

Pengembangan Instrumen Pengukur *Higher Order Thinking Skills* Matematika Siswa SMA Kelas X

Zaenal Arifin¹*, Heri Retnawati¹

¹ Program Studi Pendidikan Matematika, Program Pascasarjana, Universitas Negeri Yogyakarta.
Jalan Colombo No. 1, Karangmalang, Yogyakarta 55281, Indonesia.

* Corresponding Author. Email: arifin1169@gmail.com, Telp: +62274-550836

Received: 20 May 2017; Revised: 7 June 2017; Accepted: 10 June 2017

Abstrak

Penelitian ini bertujuan untuk menghasilkan instrumen pengukur *higher order thinking skills* (HOTS) matematika siswa kelas X yang valid dan reliabel, dan untuk mengetahui *higher order thinking skills* (HOTS) matematika siswa kelas X. Penelitian ini merupakan penelitian pengembangan dengan langkah-langkah meliputi: menyusun spesifikasi tes, menulis soal tes, menelaah soal tes, melakukan uji coba tes, menganalisis butir soal, memperbaiki tes, dan merakit tes. Uji coba instrumen dilakukan pada tiga sekolah, yaitu SMAN 4 Yogyakarta, SMAN 6 Yogyakarta, dan SMAN 10 Yogyakarta. Jumlah keseluruhan subjek uji coba sebanyak 169 siswa. Data penelitian dikumpulkan dengan menggunakan tes uraian, tes pilihan ganda dan tes jawab singkat. Instrumen tes terdiri dari 75 butir soal, tetapi hanya 46 butir soal yang diujicobakan dan dibagi menjadi tiga paket soal. Paket soal A sebanyak 15 butir, paket soal B sebanyak 16 butir, dan paket soal C sebanyak 15 butir. Hasil penelitian menunjukkan bahwa instrumen yang dibuat valid. Hal ini berdasarkan perhitungan yang menggunakan rumus Aiken, bahwa nilai V pada semua butir soal $\geq 0,3$. Di samping itu, paket soal A dan paket B menghasilkan paket soal yang reliabel, dengan masing-masing nilai *Cronbach's Alpha* sebesar 0,738 dan 0,658. Sedangkan paket soal C tidak reliabel yang ditunjukkan dengan nilai *Cronbach's Alpha* sebesar 0,488. Hasil uji coba instrumen menunjukkan HOTS matematika siswa kelas X kurang baik. Hal ini dapat diketahui dari nilai rata-rata hasil uji coba sebesar 26,38 dalam skala 100.

Kata Kunci: pengembangan instrumen, *higher order thinking skills* (HOTS), validitas, reliabilitas.

Developing an Instrument to Measure Mathematics Higher Order Thinking Skills of 10th Grade Students in Senior High School

Abstract

This research aims to know an instrument to measure mathematics higher order thinking skills (HOTS) of 10th grade Students which is valid and reliable, and to know the mathematics higher order thinking skills (HOTS) of 10th grade students. This research was the development of instruments with steps: preparing test specifications, writing the items, reviewing the items, doing trial tests, analyzing the items, correcting the items, and arranging the items. The instruments trial conducted at three schools, that is SMAN 4 Yogyakarta, SMAN 6 Yogyakarta, and SMAN 10 Yogyakarta. The total number of subjects tested were as many as 169 students. The data were gathered using essay, multiple choice, and short-answer test. The instrument consists of 75 items, but only 46 items were tested and divided into three packets. Packet test A totaling 15 items, packet test B totaling 16 items, and packet test C totaling 15 items. The research finding reveals that the instrument is valid. This is based on calculations of Aiken's formula, that value of V for all items are less than 0.3. Besides it, packet test A and packet test B are reliable, with each value of Cronbach's Alpha are 0.738 and 0.658. But for packet test C is not reliable, because the value of Cronbach's Alpha is 0.488. The results of the test instrument can be concluded that mathematics higher order thinking (HOT) of 10th grade student is poor. It can be seen from the average value of the test results of 26.38 in the scale of 100.

Keywords: instrument development, *higher order thinking skills* (HOTS), validity, reliability

How to Cite: Arifin, Z., & Retnawati, H. (2017). Pengembangan instrumen pengukur higher order thinking skills matematika siswa SMA kelas X. *PYTHAGORAS: Jurnal Pendidikan Matematika*, 12(1), 98-108. doi:<http://dx.doi.org/10.21831/pg.v12i1.14058>

Permalink/DOI: <http://dx.doi.org/10.21831/pg.v12i1.14058>

PENDAHULUAN

Pendidikan mempunyai peranan penting untuk meningkatkan kualitas sumber daya manusia (SDM), karena pendidikan merupakan salah satu sarana untuk menghasilkan perubahan pada diri manusia. Siswa, sebagai manusia pembelajar di sekolah memiliki banyak sekali potensi pada diri mereka yang merupakan sumber daya manusia. Oleh karena itu, pendidikan merupakan hal yang sangat penting yang dibutuhkan siswa untuk merubah pikiran/*mindset* mereka menjadi berkualitas.

Hal tersebut sesuai dengan tujuan pengembangan dalam Kurikulum 2013 yang berdasarkan atas landasan filosofis. Permendikbud No. 69 tahun 2013 menyatakan bahwa Kurikulum 2013 dikembangkan berdasarkan landasan filosofis yang memberikan dasar bagi pengembangan seluruh potensi peserta didik menjadi manusia Indonesia berkualitas yang tercantum dalam tujuan pendidikan nasional. Adapun tujuan pendidikan nasional telah dicantumkan oleh pemerintah dalam Undang-Undang Sistem Pendidikan Nasional RI No. 20 Tahun 2003 yaitu, "Pendidikan Nasional bertujuan untuk berkembangnya potensi peserta didik agar menjadi manusia yang beriman, bertakwa kepada Tuhan Yang Maha Esa, berakhlak mulia, sehat, berilmu, cakap, kreatif, mandiri dan menjadi warga negara yang demokratis serta bertanggung jawab".

Peran pendidikan di sekolah berfokus untuk mengembangkan sumber daya siswa yang meliputi ranah kognitif, afektif dan psikomotorik, atau sikap spiritual, sikap sosial, pengetahuan dan keterampilan. Kemampuan berpikir tingkat tinggi (*higher order thinking skills*) merupakan salah satu sumber daya manusia, yang dalam hal ini adalah pengetahuan dan keterampilan, sehingga harus ditingkatkan dan dikembangkan. Hal tersebut didukung oleh pendapat Yen & Halili (2015, p.41) bahwa HOTS merupakan tujuan utama pembelajaran pada abad 21. Oleh karena itu, salah satu indikator keberhasilan peningkatan sumber daya manusia dalam bidang pendidikan adalah berkembangnya HOTS siswa.

Pada mata pelajaran matematika, HOTS merupakan salah satu yang prioritas untuk dikembangkan. Bahkan sejak tahun 2006 pemerintah Indonesia telah menetapkan dalam Peraturan Menteri Pendidikan nomor 22 tahun 2006 tentang Standar Isi tentang perlunya matematika diberikan untuk membekali peserta

didik dengan kemampuan berpikir logis, analitis, sistematis, kritis, dan kreatif, serta kemampuan bekerjasama.

Kemampuan-kemampuan yang disebutkan dalam standar isi tersebut memiliki relevansi dengan karakteristik HOTS. Karakteristik HOTS antara lain *nonalgoritmik*, bersifat kompleks, mempunyai banyak solusi, melibatkan variasi pengambilan keputusan dan interpretasi, penerapan banyak kriteria, dan bersifat *effortful* (Resnick, 1987, p.3). Sejalan dengan pendapat Resnick tersebut, Conklin (2012, p.14,) menyebutkan bahwa karakteristik HOTS tersebut berkaitan dengan kemampuan berpikir kritis dan berpikir kreatif.

Agar HOTS siswa dapat berkembang dengan baik, maka siswa perlu dibiasakan dengan aktivitas-aktivitas yang melatih HOTS itu sendiri. Melatih HOTS siswa hal yang sangat penting karena merupakan salah satu tujuan kurikulum yang termuat dalam Kerangka Dasar dan Struktur Kurikulum SMA/MA (Mendikbud, 2013, p.4). Selain menggunakan model pembelajaran yang tepat, guru juga perlu memberikan soal-soal atau latihan yang dapat meningkatkan HOTS siswa. Hal ini sejalan dengan pendapat Walsh, Murphy, & Dunbar, (2007, p.2) dalam *Council for Curriculum, Assessment and Examinations* yang mengemukakan pentingnya panduan praktis tentang bagaimana meningkatkan kemampuan berpikir dan bagaimana menilai kemampuan berpikir siswa tersebut. Oleh karena itu, guru perlu menyiapkan instrumen berupa soal-soal yang digunakan untuk mengukur HOTS siswa. Soal-soal tersebut diberikan pada suatu tes/ujian (ulangan harian, UTS atau UKK).

Kegiatan yang dilakukan oleh guru untuk mengetahui kemampuan *higher order thinking* (HOT) siswa disebut kegiatan pengukuran, karena kegiatan apapun yang dilakukan di dunia ini tidak lepas dari pengukuran (Mardapi, 2012, P.6). Keberhasilan guru melakukan uji coba kepada siswa, seperti memberikan tes yang memuat soal HOTS, dapat diketahui melalui suatu pengukuran. Hasil pengukuran yang dilakukan oleh guru, dapat dipercaya apabila dalam beberapa kali guru tersebut melakukan pengukuran terhadap subjek yang sama, diperoleh hasil yang relatif sama (reliabel). Hal itu akan berlaku jika instrumen yang diukur tidak berubah.

Selain reliabel, instrumen yang akan dijadikan sebagai pengukur oleh guru, harus dibuktikan terlebih dahulu kevalidannya. Suatu

instrumen pengukur dapat dikatakan mempunyai validitas yang tinggi jika instrumen yang dibuat menjalankan fungsi ukurnya (Azwar, 2011, p.5). Apabila suatu instrumen memiliki kevalidan yang rendah, maka akan menghasilkan data yang tidak relevan dengan tujuan pengukuran (tidak sesuai dengan fungsi ukur). Mencermati hal tersebut, hendaknya guru dapat membuat atau mengembangkan instrumen HOTS yang valid dan reliabel. Guru dapat melakukan kegiatan pengukuran HOTS matematika siswa dengan menggunakan pendekatan *context-dependent item sets* atau seperangkat butir soal yang terdiri dari pengantar dan diikuti oleh pilihan jawaban, dan *context-dependent item sets* atau latihan menafsirkan. Materi pengantar untuk membuat butir soal tes *HOTS* diantaranya menggunakan gambar, grafik, tabel dan sebagainya yang menuntut siswa pada tingkat penerapan taksonomi tujuan pendidikan dan melibatkan proses kognitif tingkat yang lebih tinggi (Budiman, 2014, p.4).

Instrumen HOTS dikembangkan dengan memperhatikan karakteristik HOTS sebagaimana telah disebutkan. Dalam hal ini, untuk mengembangkan instrumen HOTS harus memuat indikator berpikir kritis dan berpikir kreatif. Hal tersebut perlu diperhatikan agar instrumen HOTS yang dikembangkan benar-benar layak untuk digunakan atau memenuhi kriteria valid dan reliabel. Dalam hal ini, instrumen yang dikembangkan dapat mengukur HOTS siswa dengan tepat.

Pada Ujian Nasional (UN) tahun 2013/2014 pemerintah sudah mencantumkan soal-soal yang memuat HOTS. Khususnya pada soal UN matematika pada jenjang SMA/MA, sudah memuat soal HOTS berupa soal olimpiade internasional sebanyak 2 soal dari 40 soal atau sebesar 5%. Diperkirakan untuk pelaksanaan UN tahun berikutnya, soal-soal olimpiade internasionalnya akan lebih banyak.

Hal ini merupakan salah satu yang mendasari guru untuk dapat membuat atau mengembangkan instrumen HOTS, yaitu instrumen yang mengukur beberapa aspek HOTS siswa. Soal pada UN tersebut, merupakan salah satu instrumen yang dapat dirujuk oleh guru. Tujuannya tidak lain adalah mengidentifikasi kekuatan relatif siswa dan kelemahan berpikir tingkat tinggi siswa (Collins, 2010, p.4). Di samping itu juga, guru dapat mengetahui kesiapan mereka untuk mengikuti Ujian Nasional. Jika guru tidak melakukan hal tersebut, dikhawatir-

kan potensi HOTS yang ada pada diri siswa tidak diketahui dan tidak berkembang.

Memang tidak mudah untuk membuat atau mengembangkan instrumen sebagai pengukur *higher order thinking skills* (HOTS) siswa (Thompson, 2008, p.1; Collins, 2014, p.1). Namun jika ada kemauan maka kesulitan itu akan dapat diatasi. Seberapa banyak butir pengukur *higher order thinking skills* (HOTS) harus diberikan, tergantung pada tujuan dilakukan penilaian dan jenjang pendidikan. Semakin tinggi tingkat pendidikan seseorang, semakin banyak pula butir-butir pengukur *higher order thinking skills* (Kartowagiran, 2011, p.8).

Berdasarkan data pada TIMSS yang ditulis oleh Mullis, et.al (2012, p.76), posisi Indonesia pada domain kognitif penalaran dan domain konten bilangan masing-masing berada pada peringkat terakhir dan peringkat ke-37 dari 43 negara. Berdasarkan jenis kelamin (*gender*) yang diperoleh dari data TIMSS juga, menginformasikan bahwa dari tahun 2007 ke 2011 Indonesia mengalami penurunan nilai.

Hal tersebut dapat disebabkan karena permasalahan yang terjadi di sekolah. Soal-soal yang dikerjakan siswa cenderung lebih banyak menguji aspek ingatan yang kurang melatih keterampilan berpikir tingkat tinggi siswa. Dari hasil survei TIMSS juga diperoleh bahwa kemampuan berpikir anak Indonesia secara ilmiah dianggap masih rendah dilihat. Salah satu faktor penyebabnya antara lain karena peserta didik di Indonesia kurang terlatih dalam menyelesaikan soal-soal yang mengukur HOTS. Selain itu, masalah yang dihadapi oleh guru adalah kemampuan dalam mengembangkan instrumen asesmen *HOTS* masih kurang. Ketersediaan instrumen yang didesain khusus untuk melatih HOTS juga belum banyak, sehingga perlu dikembangkan instrumen pengukur HOTS.

Berdasarkan uraian tersebut, maka membuat dan mengembangkan instrumen untuk mengukur HOTS siswa merupakan suatu hal yang penting sehingga harus menjadi perhatian oleh guru. Pentingnya membuat instrumen HOTS tersebut sejalan dengan pentingnya untuk mengukur sejauh mana *higher order thinking skills* siswa khususnya pada mata pelajaran matematika. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan instrumen untuk mengukur *higher order thinking skills* (HOTS) matematika siswa kelas X SMA yang valid dan reliabel, dan untuk mengetahui *higher order thinking skills* (HOTS) matematika siswa kelas X SMA.

METODE

Penelitian ini merupakan penelitian pengembangan yang menghasilkan sebuah produk, yaitu instrumen pengukur HOTS siswa kelas X. Model penelitian ini diadaptasi dari model pengembangan yang dibuat oleh Mardapi (2008, p.88) yaitu: (1) menyusun spesifikasi tes, (2) menulis soal tes, (3) menelaah soal tes, (4) melakukan uji coba tes, (5) menganalisis butir soal, (6) memperbaiki tes, dan (7) merakit tes. Adapun teknik pengumpulan data yang digunakan dalam penelitian ini adalah tes tertulis, sedangkan instrumen pengumpulan datanya berupa soal uraian, pilihan ganda, dan jawaban singkat.

Uji coba instrumen ini dilakukan di kelas X pada tiga sekolah, yaitu SMAN 4 Kota Yogyakarta, siswa SMAN 6 Kota Yogyakarta, dan SMAN 10 Kota Yogyakarta Tahun Ajaran 2014/2015 dengan total 169 siswa. Data selengkapnya dapat dilihat pada Tabel 1.

Tabel 1. Subjek Uji Coba

Sekolah	Kelas	Banyaknya Siswa
SMAN 4	X.E	28
	X.F	25
	X.G	22
	X.H	19
SMAN 6	X.8	22
	X.9	19
SMAN 10	X.A	18
	X.D	16

Teknik analisis data yang digunakan dalam penelitian ini adalah validitas isi, validitas konstruk, reliabilitas, tingkat kesukaran, daya beda, keefektifan pengecoh, dan kemampuan HOTS matematika siswa. Teknik untuk validitas isi yaitu meminta kepada ahli/expert, dalam hal ini sebagai validator, untuk memeriksa ketepatan dan memberikan penilaian antara kesesuaian butir soal dengan indikator-indikatornya, redaksi penulisan soal, dan kesesuaian pilihan jawaban (pengecoh) pada pilihan ganda. Penilaian tersebut dapat dilihat pada Tabel 2.

Tabel 2. Kriteria Penilaian Butir Instrumen oleh Ahli/Expert

Nilai	Keterangan
1	Tidak Relevan
2	Kurang Relevan
3	Cukup
4	Relevan
5	Sangat Relevan

Setelah diberikan penilaian oleh ahli, selanjutnya peneliti menghitung hasil penilaian menggunakan indeks validitas, diantaranya

dengan indeks yang diusulkan oleh Aiken sebagai berikut.

$$V = \frac{\sum s}{N(c - 1)}, \text{ dimana } s = r - l$$

Keterangan:

r : rating penilai

l : rating penilai kategori terendah

c : kategori tertinggi

N : jumlah penilai/responden

(Aiken, 1980, p.956).

Rentang angka *V* yang mungkin diperoleh adalah antara 0 sampai dengan 1. Semakin tinggi angka *V* (mendekati 1 atau sama dengan 1) maka nilai kevalidan sebuah item/butir soal juga semakin tinggi, dan semakin rendah angka *V* (mendekati 0 atau sama dengan 0) maka nilai kevalidan sebuah item/butir soal juga semakin rendah (Aiken, 1980, p.957).

Selain itu, untuk membuktikan validitas konstruk digunakan analisis faktor eksploratori. Analisis faktor eksploratori dapat dilihat dari persentasi varians yang dilihat dari nilai KMO (*Kaiser Meyer Olkin*). Nilai KMO diperoleh melalui *Software SPSS IBM 20*. Nilai KMO lebih dari 0,5, menunjukkan variabel dan sampel yang digunakan memungkinkan untuk dilakukan analisis lebih lanjut (Santoso, 2006, p.22). Sementara itu, reliabilitas instrumen diestimasi menggunakan teknik konsistensi internal dengan formula *Cronbach-alpha* yang dibantu *Software SPSS IBM 20*. Nilai *Cronbach's Alpha* 0,60 dan kurang dari 1 menunjukkan bahwa instrumen tersebut telah memenuhi kriteria reliabel, sedangkan jika nilai *Cronbach's Alpha* yang kurang dari 0,50 menunjukkan instrumen tidak reliabel (Basuki dan Hariyanto, 2014, p.105; Surapranata, 2009, p.114). Selanjutnya, tingkat kesukaran butir soal pilihan ganda dapat dihitung dengan rumus:

$$p = \frac{\text{jumlah siswa yang menjawab benar butir soal}}{\text{jumlah siswa yang mengikuti tes}}$$

(Nitko, 2011, p.301).

Sedangkan tingkat kesukaran soal uraian dihitung dengan menggunakan rumus:

$$p = \frac{\text{rerata skor butir soal} - \text{kemungkinan skor min}}{\text{kemungkinan skor maks} - \text{kemungkinan skor min}}$$

(Nitko, 2011, p.303).

Setelah dilakukan perhitungan, maka butir soal dapat dikategorikan menjadi butir soal yang sukar, sedang, dan mudah. Pengkategorian tersebut menggunakan kriteria sebagaimana diberikan pada Tabel 3.

Tabel 3. Kategori Tingkat Kesukaran (*p*)

Koefisien	Kategori
$p < 0,3$	Sukar
$0,3 \leq p \leq 0,7$	Sedang
$p > 0,7$	Mudah

(Surapranata, 2009, p.21)

Teknik analisis data untuk daya beda soal pilihan ganda dapat dihitung dengan rumus:

$$DB = \frac{JB_A - JB_B}{n}$$

Dengan:

DB = Indeks daya beda soal

JB_A = Jumlah peserta didik kelompok atas yang menjawab soal itu benar

JB_B = Jumlah peserta didik kelompok bawah yang menjawab soal itu benar

n = Persentase perbandingan ukuran kelompok.

(Nitko, 2011, p.304).

Sementara itu, daya beda soal uraian dihitung dengan menggunakan rumus:

$$DB = \frac{\text{mean kelompok atas} - \text{mean kelompok bawah}}{\text{skor maksimum soal}}$$

(Nitko, 2011, p.304).

Setelah dilakukan perhitungan, maka butir soal dikategorikan menjadi butir soal yang diterima, direvisi, dan ditolak (Tabel 4). Hal itu bergantung koefisien daya bedanya. Jika ada soal yang ditolak, maka dapat dibuang atau diganti dengan butir soal yang baru.

Tabel 4. Kriteria Daya Bada (DB)

Koefisien Daya Bada	Kategori
$DB > 0,30$	Diterima
$0,10 \leq DB < 0,30$	Direvisi
$DB < 0,10$	Ditolak

(Surapranata, 2009, p.47)

Selanjutnya, suatu pengecoh yang efektif harus dipilih oleh beberapa siswa atau minimal dipilih oleh 5% peserta tes (Basuki & Hariyanto, 2014, p.144). Pengecoh dikatakan berfungsi efektif jika banyak dipilih oleh peserta tes dari

kelompok bawah, sebaliknya jika banyak dipilih oleh kelompok atas, maka pengecoh tersebut tidak berfungsi sebagaimana mestinya.

Setelah instrumen memenuhi kriteria yang telah ditetapkan di atas, selanjutnya analisis data dilakukan untuk mengetahui *higher order thinking skills* (HOTS) matematika siswa. Kriteria yang digunakan adalah apabila hasil penilaian menunjukkan nilai HOTS matematika siswa lebih dari atau sama dengan 65 (dalam skala 0-100), maka HOTS matematika siswa telah memenuhi kriteria baik.

HASIL DAN PEMBAHASAN

Validitas Isi

Telaah instrumen dilakukan oleh 2 ahli/pakar pada pelajaran matematika menghasilkan bukti validitas isi instrumen. Hasil dari telaah tersebut menunjukkan bahwa instrumen awal yang dibuat peneliti belum baik. Oleh karena itu, peneliti memperbaiki/merevisi semaksimal mungkin sesuai saran-saran yang dituliskan oleh validator pada lembar-lembaran instrumen. Setelah selesai diperbaiki, instrumen tersebut diberikan kembali kepada validator untuk dinilai masing-masing butirnya. Hasil dari penilaian tersebut dianalisis menggunakan rumus Aiken untuk mengetahui kevalidan masing-masing butir soal.

Tabel 5. Persentase Kevalidan Butir

No.	Nilai V Aiken	Persentase
1.	1	2,67%
2.	0,94	9,33%
3.	0,92	5,33%
4.	0,88	24%
5.	0,83	8%
6.	0,81	24%
7.	0,79	4%
8.	0,75	18,67%
9.	0,71	1,33%
10.	0,69	1,33%
11.	0,63	1,33%
Total		100%

Tabel 6. Pembagian Butir Soal pada Masing-masing Paket Soal

Paket Soal	Butir Soal	Jumlah Butir Soal pada tiap Indikator			Jumlah
		Analisis	Evaluasi	Mencipta	
A	1, 4, 7, 22, 25, 28, 34, 37, 40, 43, 46, 62, 65, 68, 71.	4	4	7	15
B	2, 5, 8, 23, 26, 32, 35, 38, 41, 42, 47, 50, 63, 66, 69, 72.	3	5	8	16
C	3, 6, 21, 24, 27, 33, 37, 39, 42, 45, 48, 61, 64, 67, 70.	4	4	7	15

Berdasarkan hasil penilaian yang disajikan pada Tabel 5 dapat disimpulkan bahwa instrumen valid, karena nilai V Aiken pada semua butir soal mendekati 1. Oleh karena masing-masing butir soal instrumen telah memenuhi kriteria valid, maka instrumen siap untuk diujicobakan. Peneliti tidak mengujicobakan semua instrumen yang dibuat. Instrumen yang diujicobakan sebanyak 46 butir soal. Selain itu, instrumen tersebut diujicobakan pada sekolah-sekolah yang menggunakan KTSP.

Sebelum diujicobakan, peneliti membagi 46 butir soal tersebut menjadi tiga paket soal. Pembagian paket soal tersebut berdasarkan pemerataan indikator HOTS yang diukur. Paket soal A sebanyak 15 butir soal yang terdiri 3 butir pilihan ganda (butir 7, 62, dan 68), 2 butir jawaban singkat (butir 1 dan 43), dan 10 butir uraian (butir 4, 22, 25, 28, 34, 37, 40, 44, 65, dan 71). Pada paket soal B sebanyak 16 butir soal yang terdiri 3 butir pilihan ganda (butir 50, 63, dan 69), 2 butir jawaban singkat (butir 5 dan 66), dan 11 butir uraian (butir 2, 8, 23, 26, 32, 35, 38, 41, 44, 47, dan 72). Pada paket soal C sebanyak 15 butir soal yang terdiri 3 butir pilihan ganda (butir 36, 39, dan 48), 1 butir jawaban singkat (butir 33), dan 11 butir uraian (butir 3, 6, 21, 24, 27, 42, 45, 61, 64, 67, dan 70). Pembagian butir soal pada masing-masing paket soal selengkapnya dapat dilihat pada Tabel 6. Setelah menentukan butir-butir mana saja yang terdapat pada masing-masing paket soal, peneliti membagi paket soal uji coba tersebut pada masing-masing sekolah yang bertujuan untuk menentukan kelas yang mendapatkan paket soal A, B, dan C. Dari hasil pembagian soal tersebut, diputuskan bahwa paket soal A diujicobakan pada tiga kelas dengan banyaknya siswa 71 orang. Paket soal B diujicobakan pada dua kelas dengan banyaknya siswa 38 orang. Sedangkan paket soal C diujicobakan pada tiga kelas dengan banyaknya siswa 59 orang. Data selengkapnya dapat dilihat pada Tabel 7.

Tabel 7. Pembagian Paket soal

Sekolah	Kelas	Jumlah Siswa	Paket Soal
SMAN 4	X.E	28	A
	X.F	25	A
	X.G	22	C
	X.H	19	C
SMAN 6	X.8	22	B
	X.9	19	C
SMAN 10	X.A	18	A
	X.D	16	B

Selanjutnya, mengingat subjek uji coba yang berasal dari sejumlah kelas dan berada pada tiga sekolah yang berbeda, maka selama pelaksanaan uji coba peneliti dibantu oleh rekan/teman sejawat. Selain karena terdapat jadwal yang bersamaan, dikhawatirkan juga jika dilaksanakan oleh peneliti sendiri tidak tepat waktu untuk pelaksanaan uji coba di sekolah berikutnya.

Validitas Konstruk

Setelah selesai melaksanakan uji coba soal di lapangan, selanjutnya peneliti melakukan kegiatan *scoring*/penskoran. Hal ini dilakukan salah satunya untuk membuktikan validitas konstruk yang menggunakan analisis faktor eksploratori. Hasil analisis faktor eksploratori tersebut dapat dilihat pada Tabel 8.

Tabel 8. Hasil Analisis Faktor Eksploratori

Paket Soal	Varians yang dapat Dijelaskan
A	73,7%
B	54,6%
C	50,6%

Berdasarkan Tabel 8 terlihat bahwa persentasi varians yang dapat dijelaskan lebih dari 50%. Oleh karena itu dapat disimpulkan bahwa variabel dan sampel yang digunakan memungkinkan untuk dilakukan analisis lebih lanjut.

Reliabilitas instrumen

Hasil estimasi reliabilitas instrumen HOTS yang dikembangkan menunjukkan hasil yang memuaskan. Berdasarkan hasil estimasi reliabilitas instrumen pada Tabel 9 diperoleh informasi bahwa nilai koefisien *Cronbach's Alpha* pada paket soal A adalah 0,738 yang berarti paket soal A reliabel. Selain itu, nilai koefisien *Cronbach's Alpha* paket soal B sebesar 0,658 yang berarti paket soal B juga reliabel, sedangkan paket soal C diperoleh nilai koefisien *Cronbach's Alpha* adalah 0,488 yang berarti paket soal C tidak reliabel. Namun pada paket soal C mendekati nilai 0,5. Oleh karena itu, paket soal C juga dapat digunakan untuk tujuan penelitian.

Tabel 9. Reliabilitas Instrumen

Paket Soal	Nilai <i>Cronbach's Alpha</i>
A	0,738
B	0,658
C	0,488

Tingkat Kesukaran

Langkah kegiatan analisis selanjutnya adalah menentukan tingkat kesukaran untuk masing-masing butir soal pada setiap paket soal. Tingkat kesukaran butir soal pada paket A, paket B, dan paket C dapat dilihat Tabel 10.

Tabel 10. Persentase Jumlah Butir Soal Berdasarkan Tingkat Kesukarannya

Paket Soal	Kategori Tingkat kesukaran	Presentase
A	Mudah	6,67%
	Sedang	46,67%
	Sulit	46,67%
B	Mudah	12,50%
	Sedang	12,50%
	Sulit	75%
C	Mudah	0%
	Sedang	53,33%
	Sulit	46,67%

Berdasarkan hasil analisis tingkat kesukaran butir soal pada masing-masing paket soal yang diberikan pada Tabel 10, diperoleh informasi bahwa sebagian besar soal pada paket A terdistribusi pada tingkat kesukaran yang sedang dan sulit. Hanya 6,67% soal yang berada dalam kategori mudah. Pada paket B, sebagian besar (75%) butir soal termasuk kategori sulit, sedangkan sisanya yang berkategori mudah dan sedang masing-masing sebesar 12,50. Hasil yang hampir sama ditunjukkan oleh soal pada paket C, dimana seluruh butir soal terdistribusi pada kategori sulit dan sedang, atau dalam hal ini tidak ada soal yang terkategori mudah.

Daya Beda

Langkah kegiatan analisis selanjutnya adalah menentukan daya pembeda untuk masing-masing butir soal pada setiap paket soal. Hasil analisis data menunjukkan bahwa tidak semua butir soal daya pembedanya baik, sehingga ada butir-butir soal yang perlu direvisi/diperbaiki atau diganti/dibuang. Selengkapnya peneliti disajikan pada Tabel 11.

Tabel 11. Daya Pembeda

Paket Soal	Kategori Daya Beda	Persentase
A	Diterima	53,33%
	Direvisi	26,67%
	Ditolak	20%
B	Diterima	37,50%
	Direvisi	18,75%
	Ditolak	43,75%
C	Diterima	20%
	Direvisi	60%
	Ditolak	20%

Berdasarkan Tabel 11, diketahui pada paket soal A, butir soal yang diterima sebesar 53,33%, butir soal yang perlu direvisi sebesar 26,67%, dan butir soal yang ditolak sebesar 20%. Pada paket soal B, butir soal yang diterima sebesar 37,50%, butir soal yang perlu direvisi sebesar 18,75%, dan butir soal yang ditolak sebesar 43,75%. Pada paket soal C, butir soal yang diterima sebesar 20%, butir soal yang perlu direvisi sebesar 60%, dan butir soal yang ditolak sebesar 20%.

Keefektifan Pengecoh

Pada paket soal A, diketahui bahwa pengecoh tidak berfungsi pada butir 3 *option* B karena tidak ada yang memilih satu siswapun. Pada butir 12 semua *option* tidak berfungsi, sedangkan pada butir 14, pengecoh yang berfungsi hanya *option* A. Pengecoh/distraktor yang tidak berfungsi harus diganti. Selengkapnya dapat dilihat pada Tabel 12.

Tabel 12. Pengecoh pada Paket Soal A

Option	Persentase Terpilihnya Pengecoh		
	Butir 3	Butir 12	Butir 14
A	14,08	0,00	5,63
B	0,00	52,11*	1,41
C	5,63	0,00	2,82
D	7,04	1,41	1,41
E	33,80*	1,41	50,70*

Keterangan: * Kunci jawaban

Pada paket soal B, diketahui bahwa pengecoh tidak berfungsi pada butir 12 *option* C dan D. Pengecoh pada butir 13 *option* B, C, dan E tidak berfungsi, sedangkan pada butir 15, pengecoh yang tidak berfungsi adalah *option* C dan E. Selengkapnya dapat dilihat Tabel 13.

Tabel 13. Pengecoh Pada Paket Soal B

Option	Persentase Terpilihnya Pengecoh		
	Butir 12	Butir 13	Butir 15
A	5,26	28,95	5,26
B	23,68	2,63	10,53
C	2,63	0,00	2,63
D	0,00	0,00*	73,68*
E	2,63*	0,00	0,00

Keterangan: * Kunci jawaban

Pada paket soal C, diketahui bahwa pengecoh tidak berfungsi pada butir 7 *option* B dan E. Pada butir 8 semua *option* berfungsi kecuali *option* B, sedangkan pada butir 11 pengecoh yang tidak berfungsi adalah *option* D dan E. Selengkapnya dapat dilihat pada Tabel 14.

Tabel 14. Pengecoh Pada Paket Soal C

Option	Persentase Terpilihnya Pengecoh		
	Butir 7	Butir 8	Butir 11
A	3,39	54,24	71,19
B	5,08	1,69	13,56
C	67,80*	13,56	3,39*
D	10,17	15,25	1,69
E	0,00	8,47*	1,69

Keterangan: * Kunci jawaban

HOTS Matematika Siswa

Secara umum, hasil analisis dari seluruh skor siswa menunjukkan bahwa kemampuan *higher order thinking* (HOT) matematika siswa pada uji coba masih di bawah 65, yaitu dengan hasil nilai rata-rata sebesar 26,38 dalam skala 100. Nilai rata-rata ini mengidentifikasi bahwa kemampuan *higher order thinking* (HOT) matematika siswa pada uji coba kurang baik. Secara rinci nilai rata-rata siswa pada tiap-tiap sekolah dapat dilihat pada Tabel 15. Berdasarkan Tabel 15, diperoleh informasi bahwa siswa SMAN 4 Kota Yogyakarta memiliki kemampuan *higher order thinking* (HOT) yang lebih tinggi dibandingkan SMAN 10 dan SMAN 6, dan SMAN 10 memiliki kemampuan *higher order thinking* (HOT) yang lebih baik SMAN 6.

Selain itu, informasi pada Tabel 15 juga menunjukkan paket soal dengan nilai rata-rata yang paling tinggi. Hasil tersebut antara lain menunjukkan bahwa rata-rata HOTS siswa yang mengerjakan paket soal A lebih tinggi dibandingkan siswa yang mengerjakan paket soal B dan C. Rata-rata HOTS siswa pada paket A yaitu, sebesar 30,30. Nilai ini lebih tinggi dibandingkan dengan rata-rata HOTS siswa yang mengerjakan paket soal C yaitu 26,38. Semen-

tura itu, rata-rata skor HOTS siswa yang mengerjakan paket soal B adalah paling rendah yaitu 25,36. Hasil ini sejalan dengan hasil analisis distribusi tingkat kesukaran butir soal pada masing-masing pake sebagaimana yang telah diterangkan pada Tabel 10.

Tabel 14. Nilai Rata-rata Siswa pada Tiap Kelas Uji Coba

Sekolah	Kelas	Paket Soal	Nilai Rata-rata (Skala 100)
SMAN 10	X.A	A	19,14
	X.D	B	31,90
SMAN 6	X.8	B	18,81
	X.9	C	24,34
SMAN 4	X.E	A	31,68
	X.F	A	40,07
	X.G	C	33,59
	X.H	C	22,84
Nilai rata-rata keseluruhan			26,38

Hasil analisis HOTS matematika siswa sebagaimana yang telah diuraikan pada Tabel 15 juga menggambarkan bahwa HOTS siswa yang pada tiga sekolah tersebut masih rendah. Hal tersebut mengindikasikan bahwa aktivitas pembelajaran yang telah ditempuh siswa selama ini belum berorientasi pada pengembangan HOTS mereka. Hal ini sejalan dengan hasil riset yang dilakukan oleh Musfiqi & Jailani (2014); (Riadi & Retnawati, 2014) yang masing-masing mengembangkan bahan ajar dan perangkat pembelajaran yang berorientasi pada HOTS siswa. Kemampuan HOTS siswa sebelum mengikuti pembelajaran (*pretest*) yang berorientasi pada HOTS masih sangat rendah (Musfiqi & Jailani, 2014, p. 55; Riadi & Retnawati, 2014, p.134).

<u>Sebelum Revisi</u>	<u>Setelah Revisi</u>
Selidiki dan tuliskan aturan eksponen yang digunakan dalam menyelesaikan soal berikut. Bentuk sederhana dari $\frac{\sqrt[8]{81}}{\sqrt[6]{3}}$ adalah ...	Perhatikan langkah-langkah dalam menyederhanakan soal berikut ini. $\frac{\sqrt[8]{81}}{\sqrt[6]{3}} = \frac{81^{\frac{1}{8}}}{3^{\frac{1}{6}}} = \frac{(3^4)^{\frac{1}{8}}}{3^{\frac{1}{6}}} = \frac{3^{\frac{1}{2}}}{3^{\frac{1}{6}}} = 3^{\frac{1}{2}-\frac{1}{6}} = 3^{\frac{1}{3}} = \sqrt[3]{3}$ Sifat-sifat eksponen yang digunakan dalam menyederhanakan bentuk di atas adalah ... A. $\sqrt[n]{a^m} = a^{\frac{m}{n}}$ B. $\sqrt[n]{a^m} = a^{\frac{m}{n}}$ dan $(a^m)^n = a^{m.n}$ C. $(a^m)^n = a^{m.n}$, dan $\frac{a^m}{a^n} = a^{m-n}$ D. $\sqrt[n]{a^m} = a^{\frac{m}{n}}$, $(a^m)^n = a^{m.n}$, dan $\frac{a^m}{a^n} = a^{m-n}$ E. $\sqrt[n]{a^m} = a^{\frac{m}{n}}$, $(a^m)^n = a^{m.n}$, $\frac{a^m}{a^n} = a^{m-n}$, dan $a^m . a^n = a^{m+n}$

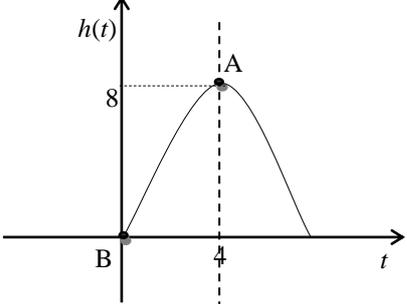
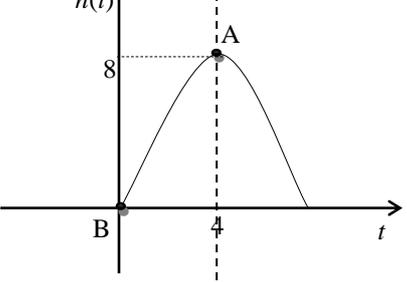
Gambar 1. Revisi yang dilakukan pada Butir 1 Pada Paket Soal A

Tahapan penelitian selanjutnya adalah perbaikan instrumen. Perbaikan instrumen dilakukan berdasarkan tahapan analisis yang sudah dilakukan, seperti tingkat kesukaran, daya pembeda, dan analisis distraktor/pengecoh. Oleh karena itu, peneliti membagi instrumen yang sudah diujicobakan ini ke dalam tiga kategori. Kategori pertama adalah kategori instrumen yang diterima, kategori kedua adalah kategori instrumen yang direvisi, dan kategori ketiga adalah kategori instrumen yang ditolak. Peneliti melakukan perbaikan terhadap butir soal yang tergolong kategori kedua pada setiap paket soal, sedangkan butir soal yang tergolong kategori ketiga, butir soal tersebut akan dibuang.

Kategori dua ini diperbaiki dengan alasan tingkat kesukaran yang mudah dan daya pembeda yang rendah. Selain itu juga, pengecoh yang kurang baik menjadi alasan bahwa butir soal tersebut harus diperbaiki. Namun, pada kategori tiga, butir soal yang ditolak akan

dibuang dengan alasan tingkat kesukaran yang terlalu mudah dan daya pembeda terlalu rendah.

Pada paket soal A, sebanyak 9 butir tergolong kategori dua, yaitu butir 1, 2, 3, 4, 6, 10, 11, 12, dan 14, sehingga butir-butir tersebut perlu direvisi/diperbaiki. Sebanyak 3 butir soal tergolong kategori tiga, yaitu butir 5, 9, dan 15, sehingga butir-butir tersebut dibuang. Pada paket soal B, diketahui sebanyak 8 butir tergolong pada pada kategori dua, yaitu butir 1, 2, 3, 5, 7, 9, 14, dan 16, sehingga butir-butir tersebut perlu direvisi. Sebanyak 7 butir tergolong pada kategori tiga, yaitu butir 4, 6, 8, 10, 12, 13, dan 15, sehingga butir-butir tersebut dibuang. Pada paket soal C, diketahui sebanyak 9 butir soal tergolong kategori dua, yaitu 1, 2, 4, 5, 7, 8, 10, 12, dan 14, sehingga butir-butir perlu direvisi, sedangkan sebanyak 3 butir tergolong kategori tiga, yaitu 3, 9, dan 11, sehingga butir-butir tersebut dibuang. Berikut ini masing-masing contoh perbaikan butir instrumen pada paket soal A, paket B, dan paket C.

Sebelum Revisi	Setelah Revisi
<p>Lina sedang bermain golf di belakang rumah. Ia memukul bola dengan lintasan bola seperti pada grafik di samping dalam meter dan detik.</p>	<p>Lina sedang bermain golf di belakang rumah. Ia memukul bola dengan lintasan bola seperti pada grafik di bawah ini yang dinyatakan dalam meter untuk ketinggian/$h(t)$ dan detik untuk waktu/t.</p>
	
<p>Apa yang dimaksud dengan titik A dan titik B pada grafik berikut.</p>	<p>Berdasarkan grafik di atas, maka yang dimaksud dengan titik A dan titik B yang tepat adalah ...</p> <ol style="list-style-type: none"> Titik A menunjukkan posisi tertinggi bola dalam lintasannya, dengan koordinat (4, 8). Titik B menunjukkan bahwa bola tersebut berada pada posisi dimana bola akan dipukul, dengan koordinat (0, 0). Titik A menunjukkan posisi tertinggi bola dalam lintasannya, dengan ketinggian 8 m dan waktu yang dibutuhkannya 4 detik. Titik B menunjukkan bahwa bola tersebut berada pada posisi dimana bola akan dipukul, artinya ketinggian bola 0 m dan waktu 0 detik. Titik A menunjukkan posisi tertinggi bola dalam lintasannya dan titik B menunjukkan bahwa bola tersebut berada pada posisi dimana bola akan dipukul. Titik A menunjukkan ketinggian 8 m dengan waktu yang dibutuhkannya 4 detik dan titik B menunjukkan bola belum dipukul. Titik A menunjukkan posisi tertinggi bola dalam lintasannya, dengan ketinggian 4 m dan waktu yang dibutuhkannya 8 detik. Titik B menunjukkan bahwa bola tersebut berada pada posisi dimana bola akan dipukul, artinya ketinggian bola 0 m dan waktu 0 detik.

Gambar 2. Revisi pada Butir 14 Paket Soal B

<u>Sebelum Revisi</u>	<u>Setelah Revisi</u>
<p>Dengan dana yang terbatas, PT. Griya Berkah akan membangun perumahan sebanyak 250 unit pada tanah seluas 45.500 m². Dengan luas tanah 150 m² akan di bangun rumah tipe A sedangkan tipe B dibangun di atas tanah seluas 200 m², dan di tengah-tengah perumahan akan dibuat jalan seluas 500 m². Ketika diadakan rapat oleh pemilik PT. Griya Berkah, seorang karyawan mengemukakan pendapat bahwa dengan 150 m² rumah tipe A dapat dibuat sebanyak 180 unit, sedangkan dengan 200 m² rumah tipe B dapat dibuat sebanyak 45 unit. Andaikan kamu adalah seorang arsitek, apakah pendapat karyawan itu tepat?</p>	<p>Dengan dana yang tidak cukup besar, PT. Griya Berkah akan membangun perumahan sebanyak 250 unit pada tanah seluas 45.500 m². Dengan luas tanah 150 m² akan di bangun rumah tipe A sedangkan dengan luas tanah 200 m² akan di bangun rumah tipe B. PT Griya Berkah juga akan membangun jalan dan di tengah-tengah perumahan seluas 500 m². Ketika diadakan rapat oleh pemilik PT. Griya Berkah, seorang karyawan menyatakan pendapatnya dan pemilik PT Griya Berkah setuju dengan pendapat karyawan tersebut. Pernyataan karyawan tersebut adalah ...</p> <p>A. Rumah tipe A dapat dibangun sebanyak 180 unit pada tanah seluas 150 m², sedangkan rumah tipe B dapat dibangun sebanyak 45 unit pada tanah seluas 200 m².</p> <p>B. Rumah tipe A dapat dibangun sebanyak 200 unit pada tanah seluas 150 m², sedangkan rumah tipe B dapat dibangun sebanyak 50 unit pada tanah seluas 200 m².</p> <p>C. Rumah tipe A dapat dibangun sebanyak 220 unit pada tanah seluas 150 m², sedangkan rumah tipe B dapat dibangun sebanyak 55 unit pada tanah seluas 200 m².</p> <p>D. Rumah tipe A dapat dibangun sebanyak 240 unit pada tanah seluas 150 m², sedangkan rumah tipe B dapat dibangun sebanyak 60 unit pada tanah seluas 200 m².</p> <p>E. Rumah tipe A dapat dibangun sebanyak 260 unit pada tanah seluas 150 m², sedangkan rumah tipe B dapat dibangun sebanyak 65 unit pada tanah seluas 200 m².</p>

Gambar 2. Revisi pada Butir 14 Paket Soal C

Setelah butir-butir soal yang direvisi telah selesai diperbaiki dan butir-butir soal yang ditolak dibuang, langkah berikutnya adalah merakit/menyusun butir-butir soal tersebut. menjadi satu kesatuan yang disebut instrumen HOTS. Pada tahap ini juga, peneliti mengelompokkan butir-butir soal ke dalam satu kesatuan instrumen Semester Ganjil dan Semester Genap. Pada setiap semester, peneliti mengelompokkan jenis/bentuk butir soal. Urutan/nomor pertama akan disusun jenis/bentuk butir soal berupa pilihan ganda, selanjutnya disusun jenis/bentuk butir soal berupa jawaban singkat, dan yang terakhir adalah jenis/bentuk butir soal uraian.

SIMPULAN

Berdasarkan hasil analisis yang telah peneliti lakukan, dapat diambil kesimpulan bahwa instrumen pengukur *higher order thinking skills* (HOTS) matematika siswa kelas X dikategorikan valid, karena semua nilai *V* di atas 0,3. Instrumen yang terdiri dari 75 butir soal ini, diujicobakan sebanyak 46 butir soal. Instrumen yang diujicobakan dibagi menjadi tiga paket soal. Paket soal A berjumlah 15 butir, paket soal B berjumlah 16 butir, dan paket soal C berjumlah 15 butir. Paket soal A dan paket B menghasilkan paket soal yang reliabel, dengan masing-masing nilai *Cronbach's Alpha* sebesar

0,738 dan 0,658, sedangkan paket soal C tidak reliabel dengan nilai *Cronbach's Alpha* sebesar 0,488. Selain itu, hasil dari uji coba instrumen dapat disimpulkan bahwa *higher order thinking skills* (HOTS) matematika siswa kelas X kurang baik. Hal ini ditunjukkan dengan nilai rata-rata hasil uji coba yang kurang dari 65, yaitu sebesar 26,38 dalam skala 100.

DAFTAR PUSTAKA

- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and psychological measurement*, 40(4), 955-959.
- Azwar, S. (2011). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Belajar.
- Basuki, I., & Hariyanto. (2014). *Asesemen pembelajaran*. Bandung: PT. Remaja Rosdakarya.
- Budiman, A., & Jailani, J. (2014). Pengembangan instrumen asesmen higher order thinking skill (HOTS) pada mata pelajaran matematika SMP kelas VIII semester 1. *Jurnal Riset Pendidikan Matematika*, 1(2). Retrieved from <http://journal.uny.ac.id/index.php/jrpm/article/view/2671>

- Collins, R. (2014). Skills for the 21st century: teaching higher-order thinking. *Curriculum & Leadership Journal. Volume 12 Issue 14*. Diambil pada tanggal 10 Juli 2015 dari http://www.curriculum.edu.au/leaderteaching_higher_order_thinking,37431.html issueID=12910
- Collins, V. (2010). *Higher order thinking (HOT) program assessment plan*. Fort Worth, Texas: University of North Texas Health Science Center.
- Conklin, W. (2012). *Higher-order thinking skills to develop 21st century learners*. Huntington Beach, California: Shell Education.
- Mendikbud. (2013). *Peraturan menteri pendidikan dan kebudayaan nomor 69, tahun 2013, tentang kerangka dasar dan struktur kurikulum sekolah menengah atas/madrasah aliyah*.
- Musfiqi, S., & Jailani, J. (2014). Pengembangan bahan ajar matematika yang berorientasi pada karakter dan higher order thinking skill (HOTS). *PYTHAGORAS: Jurnal Pendidikan Matematika*, 9(1), 45–59. Retrieved from <http://journal.uny.ac.id/index.php/pythagoras/article/view/9063/7398>
- Kartowagiran, B. (2011). *Penulisan butir soal*. Yogyakarta: Pascasarjana UNY. Makalah disampaikan pada Pelatihan penulisan dan analisis butir bagi guru SMP Provinsi D.I. Yogyakarta pada tanggal 23, 26, dan 28 Juli 2011 di Pascasarjana UNY.
- Khan, W.B. (2011). A study of lower-order and higher-order questions at secondary level. Pakistan: Canadian Center of Science and Education. *Asian Social Science Vol. 7, No. 9; September 2011*.
- Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan non tes*. Yogyakarta: Mitra Cendika.
- Mardapi, D. (2012). *Pengukuran, penilaian, & evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- Mullis, Ina V.S., et.al, (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA, USA: Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Nitko, A.J. & Brookhart, S.M. (2011). *Educational assessment of students. Xth edition*. Upper Saddle River, New Jersey: Prentice Hall Englewood Cliffs.
- Resnick, L. B. (1987). *Education and learning to think*. Washington, D.C: National Academy Press.
- Riadi, A., & Retnawati, H. (2014). Pengembangan Perangkat Pembelajaran untuk Meningkatkan HOTS pada Kompetensi Bangun Ruang Sisi Datar. *PYTHAGORAS: Jurnal Pendidikan Matematika*, 9(2), 126–135. Retrieved from <http://journal.uny.ac.id/index.php/pythagoras/article/view/9074/pdf>
- Santoso, S. (2006). *Seri solusi bisnis berbasis TI: Menggunakan SPSS untuk statistik multivariat*. Jakarta: Elex Media Komputindo.
- Surapranata, S. (2009). *Analisis, validitas, reliabilitas, dan interpretasi hasil tes implementasi kurikulum 2004*. Bandung: PT. Remaja Rosdakarya.
- Thompson, T. (2008). Mathematics teachers' interpretation of higher-order thinking in bloom's taxonomy. *International Electronic Journal of Mathematics Education, Volume 3, Number 2, July 2008*.
- Walsh, G., Murphy P., & Dunbar, C. (2007). *Thinking skills in the early years: a guide for practitioners*. Stranmillis
- Yen, T. S., & Halili, S. H. (2015). Effective teaching of higher-order thinking (hot) in education. *The Online Journal of Distance Education and e-Learning (TOJDEL). Volume 3, Issue 2*.