

# Karakteristik Instrumen Penilaian Hasil Belajar Matematika Ranah Kognitif yang Dikembangkan Mengacu pada Model Pisa

Nur Aida<sup>1</sup>, Kusacri<sup>2</sup>, A. Saepul Hamdani<sup>3</sup>

<sup>1,2,3</sup> *Jurusan Pendidikan Matematika dan IPA, Universitas Islam Negeri Sunan Ampel Surabaya*  
e-mail: [kusacri@uinsby.ac.id](mailto:kusacri@uinsby.ac.id)

**ABSTRAK.** Penelitian ini untuk mengetahui karakteristik instrumen penilaian hasil belajar yang dikembangkan berupa validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan nilai *guessing*. Selanjutnya, data penelitian dianalisis secara kuantitatif menggunakan uji validitas, uji reliabilitas, dan melalui *software* BILOG-MG untuk mengetahui karakteristik tiap instrumen penilaian hasil belajar yang dikembangkan. Hasil penelitian menunjukkan bahwa 7 butir soal yang dikembangkan dinyatakan valid dengan reliabilitas sedang dan 1 butir soal dinyatakan tidak valid. Sedangkan 7 butir soal yang dinyatakan valid meliputi daya pembeda, tingkat kesukaran, dan nilai *guessing* termasuk kategori baik, hanya butir soal nomor 1 yang mempunyai kategori tingkat kesukaran tidak baik.

**Kata kunci:** instrumen penilaian, kognitif, PISA

## PENDAHULUAN

PISA (*Programme for International Student Assessment*) merupakan studi internasional tentang prestasi literasi membaca, matematika, sains, dan keuangan siswa sekolah berusia 15 tahun (Rahmawati & Tatag, 2014). PISA diikuti oleh beberapa negara yang tergabung dalam OECD (*The Organisation for Economic Co-operation and Development*) termasuk Indonesia. PISA dilaksanakan setiap tiga tahun sekali dengan fokus literasi yang berbeda setiap pelaksanaannya, mulai dari tahun 2000, 2003, 2006, 2009, 2012, dan seterusnya. Indonesia sendiri turut berpartisipasi dalam PISA pada literasi matematika sebanyak lima kali selama tahun 2000-2012.

Namun, sejak pertama kali keikutsertaan Indonesia dalam PISA prestasi siswa-siswi belum menunjukkan hasil yang memuaskan pada semua bidang literasi termasuk matematika. Dalam kurun waktu enam tahun mulai dari tahun 2003 hingga tahun 2009 hampir 80% siswa Indonesia hanya mampu mencapai di bawah garis batas level 2 dari 6 level soal yang diujikan. Pada tahun 2003, menunjukkan prestasi Indonesia bidang matematika menempati urutan 36 dari 41 negara. Pada tahun 2006, skor perolehan siswa SMP pada matematika berkisar hanya pada angka 391 dengan skala yang sudah ditentukan yaitu 0-800, padahal rata-rata skor sebesar 500 (Suryo, 2016). Pada pelaksanaan PISA literasi matematika tahun 2009, hampir semua siswa Indonesia mencapai level 3, sedangkan hanya 0,1% siswa Indonesia mampu mencapai level 5 dan 6. Keterpurukan hasil ini semakin diperkuat oleh hasil survei PISA 2012 yang menempatkan siswa Indonesia pada peringkat 64 dari 65 negara peserta dengan pencapaian level yang masih terbilang rendah. Hampir seluruh siswa Indonesia sebesar 98,5% hanya mampu mencapai level 3 (Ahmad & Zulkardi, 2014).

Rendahnya hasil PISA menunjukkan bahwa prestasi siswa Indonesia masih belum memuaskan secara internasional. Sedangkan secara nasional, hasil belajar matematika pada jenjang persekolahan adalah rendah (Evi, 2011). Penelitian Mahdiansyah dan Rahmawati (2014) menunjukkan bahwa kemampuan matematika siswa masih rendah terutama pada ketercapaian

literasi matematika. Dengan demikian prestasi matematika siswa Indonesia baik secara nasional maupun internasional belum menggembirakan.

Hal ini disebabkan oleh ketidakbiasaan siswa mengerjakan soal-soal matematika yang menuntut kemampuan *higher order thinking* (Samritin, 2014). Thomas dan Thorne (2005) menyatakan bahwa “*Higher Order Thinking is thinking on higher level that memorizing facts or telling something back to someone exactly the way the it was told to you. When a person memorizes and gives back the information without having to think about it. That’s because it’s much like a robot; it does what it’s programmed to do, but it doesn’t think for it self*”. Dengan demikian, kemampuan berpikir tingkat tinggi merupakan keterampilan yang dapat dilatihkan (Rosnawati, 2009). Namun, siswa Indonesia pada umumnya kurang terlatih dalam menyelesaikan soal-soal dengan karakteristik *higher order thinking* yang membutuhkan kemampuan penalaran yang baik seperti pada soal-soal PISA, sehingga hal ini menjadi salah satu faktor penyebab rendahnya hasil PISA siswa Indonesia (Suryo, 2016). Tidak hanya rendahnya kemampuan penalaran siswa, Wardhani (2011) mengatakan bahwa penyebab dari lemahnya kemampuan siswa dalam memecahkan masalah PISA adalah siswa kurang terbiasa melakukan proses pemecahan masalah dengan benar. Singkatnya, hasil PISA yang rendah disebabkan oleh kurang terbiasanya siswa dalam mengerjakan soal-soal dengan karakteristik PISA.

Faktor yang mempengaruhi kurang terbiasanya siswa dalam mengerjakan soal-soal dengan karakteristik PISA adalah penyajian instrumen penilaian hasil belajar di sekolah kurang dikaitkan dengan konteks kehidupan yang dihadapi siswa (Wardani, 2011). Guru dalam pembelajaran di kelas tidak mengaitkan dengan skema yang telah dimiliki oleh siswa dan siswa kurang diberikan kesempatan untuk menemukan kembali dan mengkonstruksi sendiri ide-ide matematika (Evi, 2011). Menurut hasil survey IMSTEP-JICA (1999) di Bandung, proses pembelajaran matematika guru umumnya terlalu berkonsentrasi pada latihan menyelesaikan soal yang lebih bersifat prosedural dan mekanistik, menjelaskan konsep secara informatif, memberikan contoh soal dan memberikan soal-soal latihan (Tatang, 2007). Sebagian besar guru matematika jarang memberikan soal-soal matematika kepada siswanya dalam bentuk non rutin. Guru hanya terpaku pada soal-soal rutin yang hanya melatih siswa secara mekanistik dan sifatnya *textbook* (Edy, 2012). Kegiatan pembelajaran demikian tidak mengakomodasi pengembangan kemampuan siswa dalam pemecahan masalah, penalaran, koneksi, dan komunikasi matematis. Akibatnya, kemampuan kognitif tingkat tinggi siswa sangat lemah karena kegiatan pembelajaran yang biasa dilakukan hanya mendorong siswa untuk berpikir pada tataran tingkat rendah (Tatang, 2007).

Dalam ranah kompetensi kognitif atau pengetahuan terdapat enam tingkat proses berpikir kognitif menurut Taksonomi Bloom yang telah direvisi Anderson dan Krathwohl (Kusaeri, 2014), yaitu: mengingat (*remember*), memahami/ mengerti (*understand*), menerapkan (*apply*), menganalisis (*analyze*), mengevaluasi (*evaluate*), dan menciptakan (*create*). Proses berpikir kognitif tersebut terdapat hubungan antara tingkat kemampuan kognitif dengan level kemampuan matematika dalam PISA (Harianto, 2014).

Dengan demikian, diperlukan adanya pengembangan kemampuan kognitif yang menuntut siswa berpikir tingkat tinggi melalui pemberian stimulus berupa soal untuk melakukan proses berpikir matematika (Samritin, 2014). Stimulus yang dapat meningkatkan kemampuan berpikir tingkat tinggi yaitu dengan pemberian instrumen penilaian hasil belajar matematika ranah kognitif yang mengacu pada model PISA, karena model PISA memiliki kemampuan dasar yang setara dengan berpikir tingkat tinggi. PISA juga menggunakan pendekatan literasi yang inovatif, suatu konsep belajar yang berkaitan dengan kapasitas para siswa untuk menerapkan pengetahuan dan keterampilan dalam mata pelajaran kunci disertai dengan kemampuan untuk menelaah, memberi alasan dan mengkomunikasikannya secara efektif, serta memecahkan dan menginterpretasikan permasalahan dalam berbagai situasi (Robiyatul, 2014; Evy, Zulkardi & Darmawijoyo, 2012).

Kemampuan dasar matematika dalam PISA meliputi berpikir dan penalaran matematika, pembuktian matematika, pemodelan, menyikapi masalah dan pemecahannya, representasi, penggunaan simbol, komunikasi, dan penggunaan alat-alat matematika (OECD, 2009). Selain kemampuan dasar matematika, kemampuan yang diukur dalam PISA meliputi merumuskan

(*formulate*), menerapkan (*employ*), dan menafsirkan (*interpret*) (OECD, 2012). Mengingat rendahnya prestasi matematika secara nasional maupun internasional, maka kemampuan tersebut menjadi penting untuk disisipkan dalam instrumen penilaian hasil belajar matematika ranah kognitif yang dikembangkan dengan harapan adanya peningkatan kemampuan berpikir tingkat tinggi siswa ketika dihadapkan pada studi PISA selanjutnya.

Matematika dalam PISA 2002 memuat empat konten yang tertuang dalam soal yang diujikan. Empat konten tersebut yaitu *shape and space* (bentuk dan ruang), *change and relationship* (perubahan dan hubungan), *quantity* (bilangan), dan *uncertainty* (ketidakpastian atau *probability*) (OECD, 2012). Dalam penelitian ini, peneliti mengambil konten *quantity* (bilangan) sebagai objek soal yang akan dikembangkan. Ketiga konten lainnya tidak bisa lepas dari konten *quantity* yang selalu ada pada setiap permasalahan yang diberikan, karena *quantity* merupakan dasar dari ketiga konten yang lain. Dengan demikian konten *quantity* harus lebih dikuasai oleh siswa ketika dihadapkan pada soal matematika model PISA. *Quantity* (bilangan) berkaitan dengan hubungan bilangan dan pola bilangan, antara lain kemampuan untuk memahami ukuran, pola bilangan, dan segala sesuatu yang berhubungan dengan bilangan dalam kehidupan sehari-hari, seperti menghitung dan mengukur suatu benda tertentu. Kemampuan bernalar secara kuantitatif, menyajikan sesuatu dengan angka, memahami langkah-langkah matematika, berhitung di luar kepala, dan melakukan penaksiran juga merupakan bagian dari konten *quantity*.

Berdasarkan pemaparan tersebut di atas, maka besar harapan instrumen penilaian hasil belajar matematika ranah kognitif yang dikembangkan mengacu pada model PISA mampu meningkatkan kemampuan berpikir matematika siswa sehingga dihasilkan instrumen penilaian dengan karakteristik yang baik. Oleh karena itu, penelitian ini bertujuan untuk mengetahui karakteristik instrumen penilaian hasil belajar matematika ranah kognitif yang dikembangkan mengacu pada model PISA yang meliputi validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan nilai *guessing*.

## METODE

Penelitian ini menggunakan metode pengembangan dengan pendekatan kualitatif. Penelitian ini menekankan pada pengembangan instrumen penilaian hasil belajar matematika ranah kognitif yang dikembangkan mengacu pada model PISA untuk mengetahui karakteristik instrumen penilaian yang dikembangkan. Penelitian dilakukan dengan mengumpulkan informasi hasil tes melalui instrumen penilaian hasil belajar matematika ranah kognitif yang dikembangkan. Data yang diperoleh dari hasil tes selanjutnya dianalisis dengan uji validitas, uji reliabilitas, dan dengan bantuan *software* BILOG-MG untuk mengetahui karakteristik soal.

Prosedur pengembangan yang digunakan dalam penelitian ini mengacu pada prosedur pengembangan instrumen yang dikemukakan oleh Mardapi yang terdiri atas Sembilan langkah, yakni: (1) menyusun spesifikasi tes, (2) menulis soal tes, (3) menelaah soal tes, (4) melakukan ujicoba tes, (5) menganalisis butir soal, (6) memperbaiki tes, (7) merakit tes, (8) melaksanakan tes, (9) menafsirkan hasil tes (Samritin, 2014). Pada tahap menafsirkan hasil tes digunakan untuk mengetahui karakteristik instrumen penilaian hasil belajar matematika ranah kognitif.

Subjek coba penelitian ini adalah siswa kelas IX–E SMPN 1 Surabaya, terdiri dari 36 siswa. Jenis data yang dimaksud dalam penelitian ini hasil tes instrumen penilaian hasil belajar matematika ranah kognitif yang dikembangkan untuk mengetahui validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan nilai *guessing*.

Instrumen pengumpul data berupa tes hasil belajar matematika ranah kognitif dengan 8 butir soal PISA yang dikembangkan. Data penelitian diperoleh dari hasil tes siswa yang kemudian dianalisis menggunakan uji validitas, uji reliabilitas, dan *software* BILOG-MG. Uji validitas menggunakan rumus korelasi *product moment* lalu dilanjutkan dengan uji-t, dan uji reliabilitas menggunakan rumus *cronbach alpha*. Sedangkan analisis daya pembeda, tingkat kesukaran, dan nilai *guessing* berdasarkan pendekatan IRT (*item response theory*) tiga parameter logistik (3PL) dilakukan dengan menggunakan bantuan program BILOG-MG. Kriteria untuk melihat butir soal yang baik dalam teori respon butir digunakan pendapat yang dikemukakan oleh Hadi (2011; Huda & Mardapi, 2015) yaitu sebagai berikut:

**Tabel 1. Kriteria Butir Soal yang Baik Berdasarkan Teori Respon Butir**

Parameter	Nilai	Keterangan
Daya Pembeda	0,4 s/d 2	Baik
Tingkat Kesukaran	-2 s/d 2	Baik
<i>Pseudo Guessing</i>	0 s/d $\frac{1}{k}$ (k = jumlah alternatif jawaban)	Baik

Dalam IRT 3PL, daya pembeda disimbolkan dengan parameter “*a*” atau biasa disebut *discrimination parameters*. Pendapat lain menurut Hambleton, Swaminathan & Rogers dalam disertasi Song Gao (2011) bahwa rentang normal untuk parameter “*a*” atau daya pembeda adalah 0 – 2. Dalam prosiding seminar nasional matematika dan pendidikan matematika UMS oleh Devi Dwi Kurniawan (2015) juga menyebutkan bahwa daya pembeda yang baik berkisar antara  $0 \leq a \leq 2$ . Dengan demikian, dalam penelitian ini digunakan rentang untuk daya pembeda yang baik berkisar antara  $0 \leq a \leq 2$ . Tingkat kesukaran disimbolkan dengan parameter “*b*” atau biasa disebut *difficulty parameters*. Tingkat kesukaran yang baik yaitu memiliki rentang  $-2 \leq b \leq 2$  (Devi Dwi Kurniawan, 2015). Nilai *Guessing* disimbolkan dengan parameter “*c*” atau biasa disebut *pseudo-guessing*. *Guessing* mempunyai rentang  $0 \leq c \leq 1$  (Frank B. Baker, 2001). Berdasarkan Tabel 1 di atas, bahwa untuk *guessing* berkisar antara 0 s/d  $\frac{1}{k}$ , k merupakan jumlah alternatif jawaban. Dalam soal yang dikembangkan terdapat 4 alternatif jawaban, sehingga *guessing* yang baik berkisar antara 0 s/d 0,25.

### HASIL DAN PEMBAHASAN

Analisis empiris kevalidan menggunakan rumus korelasi *product moment* lalu dilanjutkan dengan uji-t. Analisis dilakukan terhadap 8 butir soal yang telah dikembangkan mengacu pada model PISA. Perhitungan validitas ditujukan untuk menentukan instrumen penilaian hasil belajar matematika ranah kognitif yang dikembangkan mengacu pada model PISA valid atau tidak valid terhadap hasil tes. Hasil perhitungan  $t_{hitung}$  dikorelasikan pada  $t_{tabel}$  untuk taraf signifikan  $\alpha = 5\%$  dengan derajat kebebasan ( $dk = n-2$ ). Adapun ketentuan pengujiannya adalah sebagai berikut: Jika  $t_{hitung} > t_{tabel}$  maka butir soal tersebut valid dan jika  $t_{hitung} \leq t_{tabel}$  maka butir soal tersebut tidak valid (Sugiyono, 2011). Berikut adalah rekapitulasi hasil perhitungan validitas instrumen penilaian hasil belajar matematika ranah kognitif yang dikembangkan mengacu pada model PISA terhadap hasil tes.

**Tabel 2. Rekapitulasi Hasil Perhitungan Validitas Instrumen Penilaian terhadap Hasil Tes**

Butir Soal Nomor	$t_{hitung}$	$t_{tabel}$
1	2,502	1,691
2	3,727	
3	2,455	
4	4,096	
5	4,077	
6	-0,060	
7	4,225	
8	6,059	

Berdasarkan rekapitulasi hasil perhitungan validitas menggunakan rumus korelasi *product moment* dan uji-t pada Tabel 1 di atas, diketahui bahwa nilai  $t_{hitung}$  butir soal nomor 1, 2, 3, 4, 5, 7, dan 8 berturut-turut adalah 2,502, 3,727, 2,455, 4,096, 4,077, 4,225, dan 6,059 yang lebih besar dari nilai  $t_{tabel} = 1,691$ . Dengan demikian, ketujuh butir soal tersebut dinyatakan valid. Sedangkan butir soal nomor 6 dengan nilai  $t_{hitung}$  adalah  $-0,060$  yang lebih kecil dari nilai  $t_{tabel} = 1,691$ , maka butir soal tersebut dinyatakan tidak valid. Dari butir soal yang tidak valid dilakukan analisis terhadap jawaban siswa pada butir soal nomor 6. Berdasarkan hasil jawaban siswa, diperoleh 6 siswa menjawab salah dan 30 siswa menjawab benar. Dari 6 siswa yang menjawab salah, 4 siswa memilih pilihan jawaban D sedangkan 2 siswa lainnya memilih pilihan jawaban A dan C. Penyebab kesalahan siswa dalam menjawab soal salah satunya adalah kurang memahami kalimat dalam soal sehingga terjadi perbedaan asumsi dengan apa yang diminta pada soal. Dengan demikian, soal yang mudah akan salah jika kurangnya pemahaman terhadap soal. Disisi lain, siswa juga kurang terlatih untuk menyelesaikan soal non rutin ketika pembelajaran di sekolah.

Setelah dilakukan perhitungan dan analisis kevalidan terhadap 8 butir soal tersebut, selanjutnya dilakukan perhitungan reliabilitas terkait butir soal yang valid. Rumus yang digunakan adalah *Cronbach Alpha* ( $r_{11}$ ) dengan taraf signifikan  $\alpha = 5\%$ . Adapun kriteria dari reliabilitas butir soal adalah sebagai berikut: jika  $r_{11} > r_{tabel}$  maka butir soal reliabel dan jika  $r_{11} < r_{tabel}$  maka butir soal tidak reliabel (Sandjojo, 2011). Hasil perhitungan didapat bahwa nilai reliabilitas butir soal valid yang dikembangkan adalah 0,635. Dengan menyesuaikan kategori yang terdapat pada pembahasan sebelumnya, maka butir soal tersebut dinyatakan reliabel yang berada pada kategori moderat atau sedang.

Analisis butir soal untuk mengetahui daya pembeda, tingkat kesukaran, dan nilai *guessing* dilakukan menggunakan bantuan *software* BILOG-MG. Analisis ini menggunakan model tiga parameter logistik (3PL). Hasil analisis menggunakan *software* BILOG-MG 3PL diperoleh data sebagai berikut:

ITEM	NAME	#TRIED	#RIGHT	PCT	ITEM*TEST CORRELATION		
					LOGIT/1.7	PEARSON	BISERIAL
1	ITEM1	36.0	34.0	94.4	-1.67	0.218	0.445
2	ITEM2	36.0	28.0	77.8	-0.74	0.226	0.316
3	ITEM3	36.0	31.0	86.1	-1.07	0.111	0.174
4	ITEM4	36.0	32.0	88.9	-1.22	0.360	0.597
5	ITEM5	36.0	33.0	91.7	-1.41	0.388	0.700
6	ITEM6	36.0	30.0	83.3	-0.95	-0.303	-0.451
7	ITEM7	36.0	31.0	86.1	-1.07	0.349	0.545
8	ITEM8	36.0	32.0	88.9	-1.22	0.554	0.919

**Gambar 1 Perhitungan Statistik Instrumen Penilaian Menggunakan BILOG-MG**

Gambar di atas merupakan hasil perhitungan statistik oleh BILOG-MG terhadap kedelapan butir soal yang dikembangkan. Berdasarkan hasil tersebut, kolom *#RIGHT* merupakan jumlah peserta tes yang menjawab soal dengan benar. Dengan demikian, diperoleh hasil perhitungan *correlation biserial* (korelasi biserial) yang berbeda-beda tiap butir soal. Butir soal nomor 6 mempunyai korelasi biserial sebesar  $-0,451$ , yang artinya butir soal nomor 6 tidak akan dikalibrasi karena nilai korelasi biserialnya kurang dari  $-0,15$ . Sehingga butir soal nomor 6 dikeluarkan atau tidak digunakan dalam analisis karakteristik instrumen penilaian hasil belajar matematika ranah kognitif mengacu pada model PISA. Berdasarkan pembahasan sebelumnya, butir soal nomor 6 dinyatakan tidak valid pada hasil test menurut perhitungan korelasi *product moment* dan uji-t. Pada kenyataannya sesuai hasil analisis *software* BILOG-MG, butir soal nomor 6 mempunyai korelasi biserial kurang dari  $-0,15$  yang mengakibatkan tidak dapat dikalibrasi pada analisis berikutnya. Dengan demikian, karakteristik butir soal yang dianalisis untuk mengetahui daya pembeda, tingkat kesukaran, dan nilai *guessing* hanya 7 butir yang meliputi butir soal nomor 1, 2, 3, 4, 5, 7, dan 8.

Berdasarkan data pada Gambar 1 di atas, diperoleh estimasi perhitungan daya pembeda, tingkat kesukaran, dan nilai *guessing* terhadap butir soal matematika model PISA pada konten *quantity* sebagai instrumen penilaian hasil belajar ranah kognitif yang dikembangkan sebagai berikut:

ITEM	INTERCEPT	SLOPE	THRESHOLD	LOADING	ASYMPTOTE
ITEM1	2.004 0.619*	0.901 0.385*	-2.224 0.982*	0.670 0.286*	0.250 0.097*
ITEM2	0.826 0.547*	1.216 0.653*	-0.679 0.497*	0.772 0.415*	0.251 0.096*
ITEM3	1.109 0.408*	0.770 0.309*	-1.441 0.711*	0.610 0.245*	0.255 0.098*
ITEM4	1.481 0.502*	0.937 0.411*	-1.580 0.739*	0.684 0.300*	0.244 0.095*
ITEM5	1.864 0.845*	1.141 0.580*	-1.633 0.644*	0.752 0.382*	0.253 0.097*
ITEM7	1.274 0.470*	0.952 0.421*	-1.338 0.615*	0.689 0.305*	0.244 0.095*
ITEM8	1.802 0.915*	1.293 0.723*	-1.394 0.550*	0.791 0.442*	0.238 0.094*

\* STANDARD ERROR

Keterangan:

Kolom **SLOPE** berisi parameter *a* atau *Daya Pembeda*

Kolom **THRESHOLD** berisi parameter *b* atau *Tingkat Kesukaran*

Kolom **ASYMPTOTE** berisi parameter *c* atau *Nilai Guessing*

**Gambar 2** Estimasi Perhitungan Karakteristik Butir Soal Menggunakan BILOG-MG

Berdasarkan data setimasi perhitungan karakteristik butir soal di atas diperoleh rekapitulasi hasil estimasi perhitungan karakteristik yang meliputi daya pembeda, tingkat kesukaran, dan nilai *guessing* beserta keterangannya sebagai berikut:

**Tabel 3.** Rekapitulasi Hasil Estimasi Perhitungan Karakteristik Instrumen Penilaian

Butir Soal Nomor	Daya Pembeda		Tingkat Kesukaran		Nilai <i>Guessing</i>	
	<i>a</i>	Keterangan	<i>b</i>	Keterangan	<i>c</i>	Keterangan
1	0,901	Baik	-2,224	Tidak Baik	0,250	Baik
2	1,216	Baik	-0,679	Baik	0,251	Baik
3	0,770	Baik	-1,441	Baik	0,255	Baik
4	0,937	Baik	-1,580	Baik	0,244	Baik
5	1,141	Baik	-1,633	Baik	0,253	Baik
6	Butir soal gugur saat analisis					
7	0,952	Baik	-1,338	Baik	0,244	Baik
8	1,293	Baik	-1,394	Baik	0,238	Baik

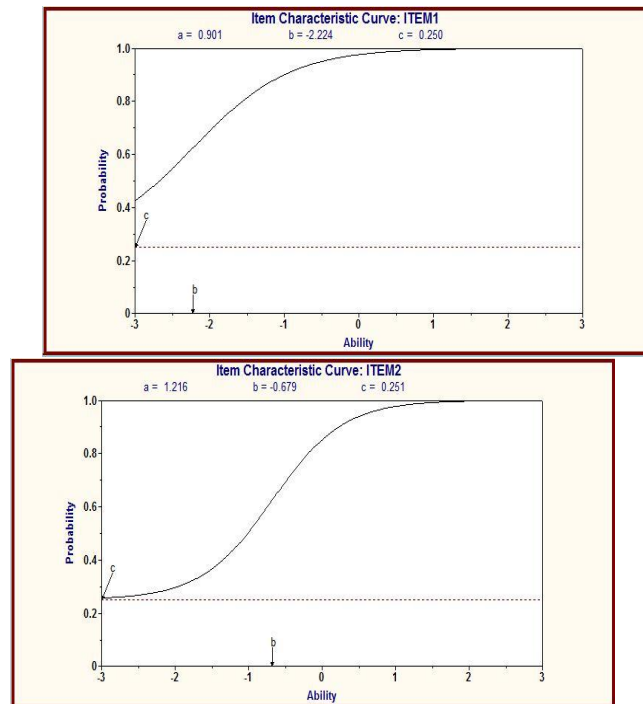
Berdasarkan tabel di atas, jumlah butir soal yang layak dianalisis menggunakan teori respon butir 3PL adalah 7 butir soal. Estimasi parameter *a* dari ke-7 butir soal tersebut dinyatakan

mempunyai nilai daya pembeda yang baik karena nilai daya pembeda yang dihasilkan berada pada rentang 0 – 2. Daya pembeda merupakan kemampuan suatu instrumen penilaian dalam membedakan antara peserta didik yang sudah menguasai materi dengan peserta didik yang belum menguasai materi. Semakin tinggi daya beda suatu *item*, semakin baik suatu *item* dalam membedakan kemampuan penempuh tes (Purwo, 2010). Sesuai hasil estimasi perhitungan tersebut, ke-7 butir soal dinyatakan baik dalam membedakan peserta didik yang sudah menguasai materi dengan peserta didik yang belum menguasai materi.

Estimasi parameter  $b$  atau tingkat kesukaran dari ke-7 butir soal tersebut, terdapat 6 butir soal termasuk kategori tingkat kesukaran yang baik karena berada pada rentang  $-2$  hingga  $2$  dan 1 butir soal memiliki tingkat kesukaran yang tidak baik. Tingkat kesukaran yang tidak baik terlihat dari hasil estimasi yang tidak berada pada rentang  $-2 \leq b \leq 2$ . Butir soal tersebut memiliki nilai tingkat kesukaran  $< -2$ , sehingga termasuk dalam kategori butir soal yang mudah. Hal tersebut sejalan dengan pernyataan Devi Dwi Kurniawan (2015) yang menyatakan bahwa nilai tingkat kesukaran  $> 2$  termasuk dalam kategori butir soal yang sulit dan nilai tingkat kesukaran  $< -2$  termasuk dalam kategori butir soal yang mudah, sedangkan nilai tingkat kesukaran yang berada pada rentang  $-2 \leq b \leq 2$  termasuk dalam kategori butir soal yang sedang. Dengan demikian, butir soal nomor 1 termasuk dalam kategori soal yang mudah, sedangkan butir soal nomor 2, 3, 4, 5, 7, dan 8 termasuk dalam kategori soal yang sedang.

Selanjutnya untuk nilai *guessing* dalam Tabel 1 berkisar antara  $0 \leq \frac{1}{k}$ ,  $k$  merupakan jumlah alternatif jawaban. Jumlah alternatif jawaban dalam penelitian ini adalah empat, sehingga *guessing* yang baik berkisar antara  $0 \leq \frac{1}{k} \leq 0,25$ . Berdasarkan perhitungan teori respon butir 3PL menggunakan BILOG-MG, ke-7 butir soal tersebut menunjukkan bahwa nilai *guessing* hanya sampai pada 0,255 dan tidak lebih dari 0,5. Hal tersebut sejalan dengan pernyataan Song Gao (2011) dalam disertasinya yang berbunyi “*The highest probability of guessing cannot be greater than 0,5; otherwise, it would not be called knowledge-based answer instead of guessing*”. Kemungkinan terbesar dari *guessing* (menebak) tidak lebih dari 0,5; dengan kata lain, tidak dapat dikatakan jawaban dari sebuah pengetahuan hanya dari *guessing* (menebak). Artinya, nilai *guessing* dalam sebuah jawaban tidak akan lebih dari 0,5 sehingga hal tersebut tidak dapat dijadikan patokan atau pedoman kebenaran jawaban dari sebuah pengetahuan tersebut hanya dengan menebak. Dengan demikian, kemampuan menebak tidak dapat benar-benar mengetahui kemampuan seseorang. Menebak dapat terjadi terutama pada soal bentuk pilihan ganda dan hal tersebut dapat menyebabkan kesalahan skor variansi terutama kesalahan pada validitas dan reliabilitas soal (Roger, 1999). Dalam menyusun soal bentuk pilihan ganda, kemampuan dalam menyusun alternatif jawaban merupakan aspek yang sangat penting. Penggunaan *distractor* yang tidak baik akan mengurangi kualitas soal. Tidak berfungsinya *distractor* dengan baik secara empirik ditunjukkan oleh hasil penelitian yang dilakukan Attali & Bar-Hillel (2003) yang menyimpulkan bahwa baik penempuh tes maupun pembuat soal memiliki kecenderungan yang sama dalam memilih jawaban atau menempatkan kunci jawaban yaitu pada jawaban yang berada di tengah. Hal tersebut, tentunya meningkatkan kesempatan penempuh tes dalam menebak jawaban. Semakin tinggi kemampuan menebak penempuh tes, tentunya akan memperlemah daya pembeda soal. Semakin rendah daya pembeda soal akan menyebabkan semakin homogen skor yang diperoleh, dan semakin homogen skor yang diperoleh akan memperlemah reliabilitas soal tersebut (Allen & Yen, 1979). Sehingga adanya soal tidak valid dan mempunyai reliabilitas rendah juga dapat disebabkan kemampuan menebak peserta didik yang tinggi dalam menjawab tes yang diberikan.

Berikut adalah contoh grafik ICC butir soal nomor 1 dan 2 pada hasil analisis menggunakan BILOG-MG:



Gambar 3 *Item Characteristic Curve (ICC)* Butir Soal Nomor 1 dan 2

Model 3PL melibatkan tiga parameter yaitu daya pembeda ( $a$ ), tingkat kesukaran ( $b$ ), dan nilai *guessing* ( $c$ ). Berdasarkan gambar di atas, diketahui nilai  $a$  butir soal nomor 1 yaitu 0,901 dengan nilai  $b$  sebesar  $-2,224$  dan nilai  $c$  sebesar 0,250. Sedangkan nilai  $a$  butir soal nomor 2 yaitu 1,216 dengan nilai  $b$  sebesar  $-0,679$  dan nilai  $c$  sebesar 0,251. Berdasarkan hal tersebut diketahui bahwa nilai  $a$  untuk butir soal nomor 1 lebih rendah dari butir soal nomor 2, sehingga ICC pada butir soal nomor 1 tampak lebih landai dibandingkan butir soal nomor 2. Dengan demikian semakin rendah nilai daya pembeda ( $a$ ) suatu butir soal maka ICC akan tampak lebih landai, sebaliknya semakin tinggi nilai daya pembeda ( $b$ ) suatu butir soal maka ICC akan tampak lebih menanjak. Nilai  $b_1 = -2,224$  dapat diartikan bahwa diperlukan kemampuan minimal  $-2,224$  untuk menjawab butir soal dengan benar. Sama halnya dengan butir soal nomor 1, butir soal nomor 2 dengan nilai  $b_2 = -0,679$  dapat diartikan bahwa diperlukan kemampuan minimal  $-0,679$  untuk menjawab butir soal dengan benar.

Nilai *guessing* ( $c$ ) pada ICC di atas, merupakan peluang menjawab benar dengan menebak tiap *item* tes yang diberikan. *Guessing* pada ICC menunjukkan jarak asimtot bawah terhadap skala kemampuan dan menyatakan probabilitas peserta ujian yang berkemampuan rendah menjawab *item* dengan benar (Hambleton, 1991). Nilai *guessing* mempunyai hubungan dengan daya pembeda *item*. Keduanya sangat menentukan besarnya fungsi informasi *item* yaitu nilai fungsi yang menyatakan ukuran ketepatan suatu pengukuran (Hambleton & Swaminathan, 1985). Fungsi informasi *item* dapat dimaknai sebagai reliabilitas dalam teori tes klasik. Nilai fungsi informasi *item* akan maksimal apabila daya pembeda bernilai setinggi mungkin dan nilai *guessing* sekecil mungkin (Hambleton, 1991). Berdasarkan Tabel 3 di atas, daya pembeda dengan nilai setinggi mungkin dan *guessing* dengan nilai sekecil mungkin terdapat pada butir soal nomor 8. Butir soal nomor 8 mempunyai daya pembeda sebesar 1,293 dan nilai *guessing* sebesar 0,238 yang artinya fungsi informasi atau reliabilitas *item* bernilai maksimal jika dibandingkan dengan butir soal yang lain.

## KESIMPULAN

Berdasarkan uraian sebelumnya, kesimpulan yang dapat disusun adalah instrumen penilaian hasil belajar matematika ranah kognitif yang dikembangkan mengacu pada model PISA menghasilkan tujuh butir soal dinyatakan valid dan satu butir soal dinyatakan tidak valid. Sedangkan reliabilitas butir soal dinyatakan reliabel dengan mendapat kategori moderat atau sedang. Daya pembeda ke-7 butir soal termasuk kategori baik karena berada pada rentang 0 sampai 2. Tingkat



kesukaran butir soal yang termasuk kategori baik dimiliki oleh 6 butir soal dan kategori tidak baik hanya 1 butir soal. Butir soal dengan kategori tidak baik dikarenakan nilai  $b$  atau tingkat kesukaran kurang dari  $-2$  dan termasuk soal yang mudah. Sedangkan butir soal dengan kategori baik termasuk soal yang sedang. Nilai *guessing* ke-7 butir soal termasuk kategori baik karena nilai  $c$  berada pada rentang  $0 \leq c \leq 1$ .

Berdasarkan temuan penelitian ini, maka dapat dibuat rekomendasi berikut. *Pertama*, ketidakbiasaan siswa dalam mengerjakan soal-soal non rutin di sekolah membuat mereka kesulitan ketika dihadapkan dengan soal *higher order thinking*. Oleh karena itu, guru matematika perlu membuat dan memberikan latihan kepada siswa berupa instrumen penilaian hasil belajar kognitif berbasis soal *higher order thinking*. Soal *higher order thinking* yang baik akan menumbuhkan kemampuan berpikir siswa menjadi lebih baik. Hal ini akan terjadi bila guru matematika rutin melatih siswanya mengerjakan soal-soal *higher order thinking*. *Kedua*, penelitian-penelitian yang menggali faktor-faktor yang mempengaruhi instrumen penilaian hasil belajar kognitif dan penerapannya dalam kemampuan tertentu belum banyak dilakukan di Indonesia. Oleh karena itu, kepada peneliti selanjutnya agar menambahkan penerapan dan pengaruh instrumen penilaian hasil belajar kognitif terhadap kemampuan tertentu, sehingga kebermanfaatannya akan lebih baik dan mampu menunjang keberhasilan mempelajari matematika khususnya soal *higher order thinking*.

#### DAFTAR PUSTAKA

- Adawiyah, Robiyatul. (2014). Skripsi Sarjana. "Pengembangan Soal Matematika Mengacu pada Standart PISA". Surabaya: UIN Sunan Ampel Surabaya.
- Aini, Rahmawati Nur dan Tatag. (2014). "Analisis Pemahaman Siswa SMP dalam Menyelesaikan Masalah Aljabar pada PISA". *Jurnal Ilmiah Pendidikan Matematika*. Vol. 03 No 2.
- Allen, M. J. & Yen, W. M. (2003). *Introduction to measurement theory*. Belmont, CA : Woodsworth, Inc.
- Attali, Y & Bar-Hillel, M. (2003). "Guess where: The position of correct answer in multiple choice test item as psychometric variable". *Journal of Educational Measurement*, 40 (2).
- Gao, Song. (2011). Doctoral Dissertation: "*The Exploration of The Relationship Between Guessing and Latent Ability in IRT Models*". Carbondale: Southern Illinois University.
- Hambleton, R. K & Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer.
- Hambleton, R. K. (1991). *Fundamentals of Item Response Theory*. New Delhi: Sage Publications.
- Hariato, dkk. (2014). "*Soal Matematika dalam PISA Kaitannya dengan Literasi Matematika dan Keterampilan Berpikir Tingkat Tinggi*". Paper presented at Seminar Nasional Matematika, Jember.
- Herman, Tatang. (2007). "Pembelajaran Berbasis Masalah untuk Meningkatkan Kemampuan Berpikir Matematis Tingkat Tinggi Siswa Sekolah Menengah Pertama". *Educationist*. Vol. 1 No. 1.
- Kohar, Ahmad Wachidul., dan Zulkardi. (2014). "Pengembangan Soal Berbasis Literasi Matematika dengan Menggunakan Kerangka PISA Tahun 2012". Paper presented at Konferensi Nasional Matematika XVII, Surabaya.
- Kusaeri, K. (2014). Acuan dan Teknik Penilaian Proses dan Hasil Belajar Dalam Kurikulum 2013. Yogyakarta: Ar-Ruz Media
- Mahdiansyah dan Rahmawati. (2014). "Literasi Matematika Siswa Pendidikan Menengah: Analisis Menggunakan Desain Tes Internasional dengan Konteks Indonesia". *Jurnal Pendidikan dan Kebudayaan*. Vol. 20 No. 4.
- OECD. (2009). *Learning Mathematics for Life : A View Perspective from PISA*. Paris: OECD Publications.
- OECD. (2014). *PISA 2012 Result: What Students know and can do: Student Performance in mathematics, reading, and science. Volume I*. Paris: OECD publishing.
- Purnomo, Suryo. (2016). Tesis. "Pengembangan Soal Matematika Model PISA Konten *Space and Shape* untuk Mengetahui Level Kemampuan Berpikir Tingkat Tinggi Berdasarkan Analisis Model Rasch". Jember: Universitas Jember.

- Rogers, H. J. (1999). Guessing in multiple choice tests. In G. N. Master & J. P. Keeves (Eds.), *Advances in Measurement in Educational Research and Assessment*. Amsterdam: Pergamom.
- Rosnawati, R. (2009). "Enam Tahapan Aktivitas dalam Pembelajaran Matematika untuk Mendayagunakan Berpikir Tingkat Tinggi Siswa". *Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA Fakultas MIPA Universitas Negeri Yogyakarta*.
- Sandjojo, Nidjo. (2011). *Metode Analisis Jalur dan Aplikasinya*. Jakarta: Pustaka Sinar Harapan.
- Sovianwati, Evi. (2011). "Pendekatan Matematika Realistik (PMR) untuk Meningkatkan Kemampuan Berpikir Siswa di Tingkat Sekolah Dasar". *Jurnal Edisi Khusus* No 2.
- Sugiyono. (2011). *Statistik untuk Penelitian*. Bandung: Alfabeta.
- Susongko, Purwo. (2010). "Perbedaan fungsi informasi item pada tes prestasi belajar matematika bentuk pilihan ganda yang menggunakan penskoran konvensional dan koreksi".
- Tandilling, Edy. (2012). "Pengembangan Instrumen untuk Mengukur Kemampuan Komunikasi Matematika, Pemahaman Matematika, dan *Self-Regulated Learning* Siswa dalam Pembelajaran Matematika di Sekolah Menengah Atas". *Jurnal Penelitian Pendidikan*. Vol. 13 No. 1.
- Wardhani, Sri., dan Rumiati. *Instrumen Penilaian Hasil Belajar Matematika SMP: Belajar dari PISA dan TIMSS* (Kementerian Pendidikan Nasional, Badan Pengembangan Sumber Daya Manusia Pendidikan dan Penjaminan Mutu Pendidikan, Pusat Pengembangan Pemberdayaan Pendidik dan Tenaga Kependidikan (PPPPTK) Matematika).