (ii) Sudut 90^0 yang membagi dua sama besar merupakan gerakan Sujud dalam sholat,*

- (iii) *Jika dalam satu rakaat gerakan shalat Ruku', I'tidal dan sujud dijumlahkan, maka sudut yang dihasilkan kurang dari 360^0 .*
- (iv) *Sudut yang dihasilkan dalam satu rakaat gerakan shalat Sujud merupakan sudut yang sama dalam gerakan shalat I'tidal"*

Redaksi butir soal nomor 4 direvisi menjadi

- " (i) *Sudut yang dibentuk oleh pertemuan pangkal dua buah garis lurus yang saling tegak lurus merupakan gerakan Ruku' dalam shalat.*
- (ii) *Jika dalam satu rakaat gerakan shalat Ruku', I'tidal dan Sujud dijumlahkan, maka sudut yang dihasilkan kurang dari 360^0 .*
- (iii) *Sudut yang dihasilkan dalam gerakan Sujud adalah setengah dari gerakan Ruku'.*
- (iv) *Sudut yang dihasilkan dalam satu rakaat gerakan shalat Sujud merupakan sudut yang sama dalam gerakan shalat I'tidal."*

Redaksi awal jawaban butir soal nomor 4

- " A. (i) dan (ii)
 B. (ii) dan (iii)
 C. (iii) dan (iv)
 D. (i) dan (iv)"

Redaksi jawaban butir soal nomor 4 direvisi menjadi

- " A. (i) dan (ii)
 B. (ii) dan (iii)
 C. (iii) dan (iv)
 D. (i) dan (iii)"

5) Butir Soal Nomor 7

Redaksi awal butir soal nomor 7

"Lilah mendapatkan kesempatan untuk menyempurnakan rukun islamnya dengan menuaikan ibadah Haji. Kesempatan ini didapatkannya setelah dia menabung selama bertahun-tahun. Diapun langsung mendaftarkan dirinya ke Kementerian Agama. Setelah melakukan pendaftaran dan adiministrasi, petugas

pendaftaran mengatakan kepada Lilah bahwa “Dalam dua puluh tahun kedepan, peluang keberangkatan Haji adalah lima per enam.” Pernyataan di bawah ini yang paling mencerminkan maksud dari petugas pendaftaran adalah . . . “

Redaksi butir soal nomor 7 direvisi menjadi

“Lilah mendapatkan kesempatan untuk menyempurnakan rukun islamnya dengan menuaikan ibadah Haji. Kesempatan ini didapatkannya setelah dia menabung selama bertahun-tahun. Diapun langsung mendaftarkan dirinya ke Kementrian Agama. Setelah melakukan pendaftaran dan adiministrasi, petugas pendaftaran mengatakan kepada Lilah bahwa “Dalam dua puluh tahun kedepan, peluang keberangkatan Haji adalah lima per enam.” Pernyataan di bawah ini yang benar mencerminkan maksud dari petugas pendaftaran adalah . . . ”

6) Butir Soal Nomor 8

Redaksi Awal butir soal nomor 8

“Masjid Al-Akbar Surabaya mengadakan kegiatan buka bersama dengan yatim piatu dan dhuafa. Acara tersebut diikuti lebih dari 1000 peserta. Panitia kegiatan mengatur tempat duduk sedemikanrupa diamana tempat duduk yang diatur membentuk suatu pola barisan. Barisan pertama terdapat 2 kursi. Barisan kedua dan ketiga terdiri dari 6 dan 12 kursi. Barisan keempat terdiri dari 20 kursi. Ahmad, Imron dan Firman merupakan peserta kegiatan buka bersama. Ahmad berada di barisan ke sepuluh. Imron berada di barisan ke dua belas. Sedangkan Firman berada di barisan ke empat belas. Banyaknya kursi pada barisan ke sepuluh, dua belas dan empat belas pada kegiatan buka bersama adalah...”

Redaksi butir soal nomor 8 direvisi menjadi

“Masjid Al-Akbar Surabaya mengadakan kegiatan buka bersama dengan yatim piatu dan dhuafa. Acara tersebut diikuti lebih dari 1000 peserta. Panitia kegiatan mengatur tempat duduk sedemikan rupa di mana tempat duduk yang diatur membentuk suatu pola barisan. Barisan pertama terdapat 2 kursi. Barisan kedua dan ketiga terdiri dari 6 dan 12 kursi. Barisan keempat terdiri dari 20 kursi. Ahmad, Imron dan

Firman merupakan peserta kegiatan buka bersama. Ahmad berada di barisan ke sepuluh. Imron berada di barisan ke dua belas. Sedangkan Firman berada di barisan ke empat belas. Banyaknya kursi pada barisan yang ditempati Ahmad, Imron dan Firman pada kegiatan buka bersama adalah...”

4. Data Uji Coba

Instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang telah ditelaah oleh para validator (Lampiran 5, Lampiran 6 dan Lampiran 7) kemudian dilakukan uji coba di empat sekolah yang melibatkan 165 siswa. Kegiatan uji coba dilaksanakan tanggal 22 November 2017 – 12 Desember 2017. Jadwal dan jumlah yang mengikuti uji coba pada setiap sekolah disajikan pada Tabel 4.2:

Tabel 4.2
Jadwal dan Banyaknya Peserta Kegiatan Uji Coba

Nama Sekolah	Kelas	Jumlah Peserta	Waktu Pelaksanaan
MTsN Gresik	9C	31	22 November 2017
	9A	33	23 November 2017
MTsN 1 Sidoarjo	9A	29	24 November 2017
MTsN 2 Surabaya	9A	36	24 November 2017
MTsN 3 Surabaya	9A	36	12 Desember 2017

Dari hasil uji coba yang telah dilakukan, diperoleh suatu data yang digunakan untuk mengetahui karakteristik instrumen penilaian yang meliputi daya pembeda, tingkat kesukaran dan nilai *guessing*.

5. Menganalisis Butir Soal Tes

Setelah peneliti memperoleh data dari uji coba lapangan, dilakukan kegiatan *scoring* atau penskoran. Hal tersebut dilakukan untuk menganalisis karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang meliputi tingkat kesukaran, daya pembeda dan nilai *guessing*. Analisis instrumen penilaian dilakukan dengan menggunakan program BILOG-MG. Model

analisis yang digunakan adalah model tiga parameter logistik (3PL).

B. Penyajian dan Analisis Data Karakteristik Instrumen Penilaian *Higher Order Thinking* untuk Pelajaran Matematika dengan Memasukkan Konteks Islam

Berdasarkan data yang diperoleh dilapangan, dilakukan kegiatan analisis untuk mengetahui daya pembeda, tingkat kesukaran dan nilai *guessing* dari instrumen penilaian yang dikembangkan. Berikut langkah yang dilakukan dalam kegiatan analisis:

1. Analisis Instrumen Penilaian Butir Soal

Langkah pertama dalam kegiatan analisis yaitu melihat keseluruhan hasil instrumen penilaian dari data peserta kegiatan uji coba dilapangan. Disajikan pada Gambar 4.1:

ITEM	NAME	#TRIED	#RIGHT	PCT	ITEM*TEST CORRELATION		
					LOGIT/1.7	PEARSON	BISERIAL
1	MATH1	165.0	84.0	50.9	-0.02	0.519	0.650
2	MATH2	165.0	49.0	29.7	0.51	0.359	0.474
3	MATH3	165.0	65.0	39.4	0.25	-0.012	-0.016
4	MATH4	165.0	65.0	39.4	0.25	0.196	0.249
5	MATH5	165.0	47.0	28.5	0.54	0.081	0.107
6	MATH6	165.0	72.0	43.6	0.15	0.187	0.236
7	MATH7	165.0	65.0	39.4	0.25	0.388	0.493
8	MATH8	165.0	110.0	66.7	-0.41	0.408	0.530

Gambar 4.1
Hasil Analisis Instrumen Penilaian Butir Soal

Gambar 4.1 di atas menunjukkan hasil keseluruhan analisis setiap butir soal dari program BILOG-MG terhadap kedelapan butir soal instrumen penilaian yang dikembangkan. Kolom *#TRIED* menunjukkan jumlah peserta kegiatan uji coba, kolom *#RIGHT* menunjukkan jumlah peserta kegiatan uji coba yang menjawab benar, kolom *PCT* menunjukkan persentase menjawab benar. Terdapat dua kolom berbeda terkait dengan *ITEM*TEST CORRELATION*. Kolom pertama, *ITEM*TEST CORRELATION PEARSON* menunjukkan korelasi skor butir soal dengan skor total yang dikorelasi dengan *product-moment*. Sedangkan kolom kedua, *ITEM*TEST CORRELATION*

BISERIAL merupakan korelasi biserial atau dalam teori tes klasikal disebut sebagai daya pembeda¹.

Terkait hasil analisis perhitungan statistik setiap butir soal dari program BILOG-MG, diperoleh hasil perhitungan *correlation biserial* (korelasi biserial) yang berbeda tiap butir soal. Butir soal nomor 3 mempunyai korelasi biserial sebesar -0.016 , artinya butir soal nomor 3 tidak akan dikalibrasi karena nilai korelasi biserialnya bernilai negatif. Sehingga butir soal nomor 3 harus dikeluarkan atau tidak digunakan dalam analisis karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam. Hal tersebut sejalan dengan Ebel & Fresibe yang menyatakan bahwa pemilihan item yang baik terkait analisis, item dengan biserial yang negatif sebaiknya dikeluarkan dari model. Dengan demikian, dilakukan revisi analisis instrumen penilaian dengan membuang butir soal nomor 3. Berikut hasil revisi analisis instrumen penilaian yang tersaji pada Gambar 4.2.

ITEM STATISTICS FOR SUBTEST PRETEST

ITEM	NAME	#TRIED	#RIGHT	PCT	LOGIT/1.7	ITEM*TEST CORRELATION	
						PEARSON	BISERIAL
1	ITEM1	165.0	84.0	50.9	-0.02	0.544	0.682
2	ITEM2	165.0	49.0	29.7	0.51	0.414	0.547
3	ITEM3	165.0	65.0	39.4	0.25	0.216	0.275
4	ITEM4	165.0	47.0	28.5	0.54	0.072	0.095
5	ITEM5	165.0	72.0	43.6	0.15	0.197	0.248
6	ITEM6	165.0	65.0	39.4	0.25	0.411	0.521
7	ITEM7	165.0	110.0	66.7	-0.41	0.392	0.509

Gambar 4.2
Hasil Revisi Analisis Instrumen Penilaian Butir Soal

Berdasarkan Gambar 4.2, terdapat 7 butir soal yang dianalisis di mana sebelumnya telah membuang satu butir soal yang bernilai negatif. Dari hasil revisi analisis tersebut, diperoleh informasi bahwa tidak ada lagi item yang memiliki korelasi biserial negatif, sehingga dapat dilakukan uji parameter logistik 3 PL.

¹Heri Retnawati, *Validitas Reabilitas & Karakteristik Butir* (Yogyakarta: Parama Publishing, 2016), 133.

2. Analisis Karakteristik Instrumen Penilaian

Langkah kedua dalam kegiatan analisis yaitu menganalisis karakteristik dari instrumen penilaian yang dikembangkan. Hasil revisi analisis karakteristik instrumen penilaian yang dikembangkan tersaji pada gambar berikut:

ITEM	INTERCEPT S.E.	SLOPE S.E.	THRESHOLD S.E.	LOADING S.E.	ASYMPTOTE S.E.	CHISQ (PROB)	DF
ITEM1	-1.021 1.050*	5.407 3.137*	0.189 0.145*	0.983 0.570*	0.166 0.061*	15.7 (0.0004)	2.0
ITEM2	-1.607 0.853*	1.749 0.914*	0.919 0.175*	0.868 0.454*	0.124 0.042*	16.3 (0.0026)	4.0
ITEM3	-1.015 0.427*	0.805 0.294*	1.261 0.362*	0.627 0.229*	0.228 0.066*	7.2 (0.4125)	7.0
ITEM4	-2.148 0.780*	0.848 0.361*	2.533 0.820*	0.647 0.275*	0.244 0.045*	8.0 (0.3323)	7.0
ITEM5	-0.882 0.403*	0.657 0.231*	1.343 0.477*	0.549 0.193*	0.276 0.075*	11.0 (0.1368)	7.0
ITEM6	-0.735 0.365*	1.333 0.441*	0.551 0.184*	0.800 0.264*	0.132 0.052*	6.0 (0.1128)	3.0
ITEM7	0.268 0.241*	1.010 0.258*	-0.265 0.258*	0.711 0.182*	0.221 0.087*	13.4 (0.0039)	3.0

* STANDARD ERROR

Gambar 4.3
Hasil Analisis Karakteristik Instrumen Penilaian

Gambar 4.3 menunjukkan hasil keseluruhan analisis karakteristik setiap butir soal dari program BILOG-MG terhadap ke-7 butir soal instrumen penilaian yang dikembangkan. Tersaji empat kolom yang menunjukkan hasil karakteristik dari instrumen penilaian meliputi kolom SLOPE berisi parameter a atau daya pembeda, kolom THRESHOLD berisi parameter b atau tingkat kesukaran, kolom ASYMTOTE berisi parameter c atau nilai *guessing*, dan kolom CHISQ berisi kesesuaian butir (fit model). Berikut penjabaran terkait karakteristik instrumen yang dikembangkan dari data yang diperoleh di lapangan:

a. SLOPE (Daya Pembeda atau Parameter a)

Parameter a menunjukkan karakteristik instrumen penilaian terkait SLOPE (daya pembeda). Sesuai dengan kajian pada BAB III, daya pembeda yang baik berkisar antara $0 \leq$

$a \leq 2$. Berikut hasil analisis daya pembeda setiap butir soal yang disajikan pada Tabel 4.3²

Tabel 4.3
Hasil Analisis Indeks Daya Pembeda

No.	Indeks Pembeda (a)	Kategori	Jumlah Item	Nomor Item
1	$a > 2$	Kurang Baik	1	1
2	$0 \leq a \leq 2$	Baik	6	2, 3, 4, 6.
Total Item			7	

Tabel 4.3 menunjukkan bahwa hanya satu instrumen penilaian butir soal yang mempunyai daya pembeda yang kurang baik, yaitu butir soal nomor 1. Sedangkan untuk butir soal yang lain, daya pembeda soal dalam kategori baik.

b. THRESHOLD (Tingkat Kesukaran atau Parameter b)

Parameter b menunjukkan karakteristik instrumen penilaian terkait THRESHOLD (tingkat kesukaran). Sesuai dengan kajian pada BAB III, tingkat kesukaran yang baik berkisar antara $-2 \leq b \leq 2$. Berikut hasil analisis tingkat kesukaran setiap butir soal yang disajikan pada Tabel 4.4³

Tabel 4.4
Hasil Analisis Indeks Tingkat Kesukaran

No.	Indeks Kesukaran (b)	Kategori	Jumlah Item	Nomor Item
1	$b > 2$	Kurang Baik	1	5
2	$-2 \leq b \leq 2$	Baik	7	1, 2, 3, 4, 6, 7,
3	$b < -2$	Kurang Baik	-	-
Total Item			7	

Tabel 4.4 diperoleh informasi bahwa instrumen penilaian butir soal nomor 5 memiliki kategori tingkat kesukaran yang

²Devi Dwi Kurniawan, "Analisis Kualitas Soal Ujian Akhir Semester Matematika Berdasarkan Teori Respon Butir". (Paper presented at Prosiding Seminar Nasional Matematika dan Pendidikan Matematika, UMS, 2015), 130.

³Ibid, halaman 130.

kurang baik dan instrumen penilaian butir soal yang lain memiliki kategori tingkat kesukaran yang baik.

c. ASYMTOTE (Nilai *Guessing* atau parameter *c*)

Parameter *c* menunjukkan karakteristik instrumen penilaian terkait ASYMTOTE (nilai *guessing*). Sesuai dengan kajian pada BAB III, nilai *guessing* yang baik berkisar antara $0 \leq c \leq 1$. Berikut hasil analisis nilai *guessing* setiap butir soal yang disajikan pada Tabel 4.5⁴.

Tabel 4.5
Hasil Analisis Indeks Nilai *Guessing*

No.	Indeks Tebakan (<i>c</i>)	Kategori	Jumlah Item	Nomor Item
1	$c > 1$	Kurang Baik	-	
2	$0 \leq c \leq 1$	Baik	7	1, 2, 3, 4, 5, 6, 7
Total Item			7	

Tabel 4.5 diperoleh informasi bahwa seluruh instrumen penilaian butir soal memiliki nilai *guessing* dengan kategori baik. Tidak ada satupun butir soal yang memiliki nilai *guessing* dengan kategori yang tidak baik.

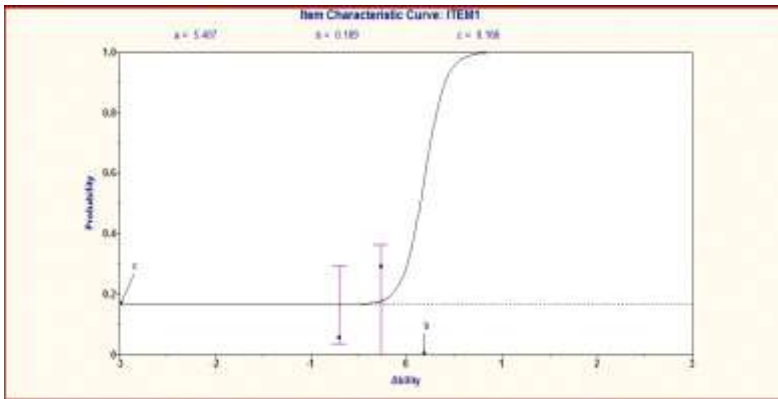
3. Analisis Karakteristik Instrumen Penilaian Butir Soal

Langkah ketiga dalam kegiatan analisis yaitu menganalisis karakteristik instrumen penilaian dari setiap butir soal yang dikembangkan. Berikut penjabaran karakteristik setiap butir soal dari hasil revisi analisis karakteristik instrumen penilaian yang dikembangkan:

a. Butir Soal Nomor 1

Hasil kegiatan uji coba lapangan, menunjukkan bahwa hanya 84 siswa yang menjawab benar sedangkan 81 siswa yang menjawab salah. Dari data butir soal nomor 1, dihasilkan grafik ICC pada analisis karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan dengan model 3PL yang tersaji pada Gambar 4.4:

⁴Frank B. Baker, *The Basics of Item Response Theory* (United States of Amerika: ERIC Clearinghouse on Assessment and Evaluation, 2001), 28.



Gambar 4.4
Item Characteristic Curve Butir Soal Nomor 1

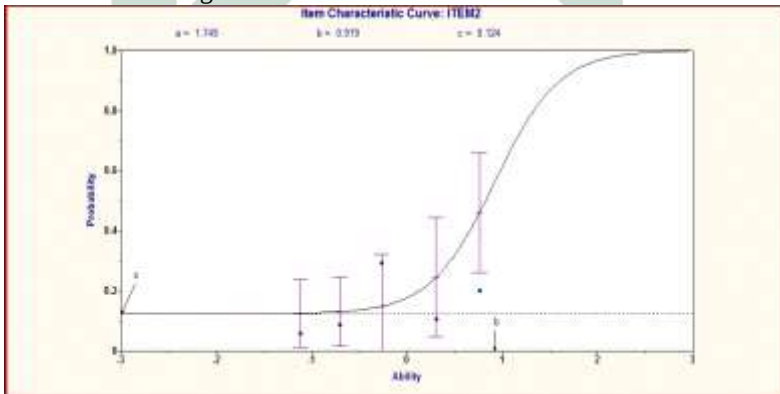
Berdasarkan Gambar 4.5, diperoleh informasi nilai a_1 yang merupakan daya pembeda butir soal nomor 1 yaitu 5.407, nilai $b_1 = 0.189$ yang merupakan tingkat kesukaran dan nilai guessing $c_1 = 0.166$. Besar kecilnya nilai dari daya pembeda yang dihasilkan, akan mempengaruhi bentuk dari grafik ICC. Semakin besar daya pembeda suatu soal maka grafik ICC akan semakin menanjak, sedangkan untuk daya pembeda dengan nilai kecil maka grafik ICC akan semakin landai. Grafik ICC yang tergambar pada butir soal nomor 1, terlihat sangat menanjak. Hal tersebut dikarenakan butir soal nomor 1 memiliki nilai daya pembeda yang sangat besar bahkan melebihi dari kriteria yang ditetapkan.

Berdasarkan penjelasan terkait analisis butir soal nomor 1, diperoleh data terkait karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan. Daya pembeda dengan nilai 5.407 termasuk dalam kategori tidak baik. Artinya, butir soal nomor 1 tidak dapat membedakan antara siswa yang sudah dan belum memahami materi. Tingkat kesukaran dengan nilai 0.189 termasuk dalam kategori baik. Artinya, siswa dengan kemampuan tinggi merasa mudah dalam mengerjakan dan sebaliknya siswa dengan kemampuan rendah akan sulit menjawab butir soal nomor 1. Sedangkan peluang kemampuan

siswa untuk *guessing* (menebak) jawaban benar butir soal nomor 1 adalah adalah 0.166. Artinya, butir soal nomor 1 dijawab benar tanpa adanya unsur kebetulan.

b. Butir Soal Nomor 2

Hasil kegiatan uji coba lapangan, menunjukkan bahwa hanya 49 siswa yang menjawab benar sedangkan 116 siswa yang menjawab salah. Dari data butir soal nomor 2, dihasilkan grafik ICC pada analisis karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan dengan model 3PL sebagai berikut:



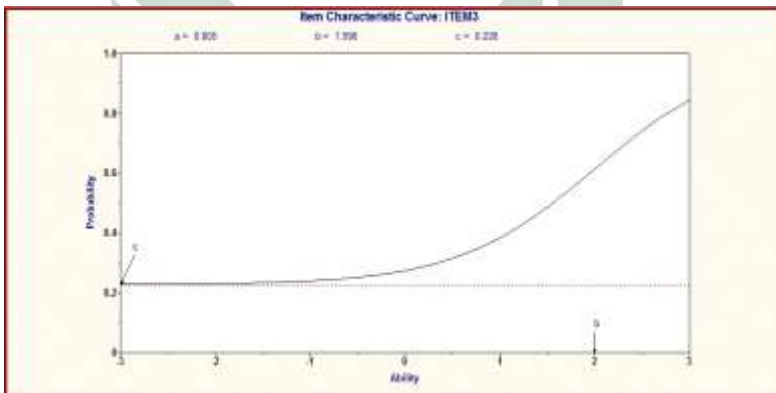
Gambar 4.5
Item Characteristic Curve Butir Soal Nomor 2

Berdasarkan Gambar 4.5, diperoleh informasi nilai a_2 yang merupakan daya pembeda butir soal nomor 2 yaitu 1.749, nilai $b_2 = 0.919$ yang merupakan tingkat kesukaran dan nilai *guessing* $c_2 = 0.124$. Besar kecilnya nilai dari daya pembeda yang dihasilkan, akan mempengaruhi bentuk dari grafik ICC. Semakin besar daya pembeda suatu soal maka grafik ICC akan semakin menanjak, sedangkan untuk daya pembeda dengan nilai kecil maka grafik ICC akan semakin landai. Grafik ICC yang tergambar pada butir soal nomor 2, terlihat menanjak. Hal tersebut dikarenakan butir soal nomor 2 memiliki nilai daya pembeda yang besar.

Berdasarkan penjelasan terkait analisis butir soal nomor 2, diperoleh data terkait karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan. Daya pembeda dengan nilai 1.749 termasuk dalam kategori baik. Artinya, butir soal nomor 2 dapat membedakan antara siswa yang sudah dan belum memahami materi. Tingkat kesukaran dengan nilai 0.919 termasuk dalam kategori baik. Artinya, siswa dengan kemampuan tinggi merasa mudah dalam mengerjakan dan sebaliknya siswa dengan kemampuan rendah akan sulit menjawab butir soal nomor 2. Sedangkan peluang kemampuan siswa untuk *guessing* (menebak) jawaban benar butir soal nomor 2 adalah 0.124. Artinya, butir soal nomor 2 dijawab benar tanpa adanya unsur kebetulan.

c. Butir Soal Nomor 3

Hasil kegiatan uji coba lapangan, menunjukkan bahwa hanya 65 siswa yang menjawab benar sedangkan 100 siswa yang menjawab salah. Dari data butir soal nomor 3, dihasilkan grafik ICC pada analisis karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan dengan model 3PL sebagai berikut:



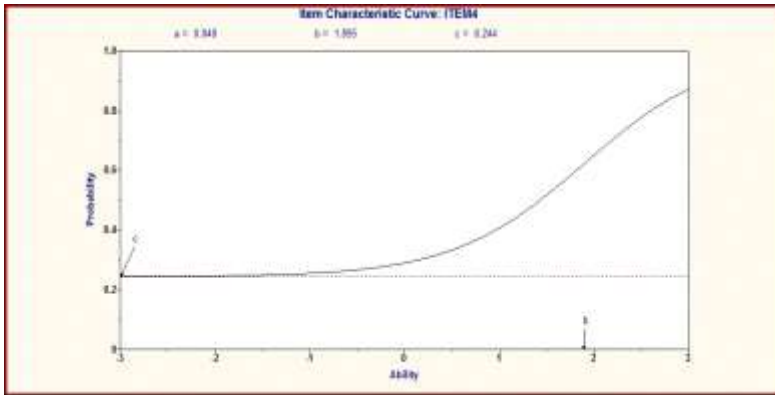
Gambar 4.6
Item Characteristic Curve Butir Soal Nomor 3

Berdasarkan gambar 4.6, diperoleh informasi nilai a_3 yang merupakan daya pembeda butir soal nomor 3 yaitu 0.805, nilai $b_3 = 1.996$ yang merupakan tingkat kesukaran dan nilai *guessing* $c_3 = 0.228$. Besar kecilnya nilai dari daya pembeda yang dihasilkan, akan mempengaruhi bentuk dari grafik ICC. Semakin besar daya pembeda suatu soal maka grafik ICC akan semakin menanjak, sedangkan untuk daya pembeda dengan nilai kecil maka grafik ICC akan semakin landai. Grafik ICC yang tergambar pada butir soal nomor 3, terlihat landai. Hal tersebut dikarenakan butir soal nomor 3 memiliki nilai daya pembeda yang kecil.

Berdasarkan penjelasan terkait analisis butir soal nomor 3, diperoleh data terkait karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan. Daya pembeda dengan nilai 0.805 termasuk dalam kategori baik. Artinya, butir soal nomor 3 dapat membedakan antara siswa yang sudah dan belum memahami materi. Tingkat kesukaran dengan nilai 1.996 termasuk dalam kategori baik. Artinya, siswa dengan kemampuan tinggi merasa mudah dalam mengerjakan dan sebaliknya siswa dengan kemampuan rendah akan sulit menjawab butir soal nomor 3. Sedangkan peluang kemampuan siswa untuk *guessing* (menebak) jawaban benar butir soal nomor 3 adalah 0.228. Artinya, butir soal nomor 3 dijawab benar tanpa adanya unsur kebetulan.

d. Butir Soal Nomor 4

Hasil kegiatan uji coba lapangan, menunjukkan bahwa hanya 47 siswa yang menjawab benar sedangkan 118 siswa yang menjawab salah. Dari data butir soal nomor 4, dihasilkan grafik ICC pada analisis karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan dengan model 3PL sebagai berikut:



Gambar 4.7
Item Characteristic Curve Butir Soal Nomor 4

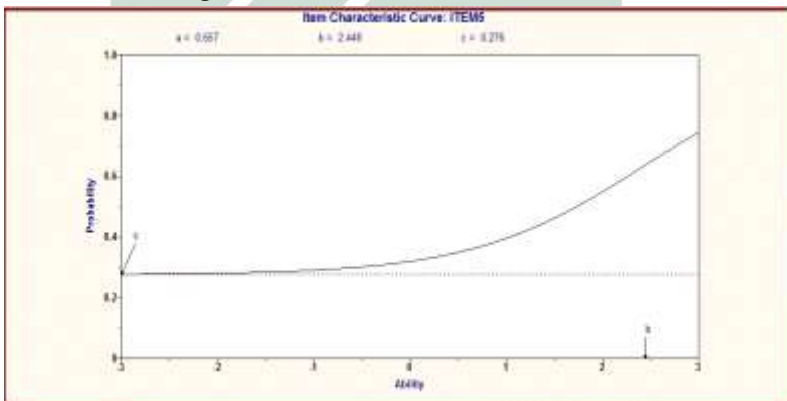
Berdasarkan gambar 4.7, diperoleh informasi nilai a_4 yang merupakan daya pembeda butir soal nomor 4 yaitu 0.848, nilai $b_4 = 1.895$ yang merupakan tingkat kesukaran dan nilai *guessing* $c_4 = 0.244$. Besar kecilnya nilai dari daya pembeda yang dihasilkan, akan mempengaruhi bentuk dari grafik ICC. Semakin besar daya pembeda suatu soal maka grafik ICC akan semakin menanjak, sedangkan untuk daya pembeda dengan nilai kecil maka grafik ICC akan semakin landai. Grafik ICC yang tergambar pada butir soal nomor 4, terlihat landai. Hal tersebut dikarenakan butir soal nomor 4 memiliki nilai daya pembeda yang kecil.

Berdasarkan penjelasan terkait analisis butir soal nomor 4, diperoleh data terkait karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan. Daya pembeda dengan nilai 0.848 termasuk dalam kategori baik. Artinya, butir soal nomor 4 dapat membedakan antara siswa yang sudah dan belum memahami materi. Tingkat kesukaran dengan nilai 1.895 termasuk dalam kategori baik. Artinya, siswa dengan kemampuan tinggi merasa mudah dalam mengerjakan dan sebaliknya siswa dengan kemampuan rendah akan sulit menjawab butir soal nomor 4. Sedangkan peluang kemampuan siswa untuk *guessing* (menebak) jawaban benar butir soal nomor

4 adalah 0.244. Artinya, butir soal nomor 4 dijawab benar tanpa adanya unsur kebetulan.

e. Butir Soal Nomor 5

Hasil kegiatan uji coba lapangan, menunjukkan bahwa hanya 72 siswa yang menjawab benar sedangkan 93 siswa yang menjawab salah. Dari data butir soal nomor 5, dihasilkan grafik ICC pada analisis karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan dengan model 3PL sebagai berikut:



Gambar 4.8
Item Characteristic Curve Butir Soal Nomor 5

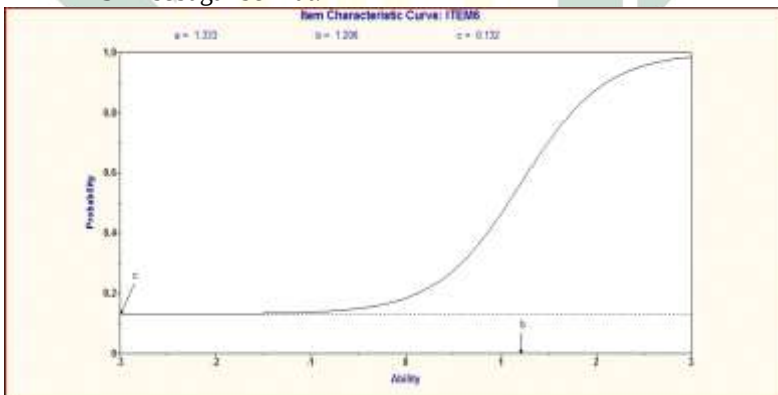
Berdasarkan gambar 4.8, diperoleh informasi nilai a_5 yang merupakan daya pembeda butir soal nomor 5 yaitu 0.657, nilai $b_5 = 2.448$ yang merupakan tingkat kesukaran dan nilai $c_5 = 0.276$. Besar kecilnya nilai dari daya pembeda yang dihasilkan, akan mempengaruhi bentuk dari grafik ICC. Semakin besar daya pembeda suatu soal maka grafik ICC akan semakin menanjak, sedangkan untuk daya pembeda dengan nilai kecil maka grafik ICC akan semakin landai. Grafik ICC yang tergambar pada butir soal nomor 5, terlihat landai. Hal tersebut dikarenakan butir soal nomor 5 memiliki nilai daya pembeda yang kecil.

Berdasarkan penjelasan terkait analisis butir soal nomor 5, diperoleh data terkait karakteristik instrumen penilaian *Higher*

Order Thinking untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan. Daya pembeda dengan nilai 0.657 termasuk dalam kategori baik. Artinya, butir soal nomor 5 dapat membedakan antara siswa yang sudah dan belum memahami materi. Tingkat kesukaran dengan nilai 2.448 termasuk dalam kategori tidak baik. Artinya, butir soal nomor 5 sulit membedakan manakah kelompok dengan kemampuan tinggi dan rendah yang benar-benar mampu menjawab butir soal nomor 5. Sedangkan peluang kemampuan siswa untuk *guessing* (menebak) jawaban benar butir soal nomor 5 adalah 0.276. Artinya, butir soal nomor 5 dijawab benar tanpa adanya unsur kebetulan.

f. Butir Soal Nomor 6

Hasil kegiatan uji coba lapangan, menunjukkan bahwa hanya 65 siswa yang menjawab benar sedangkan 100 siswa yang menjawab salah. Dari data butir soal nomor 6, dihasilkan grafik ICC pada analisis karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan dengan model 3PL sebagai berikut:



Gambar 4.9
Item Characteristic Curve Butir Soal Nomor 6

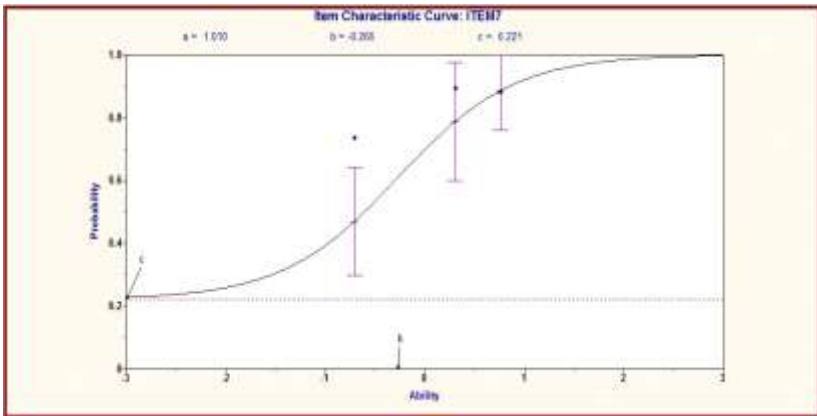
Berdasarkan gambar 4.9, diperoleh informasi nilai a_6 yang merupakan daya pembeda butir soal nomor 6 yaitu 1.333, nilai $b_6 = 1.206$ yang merupakan tingkat kesukaran dan nilai

guessing $c_6 = 0.132$. Besar kecilnya nilai dari daya pembeda yang dihasilkan, akan mempengaruhi bentuk dari grafik ICC. Semakin besar daya pembeda suatu soal maka grafik ICC akan semakin menanjak, sedangkan untuk daya pembeda dengan nilai kecil maka grafik ICC akan semakin landai. Grafik ICC yang tergambar pada butir soal nomor 6, terlihat menanjak. Hal tersebut dikarenakan butir soal nomor 6 memiliki nilai daya pembeda yang besar.

Berdasarkan penjelasan terkait analisis butir soal nomor 6, diperoleh data terkait karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan. Daya pembeda dengan nilai 1.333 termasuk dalam kategori baik. Artinya, butir soal nomor 6 dapat membedakan antara siswa yang sudah dan belum memahami materi. Tingkat kesukaran dengan nilai 1.206 termasuk dalam kategori baik. Artinya, siswa dengan kemampuan tinggi merasa mudah dalam mengerjakan dan sebaliknya siswa dengan kemampuan rendah akan sulit menjawab butir soal nomor 6. Sedangkan peluang kemampuan siswa untuk *guessing* (menebak) jawaban benar butir soal nomor 6 adalah 0.132. Artinya, butir soal nomor 6 dijawab benar tanpa adanya unsur kebetulan.

g. Butir Soal Nomor 7

Hasil kegiatan uji coba lapangan, menunjukkan bahwa hanya 110 siswa yang menjawab benar sedangkan 55 siswa yang menjawab salah. Dari data butir soal nomor 7, dihasilkan grafik ICC pada analisis karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan dengan model 3PL sebagai berikut:



Gambar 4.10
Item Characteristic Curve Butir Soal Nomor 7

Berdasarkan gambar 4.10, diperoleh informasi nilai a_7 yang merupakan daya pembeda butir soal nomor 7 yaitu 1.010, nilai $b_7 = -0.265$ yang merupakan tingkat kesukaran dan nilai guessing $c_7 = 0.221$. Besar kecilnya nilai dari daya pembeda yang dihasilkan, akan mempengaruhi bentuk dari grafik ICC. Semakin besar daya pembeda suatu soal maka grafik ICC akan semakin menanjak, sedangkan untuk daya pembeda dengan nilai kecil maka grafik ICC akan semakin landai. Grafik ICC yang tergambar pada butir soal nomor 7, terlihat menanjak. Hal tersebut dikarenakan butir soal nomor 7 memiliki nilai daya pembeda yang besar.

Berdasarkan penjelasan terkait analisis butir soal nomor 7, diperoleh data terkait karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan. Daya pembeda dengan nilai 1.010, termasuk dalam kategori baik. Artinya, butir soal nomor 7 dapat membedakan antara siswa yang sudah dan belum memahami materi. Tingkat kesukaran dengan nilai -0.265 termasuk dalam kategori baik. Artinya, siswa dengan kemampuan tinggi merasa mudah dalam mengerjakan dan sebaliknya siswa dengan kemampuan rendah akan sulit menjawab butir soal nomor 7. Sedangkan peluang kemampuan

siswa untuk *guessing* (menebak) jawaban benar butir soal nomor 7 adalah adalah 0.221. Artinya, butir soal nomor 7 dijawab benar tanpa adanya unsur kebetulan.

Berdasarkan penjabaran mengenai kegiatan analisis di atas, diperoleh rekapitulasi hasil estimasi perhitungan karakteristik instrumen penilaian meliputi daya pembeda, tingkat kesukaran dan nilai *guessing* yang tersaji pada Tabel 4.6.

Tabel 4.6
Rekapitulasi Hasil Estimasi Perhitungan
Karakteristik Instrumen Penilaian

Butir Soal Nomor	Daya Pembeda		Tingkat Kesukaran		Nilai <i>Guessing</i>		Chi-Squared		Ket
	<i>a</i>	Ket	<i>b</i>	Ket	<i>c</i>	Ket	P-value	Ket	
ITEM 1	5.407	Tidak Baik	0.189	Baik	0.166	Baik	0.0004	Tidak Fit Model	Tidak Baik
ITEM 2	1.749	Baik	0.919	Baik	0.124	Baik	0.0026	Tidak Fit Model	Tidak Baik
ITEM 3	0.805	Baik	1.996	Baik	0.228	Baik	0.4125	Fit Model	Baik
ITEM 4	0.848	Baik	1.895	Baik	0.244	Baik	0.3323	Fit Model	Baik
ITEM 5	0.657	Baik	2.448	Tidak Baik	0.276	Baik	0.1368	Fit Model	Tidak Baik
ITEM 6	1.333	Baik	1.206	Baik	0.132	Baik	0.1128	Fit Model	Baik
ITEM 7	1.010	Baik	-0.265	Baik	0.221	Baik	0.0039	Tidak Fit Model	Tidak Baik
Rerata	1.687		0.933		0.199				

Berdasarkan Tabel 4.6, jumlah butir soal yang layak dianalisis menggunakan *Item Response Theory* (IRT) model 3PL adalah 7 butir soal. Estimasi parameter butir *a* dari ke-7 butir soal, menunjukkan 6 butir soal termasuk kategori daya

pembeda yang baik dan hanya 1 butir soal yang memiliki daya pembeda tidak baik. Hal ini terlihat dari estimasi parameter a yang tidak berada diantara $0 \leq a \leq 2$.⁵ Selanjutnya, estimasi parameter butir b dari ke-7 butir soal menunjukkan 6 butir soal termasuk kategori daya pembeda yang baik dan hanya 1 butir soal yang memiliki tingkat kesukaran yang tidak baik. Hal ini terlihat dari estimasi parameter b yang tidak berada $-2 \leq b \leq 2$.⁶ Satu butir soal tersebut, dikategorikan soal yang sulit yaitu butir soal nomor 6 dengan tingkat kesukaran sebesar 2.488. Sedangkan estimasi parameter butir c , menunjukkan perbedaan dari estimasi parameter butir a dan b . Song Gao dalam disertasinya, menyatakan “*The highest probability of guessing cannot be greater than 0,5; otherwise, it would not be called knowledge-based answer instead of guessing*”⁷. Artinya, kemungkinan terbesar dari *guessing* (menebak) tidak lebih dari 0,5; dengan kata lain tidak bisa dikatakan bahwa jawaban dari sebuah pengetahuan hanya dari *guessing* (menebak). Sedangkan hasil estimasi parameter butir c dari ke-7 instrumen, tidak lebih dari 0.5. Artinya, peluang kemampuan siswa untuk *guessing* (menebak) jawaban dari ke-7 butir soal kurang dari 0.5. Terkait dengan kecocokan model, terdapat 3 butir soal yang menunjukkan butir soal tidak cocok dengan model 3PL yaitu butir soal nomor 1 (0.0004), butir soal nomor 2 (0.0026), butir soal nomor 7 (0.0039). Hal ini dikarenakan nilai signifikansi (p -value) dari chi squared di bawah taraf signifikansi 0.05.

Hasil informasi parameter di atas, dapat dijadikan informasi untuk menentukan butir-butir soal mana yang memiliki karakteristik yang baik. Butir soal yang baik hendaknya memenuhi 4 persyaratan yaitu memiliki daya pembeda yang baik ($0 \leq a \leq 2$), tingkat kesukaran yang baik ($-2 \leq b \leq 2$), nilai *guessing* yang baik ($0 \leq c \leq 1$) dan butir soal cocok dengan model. Berdasarkan hal tersebut, dari ke-7 butir soal,

⁵Devi Dwi Kurniawan, “Analisis Kualitas Soal Ujian Akhir Semester Matematika Berdasarkan Teori Respon Butir”. (Paper presented at Prosiding Seminar Nasional Matematika dan Pendidikan Matematika, UMS, 2015), 130.

⁶ Ibid, halaman 130.

⁷ Song Gao, Doctoral Dissertation: “*The Exploration of The Relationship Between Guessing and Latent Ability in IRT Models*” (Carbondale: Southern Illinois University, 2011), 37.

terdapat 3 soal yang termasuk dalam karakteristik butir yang baik yaitu butir soal nomor 3, 4 dan 6.

Hal yang berbeda ditunjukkan dari instrumen penilaian yang telah dikembangkan. Dari hasil instrumen yang telah dikembangkan menunjukkan rerata daya pembeda (*a*) sebesar 1.687 yang termasuk dalam kategori baik, rerata dari tingkat kesukaran (*b*) sebesar 0.933 yang termasuk dalam kategori baik dan rerata nilai *guessing* (*c*) sebesar 0.199 termasuk dalam kategori baik atau peluang kemampuan siswa untuk *guessing* (menebak) jawaban benar tidak lebih dari 0.5.

C. Kajian Produk Akhir

Berdasarkan hasil pengembangan dan uji coba produk yang telah dilakukan dilapangan, maka dapat dikemukakan berbagai kajian terkait produk akhir instrumen penilaian yang dihasilkan. Berikut temuan yang diperoleh:

1. Temuan Terkait Proses Pengembangan Instrumen Penilaian *Higher Order Thinking* untuk Pelajaran Matematika dengan Memasukkan Konteks Islam

Berdasarkan penyajian dan analisis data pengembangan instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam, terdapat temuan menarik pada saat proses pengembangan berlangsung. Temuan tersebut, terkait saat proses penelaah soal tes. Telaah soal tes dilakukan oleh 4 validator ahli dalam bidangnya yang meliputi tiga guru matematika dan satu dosen pendidikan matematika. Sebanyak 10 butir soal instrumen penilaian yang dikembangkan, yang ditelaah oleh para validator. Instrumen penilaian yang dikembangkan, di uji coba kepada siswa Sekolah Menengah Pertama kelas IX.

Dari validator pertama, soal dikatakan layak untuk di uji coba. Akan tetapi, validator pertama menanyakan waktu yang dibutuhkan dalam mengerjakan ke-10 soal tersebut. Dari hasil diskusi yang kami lakukan, diperoleh waktu 2×40 menit sebagai waktu dalam pengerjaan. Tidak ada revisi terkait soal yang dikembangkan.

Beralih ke validator kedua, validator kedua menyatakan instrumen penilaian yang dikembangkan layak untuk di uji coba dengan waktu 2×40 menit. Berbeda dengan validator pertama,

validator kedua berasal dari guru madrasah. Tujuan dari telaah oleh guru madrasah, untuk melihat aspek isi, bahasa dan keterbacaan instrumen penilaian yang dikembangkan terkait dengan konteks Islam.

Berdasarkan aspek isi, instrumen penilaian yang dikembangkan dinyatakan valid oleh validator 2. Hal tersebut dikarenakan instrumen soal yang dikembangkan telah sesuai dengan materi matematika yang dikaitkan dengan konteks islam yang tersaji pada butir soal. Terkait dengan konteks Islam yang disajikan pada soal, sudah sesuai dengan apa yang didapat siswa pada jenjang Madrasah Tsanawiyah. Keterkaitan antara kompetensi dasar, indikator HOT dengan konteks Islam yang disajikan pada setiap butir soal, tersaji secara runtut dan sistematis.

Berikutnya pada aspek bahasa dan keterbacaan, instrumen penilaian yang dikembangkan dibuat sesuai dengan tingkat pemahaman dan kemampuan siswa dalam memahami isi maupun bahasa dari setiap butir soal. Menurut Belawati, runtutan isi dari butir soal dapat mempermudah siswa untuk memahami isi dari butir soal. Instrumen penilaian yang disusun dengan secara runtut dan sistematis dapat membantu dan mempermudah siswa dalam penyelesaian soal. Dalam mengembangkan suatu instrumen penilaian, penggunaan bahasa menjadi salah satu faktor yang penting. Penggunaan bahasa meliputi pemilihan ragam bahasa, pemilihan kata, penggunaan kalimat efektif dan penyusunan paragraf yang bermakna, sangat berpengaruh terhadap butir soal. Walaupun isi butir soal sudah cermat, menggunakan format yang konsisten serta dikemas dengan menarik namun jika bahasa yang digunakan tidak dimengerti oleh siswa, maka butir soal tidak akan bermakna dan siswa akan kesulitan untuk menyelesaikan butir soal tersebut⁸.

Terkait dengan validator 3, banyak masukan dan saran yang disampaikan. Untuk mendapatkan kata layak di uji coba, peneliti melakukan diskusi selama 3 kali pertemuan dengan validator 3. Validator 3 memang sudah sangat berpengalaman dalam lingkungan sekolah. Selama diskusi berlangsung, validator 3 berkali-kali meninjau setiap butir soal. Bahkan, validator 3

⁸Tian Belawati, *Pengembangan Bahan Ajar* (Jakarta: Universitas Terbuka, 2003), 39.

mengujinya langsung dengan mengerjakan soal tersebut. Saat proses validator 3 mengerjakan soal, validator merasa kesulitan dalam menyelesaikan soal. Butuh waktu yang lama untuk menjawab setiap butir soal yang dikembangkan. Validator 3 mengungkapkan “saya sendiri saja membutuhkan waktu yang lama, apa lagi siswanya. Dengan waktu yang diberikan selama 2×40 menit, tidak cukup bagi siswa untuk menyelesaikan 10 soal”. Tidak hanya sampai disitu saja, validator 3 menambahkan bahwa “saya mengerjakan satu soal saja bisa sampai 10 menit, bagaimana dengan siswa saya nantinya yang diberikan soal tersebut”. Dari pernyataan yang disampaikan oleh validator 3, menjelaskan bahwa gurulah yang mengetahui segala sesuatu terkait siswanya.

Menurut Widoyoko, guru memiliki peran penting sebagai penyelenggara pendidikan disekolah⁹. Artinya, guru memiliki tanggungjawab menjadikan siswanya memiliki pengetahuan dan wawasan terkait dengan ilmu pengetahuan. Buchory menambahkan guru memiliki tugas utama yaitu mendidik, membimbing, menilai dan mengevaluasi siswa pada saat disekolah¹⁰. Peranan seorang guru sangat besar dalam lingkungan sekolah sehingga segala sesuatu terkait dengan siswanya dilihat dari kemampuan ataupun cara berpikir siswa, gurulah yang lebih tahu dari yang lain. Hal tersebut tergambar pada validator 3.

Sebagai validator ahli, validator 3 memberikan masukan terkait instrumen penilaian yang dikembangkan berdasarkan pengalamannya di lingkungan sekolah. Guru memiliki peran untuk mengevaluasi siswa terkait materi yang sudah diberikan atau biasa disebut dengan evaluasi pembelajaran. Pengalaman dalam mengevaluasi pembelajaran, dijadikan tolak ukur validator 3 untuk menilai keefektifan dari instrumen penilaian yang akan di uji coba. Berdasarkan pengalaman yang didapat, instrumen penilaian yang dikembangkan dengan waktu yang diberikan tidak terlalu efektif. Dengan 10 butir soal yang

⁹S.E.P. Widoyoko, “Peranan Evaluasi Program Pembelajaran dalam Meningkatkan Kualitas Pendidikan”. (Paper Presented at Seminar Nasional Program Pascasarjana UNY berkerja sama dengan HEPI, Yogyakarta, 2008), 6.

¹⁰Buchory, “Kiat Sukses Sertifikasi”. (Paper Presented at Seminar Nasional Pusat Studi Pendidikan IKA UNY, Yogyakarta, 2008), 1.

diberikan, siswa akan merasa waktu yang diberikan terlalu cepat. Bila dipaksakan dengan waktu yang diberikan, siswa akan menjawab dengan asal-asalan karena waktu pengerjaan yang terlalu sebentar.

Validator 3 memberikan saran terkait dengan instrumen penilaian yang dikembangkan. Saran yang diberikan meliputi dua hal yaitu penambahan waktu pengerjaan atau pengurangan butir soal pada instrumen penilaian yang dikembangkan dengan ketetapan waktu yang diberikan. Segala pertimbangan yang telah dilakukan, hasil diskusi dengan validator 3 adalah mengurangi butir soal yang terdapat pada instrumen penilaian yang dikembangkan dengan waktu pengerjaan yang telah ditentukan. Dengan demikian diberikannya waktu selama 2×40 menit, siswa dapat mengerjakan setidaknya 8 butir soal. Butir soal yang dikeluarkan dari instrumen penilaian yang dikembangkan adalah butir soal nomor 2 dan 7. Hal tersebut sejalan dengan saran yang diberikan oleh validator 3.

Telaah butir soal terakhir, dilakukan oleh salah satu dosen pendidikan matematika. Pemilihan dosen sebagai validator, melengkapi dari kriteria untuk mendapatkan kevalidan dari instrumen penilaian yang dikembangkan. Pengalaman disekolah sudah didapatkan, bahkan sudah mencapai tahap mahasiswa dalam proses pengajaran. Dengan waktu 2×40 menit, validator ke-4 menyatakan kelayakan instrumen penilaian yang dikembangkan untuk di uji coba lapangan. Catatan yang diberikan, hanya terkait redaksi butir soal pada instrumen penilaian. Terkait alokasi waktu yang diberikan, tidak ada komentar sama sekali.

Instrumen penilaian yang dikembangkan, mulanya sebanyak 10 butir soal. Melalui tahap telaah butir soal tes, terdapat 8 butir soal yang dikatakan valid untuk di uji coba lapangan. Alokasi waktu yang diberikan adalah 2×40 menit.

2. **Temuan Terkait Karakteristik Instrumen Penilaian *Higher Order Thinking* untuk Pelajaran Matematika dengan Memasukkan Konteks Islam**

Berdasarkan penyajian dan analisis data karakteristik instrumen penilaian *Higher Order Thinking* untuk Pelajaran Matematika dengan memasukkan konteks Islam, terdapat temuan menarik pada setiap butir soal instrumen penilaian yang

dikembangkan. Temuan yang dimaksudkan terkait dengan karakteristik instrumen penilaian yang meliputi daya pembeda, tingkat kesukaran dan nilai guessing. Berikut temuan terkait karakteristik pada setiap butir soal:

a. Butir Soal 1

Setelah melalui tahap revisi, berikut produk akhir terkait instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam pada butir soal nomor 1:

“Lebaran merupakan salah satu momentum bagi umat Islam untuk mudik atau pulang kampung. Orang-orang rela antri, berdesak-desakan serta macet panjang demi bisa melaksanakan tradisi pulang ke kampung halaman dan berkumpul bersama keluarga. Salah satu alat transportasi yang paling dipadati pada saat mudik adalah Kereta Api. Kereta Api Bima Eksekutif melayani perjalanan dari Kota Surabaya sampai Kota Blitar. Rute awal hingga akhir yang dilalui KA Bima adalah Stasiun Kota Surabaya, Stasiun Kota Sidoarjo, Stasiun Kota Malang dan Stasiun Kota Blitar. KA Bima menyediakan lima gerbong dengan kapasitas per gerbongnya adalah 60 orang. Pernyataan yang mungkin benar berkaitan dengan kondisi diatas adalah. . .

- A. Pada saat pemberangkatan, gerbong terisi penuh. Saat tiba di stasiun Kota Sidoarjo, pemudik tidak ada yang turun. Akan tetapi pada saat tiba di stasiun Kota Malang, pemudik di setiap gerbong turun membentuk pola kelipatan 5. Jika pada saat tiba di stasiun Kota Malang tidak ada penambahan jumlah pemudik, maka hanya ada 220 pemudik yang turun di stasiun terakhir Kota Blitar.
- B. Pada saat pemberangkatan, hanya ada satu gerbong yang kapasitasnya 45. Saat tiba di stasiun Kota Sidoarjo, terdapat penambahan 10 pemudik. Sebelum berangkat ke stasiun berikutnya, ada 50 pemudik yang turun. Setibanya di Kota Malang, jumlah pemudik yang turun sebanyak 65 orang dan penambahan jumlah pemudik sebanyak 45 orang. Sebanyak 230 jumlah pemudik yang akan turun di stasiun akhir Kota Blitar
- C. Pada saat pemberangkatan, ternyata di setiap gerbong tidak terisi penuh. Setiap gerbong terdapat kekosongan

penumpang. Saat tiba distasiun Kota Sidoarjo, hampir setiap gerbong terisi penuh dan hanya ada satu gerbong yang berisi 40 dari kapasitas yang ada. Setibanya di Kota Malang, sebanyak 45 jumlah pemudik yang turun. Tanpa ada penambahan jumlah penumpang di Kota Malang, sebanyak 215 jumlah pemudik yang akan turun di stasiun terakhir.

- D. Pemberangkatan dari stasiun Kota Surabaya, KA Bima membatasi kapasitas jumlah pemudik di setiap gerbong. Hal tersebut dilakukan karena di setiap gerbong, terdapat 5 kursi penumpang yang tidak bisa digunakan. Saat tiba di stasiun Kota Sidoarjo dan stasiun Kota Malang, terdapat 25 orang dan 55 orang jumlah pemudik yang turun. Sebanyak 240 jumlah pemudik yang akan turun di stasiun terakhir dengan penambahan penumpang sebanyak 15 orang dan 35 orang di masing-masing stasiun sebelum sampai di stasiun terakhir."*

Butir soal nomor 1, mempunyai daya pembeda sebesar 5.407. Daya pembeda merupakan kemampuan suatu instrumen penilaian dalam membedakan antara siswa yang sudah memahami materi dengan siswa yang belum memahami materi. Semakin tinggi daya beda suatu item, semakin baik suatu item dalam membedakan kemampuan penempeuh tes¹¹. Daya pembeda yang baik berkisar antara $0 \leq \alpha \leq 2$. Hasil dari butir soal nomor 1, daya pembeda soal lebih dari kisaran yang ditentukan. Hal tersebut menunjukkan, butir soal nomor 1 tidak dapat membedakan antara siswa yang sudah menguasai materi dengan belum menguasai materi. Melihat dari besarnya nilai daya pembeda soal pada butir soal nomor 1, terdapat faktor yang mungkin terjadinya hal tersebut. Menindaklanjuti hasil telaah dari para validator, butir soal nomor 1 mendapatkan banyak komentar. Butir soal nomor 1, dianggap terlalu panjang terkait dengan soal maupun jawaban yang disajikan. Bahkan validator 4, memberikan masukan terhadap jawaban pada poin C yakni membuang satu kalimat pada jawaban agar dapat lebih dipahami. Hal tersebut yang mungkin mendasari tidak

¹¹Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

mampunya soal untuk membedakan antara siswa yang sudah memahami materi dengan siswa yang belum memahami materi melihat dari konten soal maupun jawaban yang terbilang panjang. Sehingga nilai daya pembeda dari butir soal nomor 1, lebih dari kisaran yang ditentukan.

Tingkat kesukaran butir soal nomor 1, memiliki nilai sebesar 0.189. Suatu soal tes hendaknya tidak terlalu sukar dan tidak pula terlalu mudah. Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan bahwa soal tersebut baik¹². Tingkat kesukaran yang baik yaitu memiliki rentang $-2 \leq b \leq 2$. Hasil dari butir soal nomor 1, tingkat kesukaran soal berada pada rentang yang ditentukan. Siswa dengan kemampuan tinggi akan merasa mudah dalam mengerjakan dan sebaliknya siswa dengan kemampuan rendah akan sulit menjawab butir soal nomor 1.

Selanjutnya, nilai *guessing* dari butir soal nomor 1 sebesar 0.166. Soal pilihan ganda sangat rentan dengan menebak (*guessing*) jawaban benar. Kecenderungan siswa dengan kemampuan rendah, menjawab butir soal hanya menebak jawaban yang benar. Nilai *guessing* yang baik yaitu mempunyai rentang $0 \leq c \leq 1$. Song Gao menambahkan, nilai *guessing* kurang dari 0.5, tidak dapat dikatakan jawaban dari soal hanya dari menebak saja¹³. Hasil dari nilai *guessing* pada butir soal nomor 1, menunjukkan kemampuan menebak tidak dapat benar-benar mengetahui kemampuan seseorang. Kebenaran jawaban pada butir soal nomor 1, hasil dari pengetahuan yang dimiliki siswa tanpa ada unsur menebak.

Untuk mengetahui kecocokan dari model 3 parameter logistik, digunakan uji statistik. Uji statistik digunakan untuk menentukan probabilitas perbedaan yang didapatkan di antara nilai yang diperoleh. Uji statistik yang digunakan adalah uji *Chi-Squared*. Nilai signifikansi (*p-value*) dari *Chi-Squared* di bawah taraf signifikansi 0.05 ($p < 0.05$). Butir soal nomor 1, mempunyai nilai 0.0004. Dari hasil tersebut, tidak ada kecocokan dari model 3 parameter logistik terhadap butir soal nomor 1.

¹²Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

¹³ Song Gao, Doctoral Dissertation: "The Exploration of The Relationship Between Guessing and Latent Ability in IRT Models" (Carbondale: Southern Illinois University, 2011), 37.

Untuk mendapatkan karakteristik butir soal dengan kriteria yang baik, hendaknya memenuhi 4 persyaratan yaitu memiliki daya pembeda yang baik ($0 \leq a \leq 2$), tingkat kesukaran yang baik ($-2 \leq b \leq 2$), nilai *guessing* yang baik ($0 \leq c \leq 1$) dan butir soal cocok dengan model. Karakteristik butir soal nomor 1 menunjukkan kriteria yang tidak baik. Karena pada daya pembeda tidak pada rentang yang yang ditentukan dan uji kecocokan model 3 parameter logistik berada pada kurang dari taraf signifikansi.

b. Butir Soal 2

Setelah melalui tahap revisi, berikut produk akhir terkait instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam pada butir soal nomor 2:

“Rahmad, Ali dan Aisah bersahabat sejak kecil. Setelah lulus SMA, mereka berpisah karena mereka masuk perguruan tinggi yang berbeda-beda. Pada saat liburan, mereka merencanakan bertemu. Rahmad berangkat pukul 17.00, Ali berangkat selisih 15 menit dari Rahmad dan Aisah berangkat selisih 15 menit dari Ali. Mereka janji dikota D pukul 19.30. Akan tetapi, hanya Rahmad yang tiba lebih awal pukul 19.00. Transportasi yang mereka gunakan berbeda-beda. Rahmad menggunakan transportasi yang mempunyai kecepatan 300km/jam, Ali menggunakan transportasi yang mempunyai kecepatan 80km/jam dan Aisah menggunakan transportasi yang mempunyai kecepatan 40km/jam. Jika waktu sholat Isya’ memasuki pukul 18.45, maka pernyataan yang benar berkaitan dengan syarat manjamak sholat adalah . . .

- A. Rahmad boleh menjamak sholat sedangkan Ali dan Aisah tidak boleh*
- B. Ali boleh menjamak sholat sedangkan Rahmad dan Aisah tidak boleh*
- C. Rahmad dan Ali boleh menjamak sholat sedangkan Aisah tidak boleh*
- D. Ali dan Aisah boleh menjamak sholat sedangkan Rahmad tidak boleh”*

Butir soal nomor 2, mempunyai daya pembeda sebesar 1.749. Daya pembeda merupakan kemampuan suatu instrumen

penilaian dalam membedakan antara siswa yang sudah memahami materi dengan siswa yang belum memahami materi. Semakin tinggi daya beda suatu item, semakin baik suatu item dalam membedakan kemampuan penempuh tes¹⁴. Daya pembeda yang baik berkisar antara $0 \leq \alpha \leq 2$. Hasil dari butir soal nomor 2, daya pembeda soal berada kisaran yang ditentukan. Dengan demikian, butir soal nomor 2 dapat membedakan antara siswa yang sudah memahami materi dengan siswa yang belum memahami materi.

Tingkat kesukaran butir soal nomor 2, memiliki nilai sebesar 0.919. Suatu soal tes hendaknya tidak terlalu sukar dan tidak pula terlalu mudah. Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan bahwa soal tersebut baik¹⁵. Tingkat kesukaran yang baik yaitu memiliki rentang $-2 \leq b \leq 2$. Hasil dari butir soal nomor 2, tingkat kesukaran soal berada pada rentang yang ditentukan. Siswa dengan kemampuan tinggi akan merasa mudah dalam mengerjakan dan sebaliknya siswa dengan kemampuan rendah akan sulit menjawab butir soal nomor 2.

Selanjutnya, nilai *guessing* dari butir soal nomor 2 sebesar 0.124. Soal pilihan ganda sangat rentan dengan menebak (*guessing*) jawaban benar. Kecenderungan siswa dengan kemampuan rendah, menjawab butir soal hanya menebak jawaban yang benar. Nilai *guessing* yang baik yaitu mempunyai rentang $0 \leq c \leq 1$. Song Gao menambahkan, nilai *guessing* kurang dari 0.5, tidak dapat dikatakan jawaban dari soal hanya dari menebak saja¹⁶. Hasil dari nilai *guessing* pada butir soal nomor 2, menunjukkan kemampuan menebak tidak dapat benar-benar mengetahui kemampuan seseorang. Kebenaran jawaban pada butir soal nomor 2, hasil dari pengetahuan yang dimiliki siswa tanpa ada unsur menebak.

Untuk mengetahui kecocokan dari model 3 parameter logistik, digunakan uji statistik. Uji statistik digunakan untuk menentukan probabilitas perbedaan yang didapatkan di antara

¹⁴ Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

¹⁵ Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

¹⁶ Song Gao, Doctoral Dissertation: “*The Exploration of The Relationship Between Guessing and Latent Ability in IRT Models*” (Carbondale: Southern Illinois University, 2011), 37.

nilai yang diperoleh. Uji statistik yang digunakan adalah uji *Chi-Squared*. Nilai signifikansi (*p-value*) dari *Chi-Squared* di bawah taraf signifikansi 0.05 ($p < 0.05$). Butir soal nomor 2, mempunyai nilai 0.0026. Dari hasil tersebut, tidak ada kecocokan dari model 3 parameter logistik terhadap butir soal nomor 2.

Untuk mendapatkan karakteristik butir soal dengan kriteria yang baik, hendaknya memenuhi 4 persyaratan yaitu memiliki daya pembeda yang baik ($0 \leq a \leq 2$), tingkat kesukaran yang baik ($-2 \leq b \leq 2$), nilai *guessing* yang baik ($0 \leq c \leq 1$) dan butir soal cocok dengan model. Karakteristik butir soal nomor 2 menunjukkan kriteria yang tidak baik. Karena pada uji kecocokan model 3 parameter logistik berada pada kurang dari taraf signifikansi.

c. Butir Soal 3

Setelah melalui tahap revisi, berikut produk akhir terkait instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam pada butir soal nomor 3:

“Adinda, Fitri, Siti dan Aisah merupakan santriwati di ponpes Darul ‘Ulum Jombang. Mereka juga sekolah disalah satu Madrasah Tsanawiyah dipondok tersebut. Untuk menyelesaikan studi mereka, mereka harus dapat menghafal juz 30. Tes hafalan dilaksanakan satu minggu setelah Ujian Nasional. Tiga minggu sebelum tes hafalan berlangsung, jumlah surat yang Adinda hafal 5 kali banyaknya surat yang dihafal Fitri. Jumlah surat yang dihafal Fitri 2 kali lebih banyak dari Siti. Sedangkan jumlah surat yang dihafal Siti tiga lebihnya dari jumlah surat Aisah. Pernyataan yang benar berkaitan dengan kondisi diatas adalah. . .

- A. Jumlah surat yang Adinda hafal, tujuh kali jumlah surat yang Siti hafal*
- B. Jumlah surat yang Siti hafal, lima lebih banyak jumlah surat yang Aisah hafal.*
- C. Jumlah surat yang Aisah hafal, enam lebih sedikit dari jumlah surat yang Fitri hafal.*
- D. Jumlah surat yang Fitri hafal, enam lebih banyak dan dua kali jumlah surat yang Aisah hafal.”*

Nilai dari setiap karakteristik butir soal nomor 3, tidak dapat diketahui. Hal tersebut dikarenakan, butir soal nomor 3 gugur saat proses awal analisis. Gugurnya butir soal nomor 3 terkait dengan nilai korelasi biserial yang bernilai negatif. Nilai korelasi biserial pada butir soal nomor 3 adalah -0.016 .

Menurut Ebel & Fresbie, korelasi biserial menggambarkan hubungan antara skor pada item tes dan skor pada total tes untuk setiap *testee*. Biserial yang bernilai positif tinggi menggambarkan kecenderungan *testee* yang berskor tinggi untuk menjawab benar dan *testee* yang berskor salah memberikan jawaban. Nilai biserial yang negatif menggambarkan hubungan yang berlawanan, dimana *testee* yang memiliki skor tinggi salah menjawab item tersebut, sedangkan *testee* dengan skor rendah benar menjawab item yang bersangkutan. Nilai korelasi biserial yang kurang dari 0.2 merupakan item yang tidak bagus¹⁷.

Dari pendapat yang dikemukakan, butir soal yang memiliki korelasi biserial kurang dari 0.2, tidak ikut di analisis karena akan mengganggu proses analisis pada program BILOG. Program BILOG tidak dapat melanjutkan analisis bila ada hasil yang tidak signifikan. Dengan demikian, butir soal nomor 3, dikeluarkan dari proses analisis.

d. Butir Soal 4

Setelah melalui tahap revisi, berikut produk akhir terkait instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam pada butir soal nomor 4:

“Perhatikan pernyataan berikut

- (i) *Sudut yang dibentuk oleh pertemuan pangkal dua buah garis lurus yang saling tegak lurus merupakan gerakan Ruku’ dalam shalat.*
- (ii) *Jika dalam satu rakaat gerakan shalat Ruku’, I’tidal dan Sujud dijumlahkan, maka sudut yang dihasilkan kurang dari 360^0 .*
- (iii) *Sudut yang dihasilkan dalam gerakan Sujud adalah setengah dari gerakan Ruku’.*

¹⁷R.L. Ebel & D.A. Frisbie, *Essentials Of Educational Measurement* (New Jersey: Prentice-Hall, Inc, 1986), 234.

(iv) Sudut yang dihasilkan dalam satu rakaat gerakan shalat Sujud merupakan sudut yang sama dalam gerakan shalat I'tidal.

Pernyataan diatas yang benar adalah...

- A. (i) dan (ii)
- B. (ii) dan (iii)
- C. (iii) dan (iv)
- D. (i) dan (iii)"

Butir soal nomor 4, mempunyai daya pembeda sebesar 0.805. Daya pembeda merupakan kemampuan suatu instrumen penilaian dalam membedakan antara siswa yang sudah memahami materi dengan siswa yang belum memahami materi. Semakin tinggi daya beda suatu item, semakin baik suatu item dalam membedakan kemampuan penempuh tes¹⁸ Daya pembeda yang baik berkisar antara $0 \leq a \leq 2$. Hasil dari butir soal nomor 4, daya pembeda soal berada kisaran yang ditentukan. Dengan demikian, butir soal nomor 4 dapat membedakan antara siswa yang sudah memahami materi dengan siswa yang belum memahami materi.

Tingkat kesukaran butir soal nomor 4, memiliki nilai sebesar 1.996. Suatu soal tes hendaknya tidak terlalu sukar dan tidak pula terlalu mudah. Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan bahwa soal tersebut baik¹⁹. Tingkat kesukaran yang baik yaitu memiliki rentang $-2 \leq b \leq 2$. Hasil dari butir soal nomor 4, tingkat kesukaran soal berada pada rentang yang ditentukan. Siswa dengan kemampuan tinggi akan merasa mudah dalam mengerjakan dan sebaliknya siswa dengan kemampuan rendah akan sulit menjawab butir soal nomor 4.

Selanjutnya, nilai *guessing* dari butir soal nomor 4 sebesar 0.228. Soal pilihan ganda sangat rentan dengan menebak (*guessing*) jawaban benar. Kecenderungan siswa dengan kemampuan rendah, menjawab butir soal hanya menebak jawaban yang benar. Nilai *guessing* yang baik yaitu mempunyai rentang $0 \leq c \leq 1$. Song Gao menambahkan, nilai *guessing*

¹⁸Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

¹⁹Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

kurang dari 0.5, tidak dapat dikatakan jawaban dari soal hanya dari menebak saja²⁰. Hasil dari nilai *guessing* pada butir soal nomor 4, menunjukkan kemampuan menebak tidak dapat benar-benar mengetahui kemampuan seseorang. Kebenaran jawaban pada butir soal nomor 4, hasil dari pengetahuan yang dimiliki siswa tanpa ada unsur menebak.

Untuk mengetahui kecocokan dari model 3 parameter logistik, digunakan uji statistik. Uji statistik digunakan untuk menentukan probabilitas perbedaan yang didapatkan di antara nilai yang diperoleh. Uji statistik yang digunakan adalah uji *Chi-Squared*. Nilai signifikansi (*p-value*) dari *Chi-Squared* di bawah taraf signifikansi 0.05 ($p < 0.05$). Butir soal nomor 4, mempunyai nilai 0.4125. Dari hasil tersebut, ada kecocokan dari model 3 parameter logistik terhadap butir soal nomor 4.

Untuk mendapatkan karakteristik butir soal dengan kriteria yang baik, hendaknya memenuhi 4 persyaratan yaitu memiliki daya pembeda yang baik ($0 \leq a \leq 2$), tingkat kesukaran yang baik ($-2 \leq b \leq 2$), nilai *guessing* yang baik ($0 \leq c \leq 1$) dan butir soal cocok dengan model. Karakteristik butir soal nomor 4 menunjukkan kriteria yang baik.

e. Butir Soal 5

Setelah melalui tahap revisi, berikut produk akhir terkait instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam pada butir soal nomor 5:

"Hari ini Bu Nisa' membagikan hasil tes Hafalan Al-Qur'an Juz 30 kepada siswanya yang berjumlah 6 orang. Ahmad mendapatkan nilai 2 kali lebih banyak dari pada Kevin. Sedangkan Aini mendapatkan nilai satu lebih banyak dari Ahmad. Fitri sahabat Aini, mendapatkan dua kurang dari sahabatnya. Ali tidak mengikuti tes hafalan ulang, karena nilainya sama seperti nilai standart kelulusan Ujian Pondok Pesantren. Faizien hampir mendapatkan nilai sempurna, akan tetapi terdapat hafalan yang kurang lancar pada awal surat juz 30. Sehingga Faizien mendapatkan satu lebih sedikit dari nilai sempurna. Nilai yang didapatkan Kevin pada saat tes

²⁰Song Gao, Doctoral Dissertation: "*The Exploration of The Relationship Between Guessing and Latent Ability in IRT Models*" (Carbondale: Southern Illinois University, 2011), 37.

hafalan Al-Qur'an bernilai 4 dan nilai standart kelulusan Pondok Pesantren adalah tiga lebih banyak dari nilai kevin. Modus dari nilai hasil tes hafalan Al-Qur'an Juz 30 yang dibagikan Bu Nisa' kepada siswanya adalah. . .

- A. 7
- B. 9
- C. 7 dan 9
- D. Tidak mempunyai modus"

Butir soal nomor 5, mempunyai daya pembeda sebesar 0.848. Daya pembeda merupakan kemampuan suatu instrumen penilaian dalam membedakan antara siswa yang sudah memahami materi dengan siswa yang belum memahami materi. Semakin tinggi daya beda suatu item, semakin baik suatu item dalam membedakan kemampuan penempuh tes²¹. Daya pembeda yang baik berkisar antara $0 \leq a \leq 2$. Hasil dari butir soal nomor 5, daya pembeda soal berada kisaran yang ditentukan. Dengan demikian, butir soal nomor 5 dapat membedakan antara siswa yang sudah memahami materi dengan siswa yang belum memahami materi.

Tingkat kesukaran butir soal nomor 5, memiliki nilai sebesar 1.895. Suatu soal tes hendaknya tidak terlalu sukar dan tidak pula terlalu mudah. Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan bahwa soal tersebut baik²². Tingkat kesukaran yang baik yaitu memiliki rentang $-2 \leq b \leq 2$. Hasil dari butir soal nomor 5, tingkat kesukaran soal berada pada rentang yang ditentukan. Siswa dengan kemampuan tinggi akan merasa mudah dalam mengerjakan dan sebaliknya siswa dengan kemampuan rendah akan sulit menjawab butir soal nomor 5.

Selanjutnya, nilai *guessing* dari butir soal nomor 5 sebesar 0.244. Soal pilihan ganda sangat rentan dengan menebak (*guessing*) jawaban benar. Kecenderungan siswa dengan kemampuan rendah, menjawab butir soal hanya menebak jawaban yang benar. Nilai *guessing* yang baik yaitu mempunyai rentang $0 \leq c \leq 1$. Song Gao menambahkan, nilai *guessing*

²¹Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

²²Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

kurang dari 0.5, tidak dapat dikatakan jawaban dari soal hanya dari menebak saja.²³ Hasil dari nilai *guessing* pada butir soal nomor 5, menunjukkan kemampuan menebak tidak dapat benar-benar mengetahui kemampuan seseorang. Kebenaran jawaban pada butir soal nomor 5, hasil dari pengetahuan yang dimiliki siswa tanpa ada unsur menebak.

Untuk mengetahui kecocokan dari model 3 parameter logistik, digunakan uji statistik. Uji statistik digunakan untuk menentukan probabilitas perbedaan yang didapatkan di antara nilai yang diperoleh. Uji statistik yang digunakan adalah uji *Chi-Squared*. Nilai signifikansi (*p-value*) dari *Chi-Squared* di bawah taraf signifikansi 0.05 ($p < 0.05$). Butir soal nomor 5, mempunyai nilai 0.3323. Dari hasil tersebut, ada kecocokan dari model 3 parameter logistik terhadap butir soal nomor 5.

Untuk mendapatkan karakteristik butir soal dengan kriteria yang baik, hendaknya memenuhi 4 persyaratan yaitu memiliki daya pembeda yang baik ($0 \leq a \leq 2$), tingkat kesukaran yang baik ($-2 \leq b \leq 2$), nilai *guessing* yang baik ($0 \leq c \leq 1$) dan butir soal cocok dengan model. Karakteristik butir soal nomor 5 menunjukkan kriteria yang baik.

f. Butir Soal 6

Setelah melalui tahap revisi, berikut produk akhir terkait instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam pada butir soal nomor 6:

“Ahmad, Alawi dan Ulfa masing-masing mempunyai uang sebanyak Rp. 600.000,00, Rp. 650.000,00 dan Rp. 700.000,00. Pada suatu hari, ketiga sahabat tersebut melakukan musahammah (investasi) kepada perusahaan ternama dikotanya. Komisi yang diberikan oleh perusahaan sebanyak 10% per bulan. Ketiga sahabat bernadzar, uang hasil nisbah (komisi) yang didapatkan dalam satu tahun akan disumbangkan kepada masjid di desanya. Ahmad menyumbang sebanyak 5%, alawi sebanyak 15%, dan ulfa sebanyak 10% dari komisi yang didupatkannya selama 1

²³Song Gao, Doctoral Dissertation: “*The Exploration of The Relationship Between Guessing and Latent Ability in IRT Models*” (Carbondale: Southern Illinois University, 2011), 37.

tahun. Pernyataan yang benar berkaitan dengan kondisi diatas adalah. . .

- A. Sisa uang Ahmad lebih besar dari Alawi dan lebih sedikit dari Ulfa*
- B. Sisa uang Ahmad lebih besar daripada Ulfa dan lebih sedikit daripada Alawi*
- C. Sisa uang Alawi lebih sedikit daripada Ulfa dan lebih banyak daripada Ahmad*
- D. Sisa uang Ulfa lebih banyak daripada Alawi dan lebih sedikit daripada Ahmad”*

Butir soal nomor 6, mempunyai daya pembeda sebesar 0.657. Daya pembeda merupakan kemampuan suatu instrumen penilaian dalam membedakan antara siswa yang sudah memahami materi dengan siswa yang belum memahami materi. Semakin tinggi daya beda suatu item, semakin baik suatu item dalam membedakan kemampuan penempuh tes²⁴. Daya pembeda yang baik berkisar antara $0 \leq \alpha \leq 2$. Hasil dari butir soal nomor 6, daya pembeda soal berada kisaran yang ditentukan. Dengan demikian, butir soal nomor 6 dapat membedakan antara siswa yang sudah memahami materi dengan siswa yang belum memahami materi.

Tingkat kesukaran butir soal nomor 6, memiliki nilai sebesar 2.448. Suatu soal tes hendaknya tidak terlalu sukar dan tidak pula terlalu mudah. Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan bahwa soal tersebut baik²⁵. Tingkat kesukaran yang baik yaitu memiliki rentang $-2 \leq b \leq 2$. Hasil dari butir soal nomor 6, tingkat kesukaran soal lebih dari kisaran yang ditentukan. Hal tersebut menunjukkan, siswa dengan kemampuan tinggi merasa kesulitan dalam mengerjakan dan sebaliknya siswa dengan kemampuan rendah merasa mudah menjawab butir soal nomor 6. Dilihat dari perolehan menjawab benar, sebanyak 72 siswa menjawab benar pada butir soal nomor 6. Mungkin saja, siswa yang menjawab kebanyakan adalah siswa dengan kemampuan rendah.

²⁴Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

²⁵Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

Selanjutnya, nilai *guessing* dari butir soal nomor 6 sebesar 0.276. Soal pilihan ganda sangat rentan dengan menebak (*guessing*) jawaban benar. Kecenderungan siswa dengan kemampuan rendah, menjawab butir soal hanya menebak jawaban yang benar. Nilai *guessing* yang baik yaitu mempunyai rentang $0 \leq c \leq 1$. Song Gao menambahkan, nilai *guessing* kurang dari 0.5, tidak dapat dikatakan jawaban dari soal hanya dari menebak saja²⁶. Hasil dari nilai *guessing* pada butir soal nomor 6, menunjukkan kemampuan menebak tidak dapat benar-benar mengetahui kemampuan seseorang. Kebenaran jawaban pada butir soal nomor 6, hasil dari pengetahuan yang dimiliki siswa tanpa ada unsur menebak.

Untuk mengetahui kecocokan dari model 3 parameter logistik, digunakan uji statistik. Uji statistik digunakan untuk menentukan probabilitas perbedaan yang didapatkan di antara nilai yang diperoleh. Uji statistik yang digunakan adalah uji *Chi-Squared*. Nilai signifikansi (*p-value*) dari *Chi-Squared* di bawah taraf signifikansi 0.05 ($p < 0.05$). Butir soal nomor 6, mempunyai nilai 0.1368. Dari hasil tersebut, tidak ada kecocokan dari model 3 parameter logistik terhadap butir soal nomor 6.

Untuk mendapatkan karakteristik butir soal dengan kriteria yang baik, hendaknya memenuhi 4 persyaratan yaitu memiliki daya pembeda yang baik ($0 \leq a \leq 2$), tingkat kesukaran yang baik ($-2 \leq b \leq 2$), nilai *guessing* yang baik ($0 \leq c \leq 1$) dan butir soal cocok dengan model. Karakteristik butir soal nomor 6 menunjukkan kriteria yang tidak baik. Karena pada tingkat kesukaran tidak pada rentang yang yang ditentukan.

g. Butir Soal 7

Setelah melalui tahap revisi, berikut produk akhir terkait instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam pada butir soal nomor 7:

“Lilah mendapatkan kesempatan untuk menyempurnakan rukun islamnya dengan menuaikan ibadah Haji. Kesempatan ini didapatkannya setelah dia menabung selama bertahun-tahun. Diapun langsung mendaftarkan dirinya ke Kementerian Agama.

²⁶Song Gao, Doctoral Dissertation: “*The Exploration of The Relationship Between Guessing and Latent Ability in IRT Models*” (Carbondale: Southern Illinois University, 2011), 37.

Setelah melakukan pendaftaran dan administrasi, petugas pendaftaran mengatakan kepada Lilah bahwa “Dalam dua puluh tahun kedepan, peluang keberangkatan Haji adalah lima per enam.” Pernyataan di bawah ini yang benar mencerminkan maksud dari petugas pendaftaran adalah . . .

- A. $\frac{5}{6} \times 20 = 16,7$, sehingga antara 16 dan 17 dari tahun sekarang Lilah akan menuaikan ibadah Haji.
- B. $\frac{5}{6}$ lebih besar daripada $\frac{1}{2}$, sehingga dapat diyakini bahwa Lilah akan menuaikan ibadah Haji dalam 20 tahun kedepan.
- C. Peluang keberangkatan Lilah menuaikan ibadah Haji dalam 20 tahun kedepan lebih tinggi daripada peluang tidak terjadinya keberangkatan.
- D. Kita dapat menyakini bahwa Haji merupakan panggilan. Meskipun kita sudah mempersiapkan segalanya dari mulai pendaftaran, menunggu sampai tahun keberangkatan, kemudian melunasi seluruh biaya yang ada dan sampai waktunya kita masuk daftar jama'ah Haji yang berangkat tahun ini, belum tentu kita benar-benar dipanggil oleh Allah ke Baitullah.”

Butir soal nomor 7, mempunyai daya pembeda sebesar 1.333. Daya pembeda merupakan kemampuan suatu instrumen penilaian dalam membedakan antara siswa yang sudah memahami materi dengan siswa yang belum memahami materi. Semakin tinggi daya beda suatu item, semakin baik suatu item dalam membedakan kemampuan penempuh tes²⁷. Daya pembeda yang baik berkisar antara $0 \leq \alpha \leq 2$. Hasil dari butir soal nomor 7, daya pembeda soal berada kisaran yang ditentukan. Dengan demikian, butir soal nomor 7 dapat membedakan antara siswa yang sudah memahami materi dengan siswa yang belum memahami materi.

Tingkat kesukaran butir soal nomor 7, memiliki nilai sebesar 1.206. Suatu soal tes hendaknya tidak terlalu sukar dan tidak pula terlalu mudah. Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan

²⁷Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

bahwa soal tersebut baik²⁸. Tingkat kesukaran yang baik yaitu memiliki rentang $-2 \leq b \leq 2$. Hasil dari butir soal nomor 7, tingkat kesukaran soal berada pada rentang yang ditentukan. Siswa dengan kemampuan tinggi akan merasa mudah dalam mengerjakan dan sebaliknya siswa dengan kemampuan rendah akan sulit menjawab butir soal nomor 7.

Selanjutnya, nilai *guessing* dari butir soal nomor 7 sebesar 0.132. Soal pilihan ganda sangat rentan dengan menebak (*guessing*) jawaban benar. Kecenderungan siswa dengan kemampuan rendah, menjawab butir soal hanya menebak jawaban yang benar. Nilai *guessing* yang baik yaitu mempunyai rentang $0 \leq c \leq 1$. Song Gao menambahkan, nilai *guessing* kurang dari 0.5, tidak dapat dikatakan jawaban dari soal hanya dari menebak saja.²⁹ Hasil dari nilai *guessing* pada butir soal nomor 7, menunjukkan kemampuan menebak tidak dapat benar-benar mengetahui kemampuan seseorang. Kebenaran jawaban pada butir soal nomor 7, hasil dari pengetahuan yang dimiliki siswa tanpa ada unsur menebak.

Untuk mengetahui kecocokan dari model 3 parameter logistik, digunakan uji statistik. Uji statistik digunakan untuk menentukan probabilitas perbedaan yang didapatkan di antara nilai yang diperoleh. Uji statistik yang digunakan adalah uji *Chi-Squared*. Nilai signifikansi (*p-value*) dari *Chi-Squared* di bawah taraf signifikansi 0.05 ($p < 0.05$). Butir soal nomor 7, mempunyai nilai 0.1128. Dari hasil tersebut, ada kecocokan dari model 3 parameter logistik terhadap butir soal nomor 7.

Untuk mendapatkan karakteristik butir soal dengan kriteria yang baik, hendaknya memenuhi 4 persyaratan yaitu memiliki daya pembeda yang baik ($0 \leq a \leq 2$), tingkat kesukaran yang baik ($-2 \leq b \leq 2$), nilai *guessing* yang baik ($0 \leq c \leq 1$) dan butir soal cocok dengan model. Karakteristik butir soal nomor 7 menunjukkan kriteria yang baik.

h. Butir Soal 8

²⁸Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

²⁹Song Gao, Doctoral Dissertation: “*The Exploration of The Relationship Between Guessing and Latent Ability in IRT Models*” (Carbondale: Southern Illinois University, 2011), 37.

Setelah melalui tahap revisi, berikut produk akhir terkait instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam pada butir soal nomor 8:

“Masjid Al-Akbar Surabaya mengadakan kegiatan buka bersama dengan yatim piatu dan dhuafa. Acara tersebut diikuti lebih dari 1000 peserta. Panitia kegiatan mengatur tempat duduk sedemikian rupa di mana tempat duduk yang diatur membentuk suatu pola barisan. Barisan pertama terdapat 2 kursi. Barisan kedua dan ketiga terdiri dari 6 dan 12 kursi. Barisan keempat terdiri dari 20 kursi. Ahmad, Imron dan Firman merupakan peserta kegiatan buka bersama. Ahmad berada di barisan ke sepuluh. Imron berada di barisan ke dua belas. Sedangkan Firman berada di barisan ke empat belas. Banyaknya kursi pada barisan yang ditempati Ahmad, Imron dan Firman pada kegiatan buka bersama adalah...

- A. 110, 132, 156
- B. 132, 156, 182
- C. 110, 156, 210
- D. 156, 182, 210”

Butir soal nomor 8, mempunyai daya pembeda sebesar 1.010. Daya pembeda merupakan kemampuan suatu instrumen penilaian dalam membedakan antara siswa yang sudah memahami materi dengan siswa yang belum memahami materi. Semakin tinggi daya beda suatu item, semakin baik suatu item dalam membedakan kemampuan penempuh tes³⁰. Daya pembeda yang baik berkisar antara $0 \leq \alpha \leq 2$. Hasil dari butir soal nomor 8, daya pembeda soal berada kisaran yang ditentukan. Dengan demikian, butir soal nomor 8 dapat membedakan antara siswa yang sudah memahami materi dengan siswa yang belum memahami materi.

Tingkat kesukaran butir soal nomor 8, memiliki nilai sebesar -0.265 . Suatu soal tes hendaknya tidak terlalu sukar dan tidak pula terlalu mudah. Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan

³⁰Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

bahwa soal tersebut baik³¹. Tingkat kesukaran yang baik yaitu memiliki rentang $-2 \leq b \leq 2$. Hasil dari butir soal nomor 8, tingkat kesukaran soal berada pada rentang yang ditentukan. Siswa dengan kemampuan tinggi akan merasa mudah dalam mengerjakan dan sebaliknya siswa dengan kemampuan rendah akan sulit menjawab butir soal nomor 8.

Selanjutnya, nilai *guessing* dari butir soal nomor 8 sebesar 0.221. Soal pilihan ganda sangat rentan dengan menebak (*guessing*) jawaban benar. Kecenderungan siswa dengan kemampuan rendah, menjawab butir soal hanya menebak jawaban yang benar. Nilai *guessing* yang baik yaitu mempunyai rentang $0 \leq c \leq 1$. Song Gao menambahkan, nilai *guessing* kurang dari 0.5, tidak dapat dikatakan jawaban dari soal hanya dari menebak saja.³² Hasil dari nilai *guessing* pada butir soal nomor 8, menunjukkan kemampuan menebak tidak dapat benar-benar mengetahui kemampuan seseorang. Kebenaran jawaban pada butir soal nomor 8, hasil dari pengetahuan yang dimiliki siswa tanpa ada unsur menebak.

Untuk mengetahui kecocokan dari model 3 parameter logistik, digunakan uji statistik. Uji statistik digunakan untuk menentukan probabilitas perbedaan yang didapatkan di antara nilai yang diperoleh. Uji statistik yang digunakan adalah uji *Chi-Squared*. Nilai signifikansi (*p-value*) dari *Chi-Squared* di bawah taraf signifikansi 0.05 ($p < 0.05$). Butir soal nomor 8, mempunyai nilai 0.0039. Dari hasil tersebut, tidak ada kecocokan dari model 3 parameter logistik terhadap butir soal nomor 8.

Untuk mendapatkan karakteristik butir soal dengan kriteria yang baik, hendaknya memenuhi 4 persyaratan yaitu memiliki daya pembeda yang baik ($0 \leq a \leq 2$), tingkat kesukaran yang baik ($-2 \leq b \leq 2$), nilai *guessing* yang baik ($0 \leq c \leq 1$) dan butir soal cocok dengan model. Karakteristik butir soal nomor 8 menunjukkan kriteria yang tidak baik. Karena pada uji kecocokan model 3 parameter logistik berada pada kurang dari taraf signifikasi.

³¹Zaenal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2009).

³²Song Gao, Doctoral Dissertation: “*The Exploration of The Relationship Between Guessing and Latent Ability in IRT Models*” (Carbondale: Southern Illinois University, 2011), 37.

BAB V

PENUTUP

A. Simpulan

Penelitian ini menghasilkan instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam. Berdasarkan hasil analisis yang telah peneliti lakukan, dapat diambil kesimpulan sebagai berikut:

1. Proses pengembangan instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam dikembangkan dengan model pengembangan Mardapi. Terdapat temuan menarik pada saat proses pengembangan yakni pada langkah menelaah soal tes dan menganalisis butir soal tes. Pada langkah menelaah soal tes, awalnya jumlah instrumen penilaian yang dikembangkan sejumlah 10 butir soal. Melalui tahap telaah butir soal tes, terdapat 8 butir soal yang dikatakan valid dengan alokasi waktu selama 2×40 menit dan layak untuk di uji coba. Berikutnya pada langkah menganalisis butir soal tes. Terdapat dua temuan topik yang menarik pada langkah ini. Pertama, butir soal nomor 3 dikeluarkan dalam tahap analisis karakteristik instrumen dikarenakan korelasi biserial yang bernilai negatif. Kedua, hasil dari estimasi perhitungan karakteristik instrumen penilaian didapatkan 3 karakteristik butir soal yang baik yaitu butir soal nomor 3,4 dan 6 yang lolos uji kecocokan model.
2. Karakteristik instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam yang dikembangkan adalah sebagai berikut
 - a. Daya pembeda instrumen penilaian yang termasuk dalam kategori baik dimiliki oleh 6 instrumen penilaian dan kategori tidak baik hanya 1 instrumen penilaian. Instrumen penilaian dengan kategori baik berada pada rentang 0 sampai 2.
 - b. Tingkat kesukaran instrumen penilaian yang termasuk dalam kategori baik dimiliki oleh 6 instrumen penilaian dan kategori tidak baik hanya 1 instrumen penilaian.

Instrumen penilaian dengan kategori baik berada rentang -2 sampai 2.

- c. Nilai *guessing* dari ke-7 instrumen penilaian termasuk kategori baik karena berada pada rentang 0 sampai 1.

B. Saran

Berdasarkan simpulan yang disampaikan peneliti, terdapat beberapa saran yang diberikan diantaranya:

1. Pengembangan instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam hanya sampai pada langkah mengetahui karakteristik instrumen penilaian yang dikembangkan. Bagi peneliti yang akan mengembangkan penelitian ini, disarankan melakukan penelitian lanjutan mengenai penerapan dan pengaruh instrumen penilaian yang dikembangkan terhadap kemampuan tertentu.
2. Produk yang berupa instrumen penilaian *Higher Order Thinking* untuk pelajaran matematika dengan memasukkan konteks Islam ini, dapat dijadikan contoh atau acuan guru matematika Madrasah Tsanawiyah (MTs) untuk membuat instrumen penilaian soal matematika yang dikaitkan dengan konteks Islam.
3. Bagi peneliti yang ingin mengembangkan penelitian ini, disarankan bentuk soal yang dikembangkan nantinya lebih bervariasi tidak hanya pada bentuk soal pilihan ganda saja.

DAFTAR PUSTAKA

- Abosalem, Yousef. "Assessment Techniques And Students' Higher-Order Thinking Skills". *International Journal of Secondary Education*. Vol. 4 No. 1, Maret, 2016.
- Aida, Nur., Skipsi: "Pengembangan Instrumen Penilaian Hasil Belajar Kognitif Matematika Berdasarkan Model Pisa Pada Konten Quantity". Surabaya: Universitas Islam Negeri Sunan Ampel Surabaya, 2017.
- Arifin, Zaenal. *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya, 2009.
- Arifin, Zaenal., - Heri Retnawati, "Analisis Instrumen Pengukur Higher Order Thinking Skills (HOTS) Matematika Siswa SMA". Paper presented at Seminar Nasional Matematika Dan Pendidikan Matematika, Universitas Negeri Yogyakarta, 2015.
- Arikunto, Suharsimi. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara, 2010.
- Baker, Frank B. *The Basics of Item Response Theory*. United States of Amerika: ERIC Clearinghouse on Assessment and Evaluation, 2001.
- Belawati, Tian. *Pengembangan Bahan Ajar*. Jakarta: Universitas Terbuka, 2003.
- Buchory., "Kiat Sukses Sertifikasi". Paper Presented at Seminar Nasional Pusat Studi Pendidikan IKA UNY, Yogyakarta, 2008.
- Budiman, Agus., – Jailani. "Pengembangan Instrumen Asesmen Higher Order Thinking Skill (HOTS) Pada Mata Pelajaran Matematika SMP Kelas VIII Semester 1". *Jurnal Riset Pendidikan Matematika*. Vol. 1 No. 2, November 2014.
- Ebel, R.L., - D.A. Frisbie. *Essentials Of Educational Measurement*. New Jersey: Prentice-Hall, Inc, 1986.
- Fatkhudin, Aslam., et.al. "Item Response Theory Model Empat Parameter Logistik Pada Computerized Adaptive Test". *Jurnal Sistem Informasi Bisnis*. Vol. 2, Maret 2014.
- Gao., Song, Doctoral Dissertation: "The Exploration of The Relationship Between Guessing and Latent Ability in IRT Models". Carbondale: Southern Illinois University, 2011.
- Hambleton, R. K., & Swaminathan, H. *Item response theory: Principles and applications*. Springer Science & Business Media, 2013.

- Haniffah, Distariana., - Janet Trineke Manoy. "Identifikasi Tipe Berpikir Dengan Soal Higher Order Thinking (Hot) Ditinjau Berdasarkan Kemampuan Matematika". *Mathedunesa Jurnal Ilmiah Pendidikan Matematika*. Vol. 3 No.3, 2014.
- Hartini., Disertasi Doktor: "*Pengembangan Higher Order Thinking Multiple Choice Test Untuk Mengukur Keterampilan Berpikir Kritis Ipa Kelas VII SMP/Mts*". Yogyakarta: Universitas Negeri Yogyakarta, 2013.
- Hartini – Sukardjo. "Pengembangan Higher Order Thinking Multiple Choice Test Untuk Mengukur Keterampilan Berpikir Kritis IPA Kelas VII SMP/MTs". *Jurnal Inovasi Pendidikan IPA*. Vol. 1 No. 1, April, 2015.
- IAIN Sunan Ampel Press. *Konsep Dasar Fiqih Muamalah*. Sidoarjo: CV Cahaya Intan, 2014.
- Jamil, Ahmad., et.al. *Sarana Penunjang Aktif Belajar Al Fath*. Gresik: CV. Putra Kembar Jaya, 2007.
- Kawuwung, Femmy. "Profil Guru, Pemahaman Kooperatif NHT, dan Kemampuan Berpikir Tingkat Tinggi Di Smp Kabupaten Minahasa Utara". *El-Hayah*. Vol. 1 No. 4, Maret, 2011.
- Kementerian Pendidikan Dan Kebudayaan. *Modul Teknik Penulisan Soal Berpikir Orde Lebih Tinggi/Higher Order Thinking (HOT)*, 2016.
- King, Fj., et.al. "Higher Order Thinking Skills". *Center for Advancement of Learning and Assessment*.1998.
- Komarudin, A, "*Analisis Tipe Berpikir Dengan Soal Higher Order Thinking Ditinjau Berdasarkan Kemampuan Matematika Siswa*". Paper presented at Seminar Nasional Matematika Dan Pendidikan Matematika Universitas Yogyakarta, Yogyakarta, 2015.
- Krathwohl, D. R. "A revision of Bloom's taxonomy: An overview". *Theory into practice*, Vol. 41 No. 4, 2002.
- Kubiszyn – Borich, *Educational testing and measurement classroom application and practice (seventh edition.)*. New York: John Wiley & Sons, Inc, 2002.
- Kurniawan, Devi Dwi, "*Analisis Kualitas Soal Ujian Akhir Semester Matematika Berdasarkan Teori Respon Butir*". Paper presented at Prosiding Seminar Nasional Matematika dan Pendidikan Matematika, UMS, 2015.

- Kusaeri, K. (2014). *Acuan dan Teknik Penilaian Proses dan Hasil Belajar dalam Kurikulum 2013*. Yogyakarta: AR-RUZZ MEDIA.
- Kusaeri, K. (2012). *Pengembangan Tes Diagnostik Dengan Menggunakan Model Dina Untuk Mendapatkan Informasi Salah Konsepsi Dalam Aljabar* (Doctoral dissertation, Universitas Negeri Yogyakarta).
- Kusumaningrum, Maya., - Abdul Aziz Saefudin, “*Mengoptimalkan Kemampuan Berpikir Matematika Melalui Pemecahan Masalah Matematika*”. Paper presented at Seminar Nasional Matematika dan Pendidikan Matematika, Yogyakarta, 2012.
- Manfaat, Budi., – Zara Zahra Anasha, “*Analisis Kemampuan Berpikir Kritis Matematik Siswa Dengan Menggunakan Graded Response Models (GRM)*”. Paper Presented at Seminar Nasional Matematika dan Pendidikan Matematika, Yogyakarta, 2013.
- Mardani. *Hukum Kewarisan Islam Di Indonesia*. Jakarta: PT RajaGrafindo Persada, 2014.
- Mardapi, Djemari. *Teknik Penyusunan Instrumen Tes dan Non Tes*. Yogyakarta: Mitra Cendika, 2008.
- McMahon, G. “Critical thinking and ICT integration in a Western Australian secondary school”. *Educational Technology & Society*, Vol. 12, No. 4, 2009.
- Mufid, Fathul. “Posisi Al-Qur’an Dalam Struktur Dan Sumber Ilmu Islam”. *Jurnal ADDIN Media Dialektika Ilmu Islam*. Vol. 2 No. 2, Juli-Desember, 2010.
- Mufid, Fathul. “Integrasi Ilmu-Ilmu Islam”. *Equilibrium*. Vol. 1 No. 1, Juni, 2013.
- Nahadi, et.al. “Pengembangan Dan Analisis Soal Ulangan Kenaikan Kelas Kimia SMA Kelas X Berdasarkan Classical Test Theory Dan Item Response Theory”. *Jurnal Pengajaran MIPA*. Vol. 16 No. 2, Oktober, 2011.
- Nitko – Brookhart. *Educational assessment of student 6th*. Boston: Pearson Education, Inc, 2011.
- Nurani, Dewi., - Tri Wahyuni. *Matematika Konsep dan Aplikasinya 1*. Jakarta:Pusat Perbukuan Dep Pendidikan Nasional, 2008.
- Pratiwi, Indah Hesti, Disertasi Doktor: “*Kemampuan Guru Mata Pelajaran Ipa Dalam Pembuatan Soal HOT (Higher Order Thinking) Dan Kesesuaian Penulisan Soal Di Smp Negeri 1 Kragan Rembang*”. Surakarta: Universitas Muhammadiyah Surakarta, 2015.

- Qalyubi, Al., - Al-Umairah. *Hasyitani' ala Minhajati Thalibin*. Beirut: Dar al-Fikr, juz 3, tth.
- Retnawati, Heri. *Validitas, Reabilitas & Karakteristik* Butir. Yogyakarta: Parama Publishing, 2016.
- Rofiah, Emi., et.al. "Penyusunan Instrumen Tes Kemampuan Berpikir Tingkat Tinggi Fisika Pada Siswa SMP". *Jurnal Pendidikan Fisika*. Vol. 1 No. 2, September, 2013.
- Rosyada, Dede. *Hukum Islam dan Pranata Sosial*. Jakarta: RajaGrafindo Persada, 1993.
- Samritin, Disertasi Doktor: "Pengembangan Instrumen Penilaian Kemampuan Higher Order Thinking Siswa SMP dalam Mata Pelajaran Matematika". Yogyakarta: Universitas Negeri Yogyakarta, 2014.
- Sari, Rosalia Hera Novita, "*Literasi Matematika: Apa, Mengapa dan Bagaimana?*". Paper presented at Seminar Nasional Matematika Dan Pendidikan Matematika, Universitas Negeri Yogyakarta, 2015.
- Sofiyah, Siti., et.al. "Pengembangan Paket Tes Kemampuan Berpikir Tingkat Tinggi Matematika Berdasarkan Revisi Taksonomi Bloom Pada Siswa Kelas V SD". *Artikel Ilmiah Mahasiswa*. Vol. 1 No.1, 2015.
- Suhandoyo, Guntur., - Pradnyo Wijayanti. "Profil Kemampuan Berpikir Kreatif Siswa Dalam Menyelesaikan Soal Higher Order Thinking Ditinjau Dari Adversity Quotient (AQ)". *Jurnal Ilmiah Pendidikan Matematika*. Vol. 3 No. 5, 2016.
- Thompson, Tony. "An Analysis of Higher-Order Thinking on Algebra I End-of Course Tests". *International Journal for Mathematics Teaching and Learning*, 2011.
- Wang, S., & Wang, H. (2010). Supporting Higher-Order Thinking in E-Learning Environment.
- Widoyoko, S.E.P., "*Peranan Evaluasi Program Pembelajaran dalam Meningkatkan Kualitas Pendidikan*". Paper Presented at Seminar Nasional Program Pascasarjana UNY berkerja sama dengan HEPI, Yogyakarta, 2008.
- Wulan, A. R. (2007). Pengertian dan esensi konsep evaluasi, asesmen, tes, dan pengukuran., accessed 4 Juni 2017; http://file.upi.edu/direktori/fpmipa/jurpend_biologi/anaratnawulan/pengertianasesmen.pdf.