

PENGEMBANGAN INSTRUMEN PENGUKUR BERPIKIR KRITIS MATEMATIKA SISWA SMA KELAS X

Zaenal Arifin ¹⁾

¹⁾Dosen Program Studi Pendidikan Matematika Universitas Majalengka
Jl. Universitas Majalengka No.1, Majalengka
Email: arifin1169@gmail.com

ABSTRAK

Penelitian ini bertujuan: (1) untuk menghasilkan instrumen pengukur *higher order thinking skills* (HOTS) matematika siswa kelas X yang valid dan reliabel, dan (2) untuk mengetahui kemampuan *higher order thinking* (HOT) matematika siswa kelas X dilihat dari hasil uji coba siswa. Penelitian ini merupakan penelitian pengembangan instrumen dengan langkah meliputi: (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. Uji coba instrumen dilakukan di kelas X pada tiga sekolah, yaitu SMAN 4 Kota Yogyakarta sebanyak empat kelas, SMAN 6 Kota Yogyakarta sebanyak dua kelas, dan SMAN 10 Kota Yogyakarta sebanyak dua kelas. Jumlah keseluruhan subjek uji coba sebanyak 169 siswa. Teknik dalam mengumpulkan data pada penelitian pengembangan ini adalah tes tertulis. Sedangkan instrumen pengumpulan datanya berupa soal uraian, pilihan ganda dan jawaban singkat. Instrumen yang terdiri dari 75 butir soal ini, diujicobakan sebanyak 46 butir soal. Instrumen yang diujicobakan dibagi menjadi tiga paket soal. Soal paket A sebanyak 15 butir, soal paket B sebanyak 16 butir, dan soal paket 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, soal paket A dan paket B menghasilkan soal paket yang reliabel, dengan masing-masing nilai *Cronbach's Alpha* sebesar 0,738 dan 0,658. Sedangkan soal paket C tidak reliabel yang ditunjukkan dengan nilai *Cronbach's Alpha* sebesar 0,488. Hasil dari uji coba instrumen dapat disimpulkan bahwa kemampuan *higher order thinking* (HOT) 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 CRITICAL THINKING OF MATHEMATICS IN CLASS X SENIOR HIGH SCHOOL.

Zaenal Arifin¹⁾

¹⁾Dosen Program Studi Pendidikan Matematika Universitas Majalengka
Jl. Universitas Majalengka No.1, Majalengka
Email: arifin1169@gmail.com

ABSTRACT

This research aims to: (1) to know an instrument to measure higher order thinking skills (HOTS) of mathematics of class X student which is valid and reliable, and (2) to know the higher order thinking skills (HOTS) of mathematics of class X students seen from the trial test. This research is the development of instruments with steps: (1) preparing test specifications, (2) writing the items, (3) reviewing the items, (4) doing trial tests, (5) analyzing the items, (6) correcting the items, and (7) arranging the items. The instruments trial was done in class X at three schools, that is SMAN 4 of Yogyakarta as many as four classes, SMAN 6 Yogyakarta as many as two classes, and SMAN 10 Yogyakarta as many as two classes. The total number of subjects tested was as many as 169 students. The technique in collecting data on this research is a test, while the data collection instruments in this research are essay, multiple choice and short-answer items. The instrument consists of 75 items, but only 46 items were tested. Instruments tested were 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 the ability of higher order thinking (HOT) of mathematics of class X student is poor. It can be seen from the average value of the test results of 26.38 in the scale of 100.

Key word: *instrument development, higher order thinking skills (HOTS), validity, reliability.*

PENDAHULUAN

Latar Belakang

Peran pendidikan di sekolah yang terdapat dalam kurikulum, berfokus untuk mengembangkan sumber daya manusia seperti 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. Oleh karena itu, salah satu indikasi keberhasilan peningkatan sumber daya manusia dalam bidang pendidikan adalah siswa memiliki kemampuan berpikir tingkat tinggi (HOTS) yang baik, karena tujuan utama pembelajaran pada abad 21 ini, adalah untuk mengembangkan dan meningkatkan HOT siswa (Yen & Halili, 2015: 41).

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: 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.

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 di atas, 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: 4). Disamping itu juga, guru dapat mengetahui kesiapan mereka untuk mengikuti Ujian Nasional. Jika guru tidak melakukan hal tersebut, dikhawatirkan potensi HOTS yang ada pada diri siswa tidak diketahui dan tidak berkembang.

Salah satu dugaan peneliti mengenai hal tersebut adalah karena siswa tidak terbiasa mengerjakan soal-soal yang memuat HOTS, baik dalam proses pembelajaran atau ketika tes seperti ulangan harian, UTS, maupun UKK. Hal ini didukung oleh hasil prasurvei berupa wawancara yang dilakukan peneliti di dua Kabupaten yang berbeda (Kabupaten Cirebon dan Kabupaten Sleman), yaitu guru belum terbiasa memberikan soal-soal yang memuat HOTS kepada siswa. Peneliti berpendapat bahwa guru belum terbiasa memberikan soal-soal yang memuat HOTS kepada siswa. Asumsi ini didukung penelitian yang dilakukan oleh Khan (2011: 1), bahwa guru jarang sekali memberikan soal-soal yang memiliki level tinggi. Oleh karena itu, Peneliti menyimpulkan bahwa belum banyak mengenai acuan atau contoh mengenai instrumen yang memuat HOTS.

Berdasarkan hal tersebut, peneliti berpikir bahwa membuat dan mengembangkan instrumen HOTS merupakan sesuatu yang perlu dilakukan. Permasalahan penelitian dapat dirumuskan seperti berikut: (1) Apakah instrumen pengukur *higher order thinking skills* (HOTS) matematika siswa kelas X valid dan reliabel? (2) Bagaimana kemampuan *higher order thinking* (HOT) matematika siswa kelas X dilihat dari hasil uji coba siswa?

Tujuan Pengembangan

Penelitian pengembangan ini mempunyai beberapa tujuan, yaitu:

1. Untuk menghasilkan instrumen pengukur *higher order thinking skills* (HOTS) matematika siswa kelas X yang valid dan reliabel.
2. Untuk mengetahui kemampuan *higher order thinking* (HOT) matematika siswa kelas X dilihat dari hasil uji coba siswa.

Kajian Teori

a. Karakteristik *Higher Order Thinking Skills* (HOTS)

Menurut Conklin (2012: 14), *higher order thinking skills* meliputi berpikir kritis dan berpikir kreatif. Berpikir kritis adalah istilah yang umumnya dikaitkan dengan kemampuan berpikir tingkat tinggi yang ditandai dengan analisis yang cermat dan pertimbangan. Setiap orang mengharapkan lembaga pendidikan dapat membuat siswa untuk berpikir kritis. Berpikir kritis adalah *self-guided*, yakni berpikir disiplin diri yang mencoba untuk mencapai kualitas tingkat tertinggi dengan cara berpikiran adil (Conklin, 2012: 14).

Orang-orang yang berpikir kritis secara konsisten berusaha untuk hidup secara rasional, cukup, dan empati. Orang-orang yang berpikir kritis, ketika dihadapkan suatu permasalahan, pemikiran kritisnya akan mendorong terus untuk tetap belajar menyelesaikan permasalahan tersebut. Mereka berkomitmen untuk berpikir logis tentang permasalahan sampai akhirnya selesai. Mereka berjuang untuk menyingkirkan bias yang datang ke dalam pemikirannya dan berusaha untuk melihat situasi dengan cara baru sehingga dapat dianalisis dan dievaluasi secara logis, dan mereka merefleksikan apa yang sudah mereka pelajari.

Keterampilan berpikir kritis dan keterampilan berpikir kreatif ini datang bersama-sama di bawah kemampuan berpikir tingkat tinggi yang didasarkan pada pemikiran tingkat yang lebih rendah. Kemampuan berpikir tentang topik pada tingkat yang lebih tinggi, harus terlebih dahulu memahami topik yang diberikan. Seorang siswa ketika berpikir tingkat tinggi, harus mengetahui fakta-fakta dasar, memahami konsep, dan menerapkan apa yang mereka ketahui, sehingga mereka dapat memilih topik terpisah melalui analisis, membuat penilaian, atau sesuatu yang baru berdasarkan ide.

b. Ciri-ciri dan Indikator *Higher Order Thinking Skills* (HOTS)

Higher order thinking skills adalah kemampuan berpikir yang meliputi berpikir kritis yang diperoleh melalui proses kognitif menganalisis, mengevaluasi dan mencipta. Proses menganalisis melibatkan proses mental siswa untuk memecah-mecah materi menjadi bagian-bagian kecil. Berbeda dengan proses mengevaluasi dimana siswa melibatkan proses mentalnya untuk membuat keputusan berdasarkan kriteria/standar tertentu, sedangkan proses mencipta siswa melibatkan proses mentalnya untuk menyusun elemen-elemen menjadi sebuah keseluruhan yang fungsional (Anderson & Krathwohl, 2012: 121-128). Untuk mengukur kemampuan berpikir tingkat tinggi seseorang, dapat menggunakan instrumen HOTS dengan indikator sebagai berikut:

Tabel 1
Indikator HOTS

HOTS	Indikator	Sub Indikator
Berpikir kritis	1. Analisis	1.1 Membedakan 1.2 Mengorganisasikan 1.3 Mengatribusikan
	2. Evaluasi	2.1 Memeriksa 2.2 Mengkritik

c. *Higher Order Thinking Skill* pada Matematika

Terkait dengan proses terbentuknya, matematika merupakan pengetahuan yang dimiliki manusia. Pengetahuan ini timbul karena kebutuhan manusia untuk memahami alam sekitar. Alam dijadikan sumber-sumber ide untuk memperoleh konsep matematika melalui abstraksi dan idealisasi. Untuk dapat menggunakan matematika dalam memahami alam sekitar, diperlukan suatu kemampuan berpikir tingkat tinggi.

Selain itu, materi matematika sangat kompleks, sehingga membutuhkan keterampilan berpikir yang sangat tinggi. Seperti yang dikemukakan James & James (1992: 262), matematika terdiri dari tiga cabang utama, yakni aljabar, geometri dan analisis. Dari ketiga cabang ini, dalam pembelajaran matematika, aljabar dipelajari oleh siswa terlebih dahulu pada pendidikan formal.

Oleh karena itu, kemampuan berpikir tingkat tinggi akan membantu materi yang kompleks dengan memecah-mecah menjadi bagian-bagian, mendeteksi hubungan, mengkombinasikan informasi yang baru dengan informasi yang telah dikenal secara kreatif yang terhubung dalam konteks, mengombinasikan dan menggunakan langkah sebelumnya untuk mengevaluasi atau membuat keputusan. Kemampuan berpikir tingkat tinggi (*higher order thinking skills*) ini dilatihkan kepada siswa melalui pembelajaran, yaitu dengan cara memberikan konteks masalah nyata (Collins, 2014: 1; Miri, et.al, 2007: 1).

Berpikir tingkat tinggi (*higher order thinking*) mencakup pemikiran kritis dapat digunakan untuk menyelesaikan permasalahan nyata. Kemampuan berpikir pada matematika diaktivasi ketika siswa sedang menyelesaikan suatu permasalahan nyata. Jika tidak dengan masalah nyata, maka itu bukan *higher order thinking* (Thompson, 2012: 1).

d. Instrumen HOTS Matematika Siswa

Menurut Azwar (2010: 2), menyatakan bahwa para ahli psikometri telah menetapkan kriteria bagi alat ukur, seperti instrumen, untuk dinyatakan sebagai alat ukur yang baik. Kriteria tersebut antara lain adalah valid, reliabel, standar, ekonomis dan praktis. Menurut Gronlund (2009: 70) menyatakan bahwa karakteristik utama yang harus dimiliki oleh sebuah alat ukur dapat diklasifikasikan menjadi karakter validitas, reliabilitas dan tingkat kegunaannya. Secara singkat dan jelas, dalam membuat instrumen penelitian pengembangan, Setyosari (2013: 207) dan Depdiknas (2008: 4) menyatakan bahwa validitas dan reliabilitas menjadi dua hal yang harus benar-benar diperhatikan. Berdasarkan hal di atas, peneliti menyimpulkan bahwa validitas dan reliabilitas menjadi dua hal terpenting yang harus diperhatikan oleh pembuat instrumen.

1) Validitas

Gronlund (2009: 70) menyebutkan bahwa validitas adalah ketepatan interpretasi yang diperoleh dari hasil penilaian. Lebih rinci Azwar (2010: 5) menyebutkan bahwa validitas berasal dari kata *validity* yang memiliki arti sejauh mana ketepatan dan kecermatan suatu alat ukur dalam melakukan fungsi ukurnya. Namun secara khusus Allen & Yen (1979: 97) menyatakan bahwa validitas dari suatu perangkat tes dapat diartikan merupakan kemampuan suatu tes untuk mengukur apa yang seharusnya diukur.

Ada tiga tipe validitas, yaitu validitas isi, validitas konstruk dan validitas kriteria (Allen & Yen, 1979: 97). Selain itu, menurut (Azwar, 2000: 45-47) terdapat dua macam validitas isi, yaitu validitas kenampakan dan validitas logika. Validitas isi berarti sejauh mana suatu perangkat tes mencerminkan keseluruhan kemampuan yang hendak diukur (Azwar, 2000: 45), yang berupa analisis rasional terhadap domain yang hendak diukur. Validitas kenampakan didasarkan pada pertanyaan apakah suatu butir-butir dalam perangkat tes mengukur aspek yang relevan dengan domainnya. Validitas logika berkaitan dengan keseksamaan batasan pada domain yang hendak diukur, dan merupakan jawaban apakah

keseluruhan butir merupakan sampel representatif dari keseluruhan butir yang mungkin dibuat.

Validitas kriteria, disebut juga validitas prediktif, merupakan kesahihan suatu perangkat tes dalam membuat prediksi, dapat meramalkan keberhasilan siswa pada masa yang akan datang. Validitas prediktif suatu perangkat tes dapat diketahui dari korelasi antara perangkat tes dengan kriteria tertentu yang dikehendaki, yang disebut dengan variabel kriteria (Allen & Yen, 1979: 97; Azwar, 2000: 51).

2) Reliabilitas

Reynold (2006: 91) menyatakan bahwa reliabilitas mengacu pada kekonsistenan atau kestabilan hasil penilaian. Namun secara singkat Cohen (2007: 146) menyatakan bahwa reliabilitas sebagai kestabilan. Mengenai reliabilitas, Ebel & Frisbie (1991: 76), menyatakan bahwa jika tesnya memiliki konsistensi yang tinggi, maka tes tersebut akurat, *reproducible*, dan *generalizable* terhadap kesempatan testing dan instrumen yang sama.

Allen & Yen (1979: 62) menyatakan bahwa tes dikatakan reliabel jika skor amatan mempunyai korelasi yang tinggi dengan skor yang sebenarnya. Selanjutnya dinyatakan bahwa reliabilitas merupakan koefisien korelasi antara dua skor amatan yang diperoleh dari hasil pengukuran menggunakan tes yang paralel. Dengan demikian, pengertian yang dapat diperoleh dari pernyataan tersebut adalah suatu tes itu reliabel jika hasil pengukuran mendekati keadaan peserta tes yang sebenarnya.

Oleh karena itu, dalam bidang pendidikan, pengukuran tidak dapat langsung dilakukan pada ciri atau karakter yang akan diukur. Ciri atau karakter ini bersifat abstrak. Hal ini menyebabkan sulitnya memperoleh alat ukur yang stabil untuk mengukur karakteristik seseorang (Mehrens & Lehmann, 1973: 103).

Berdasarkan uraian di atas, maka dalam pembuatan alat ukur dalam dunia pendidikan harus dilakukan secermat mungkin dan disesuaikan dengan kaidah-kaidah yang telah ditentukan oleh ahli-ahli pengukuran di bidang pendidikan. Kereliabilisan suatu alat ukur, yang berupa suatu indeks reliabilitas, dapat dilakukan penelaahan secara statistik. Nilai ini biasa dinamakan dengan koefisien reliabilitas (*reliability coefficient*).

e. Jenis Instrumen Tes *Higher Order Thinking Skills* (HOTS)

Cukup banyak instrumen yang dapat digunakan sebagai pengukur hasil belajar siswa. Namun, jenis instrumen tes yang digunakan untuk mengukur HOTS siswa ialah pilihan ganda, jawaban singkat, atau uraian (Anderson & Krathwohl, 2012: 121-133). Selengkapnya dapat dilihat pada Tabel berikut ini.

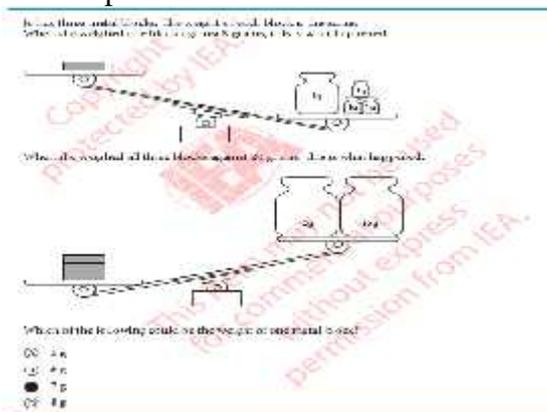
Tabel 2
Jenis Instrumen Tes Pengukur HOTS

HOTS	Indikator	Sub indikator	Jenis instrumen tes
Berpikir kritis	1. Analisis	1.1 Membedakan	• Jawaban singkat, pilihan ganda, dan uraian
		1.2 Mengorganisasikan	• Jawaban singkat, pilihan ganda, dan uraian
		1.3 Mengatribusikan	• Pilihan ganda dan uraian
	2. Evaluasi	2.1 Memeriksa	• Pilihan ganda dan uraian
		2.2 Mengkritik	• Pilihan ganda dan uraian

f. Membedakan Instrumen HOTS (*Higher Order Thinking Skills*) dan Instrumen LOTS (*Lower Order Thinking Skills*)

Instrumen yang memuat HOTS dapat diselesaikan oleh siswa yang memiliki kemampuan berpikir tingkat tinggi, yaitu berpikir kritis dan berpikir kreatif. Soal yang dibuat

oleh TIMSS dan PISA merupakan contoh soal/instrumen yang memuat HOTS, seperti soal pada gambar berikut ini yang termasuk soal nonrutin. Siswa tidak akan menemukan penyelesaiannya jika siswa tidak berpikir kritis



Gambar 1
Contoh Soal yang Memuat HOTS
 (Sumber: Mullis, 2011: 131)

Berbeda dengan soal yang tidak memuat HOTS atau LOTS. Berikut ini soal yang berjenis LOTS.

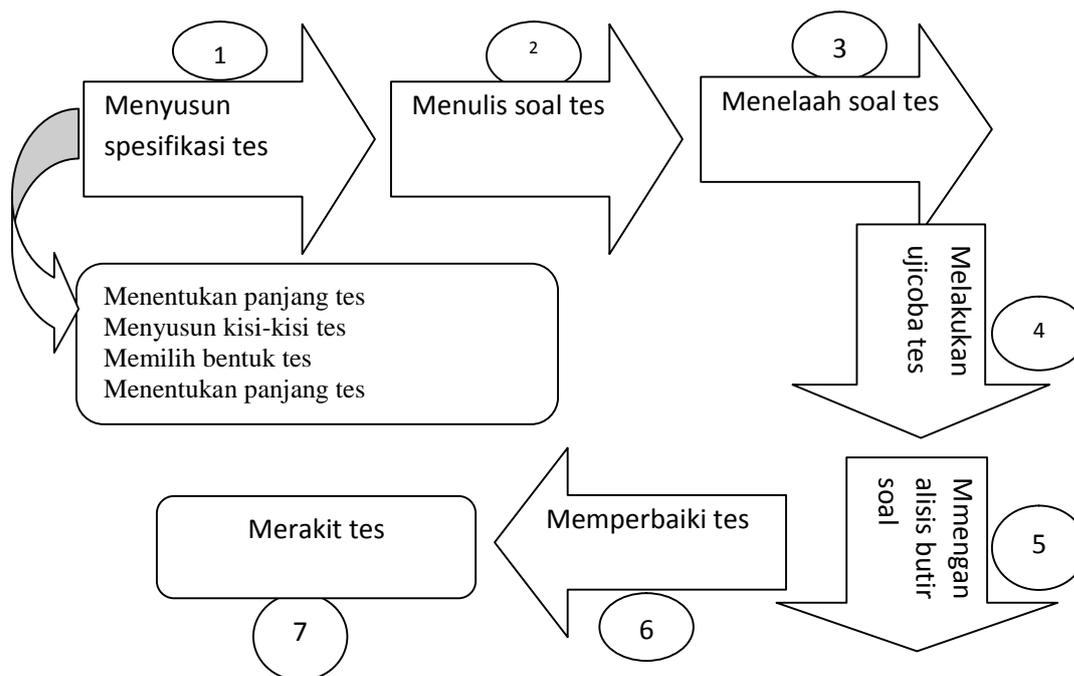
1 balok sama dengan 5 kg. Jika Pak Husen memiliki 5 balok, maka berapa berat semua balok tersebut?

Soal di atas sederhana sekali, karena semua siswa dapat menjawabnya. Siswa hanya melakukan proses perkalian saja, lalu diperoleh hasilnya. Contoh lain misalnya, dalam dua persamaan, siswa diminta untuk membedakan variabel dari persamaan tersebut dan mengaplikasikannya kepada sesuatu yang baru. Siswa yang hanya dapat membedakan saja, itulah yang disebut LOTS (*lower order thinking skills*), sedangkan siswa yang dapat membedakan dan mengaplikasikannya kepada sesuatu yang baru, itulah HOTS (*higher order thinking skills*) (Saveedra, 2011: 9).

METODE PENELITIAN

Jenis dan Sumber Data

Jenis penelitian ini termasuk *research and development* (R&D), yaitu peneliti mengembangkan instrumen pengukur *higher order thinking skills* (HOTS) matematika siswa. Penelitian ini akan menghasilkan sebuah produk berupa instrumen yang dapat dijadikan contoh atau acuan oleh guru untuk dapat mengukur HOTS siswa dalam pelajaran matematika. Model penelitian ini diadaptasi dari model pengembangan yang dibuat oleh Mardapi (2008: 88), yaitu:



Gambar 2
Model Pengembangan Instrumen Pengukur HOTS Matematika Siswa Kelas X

Bentuk tes atau jenis instrumen yang digunakan pada penelitian pengembangan ini adalah pilihan ganda, jawaban singkat, dan uraian. Bentuk tes ditentukan berdasarkan indikator yang dibuat oleh peneliti yang berdasarkan pada Tabel 3 mengenai jenis instrumen tes pengukur HOTS. Uji coba instrumen ini dilakukan di kelas X pada tiga sekolah, yaitu SMAN 4 Kota Yogyakarta, SMAN 6 Kota Yogyakarta, dan SMAN 10 Kota Yogyakarta.

Setelah disesuaikan, diperoleh instrumen sebanyak 46 butir soal yang akan diujicobakan. Butir soal sebanyak itu, peneliti bagi menjadi tiga paket soal, yaitu soal paket A, soal paket B, dan soal paket C. Alasan pembagian soal menjadi tiga paket karena tidak mungkin seorang siswa mengerjakan soal sebanyak 46 butir dalam waktu 2 jam pelajaran. Mengenai butir berapa saja yang terdapat pada masing-masing paket soal, dapat dilihat pada Tabel di bawah ini.

Tabel 3
Pembagian Soal Paket dan Waktu Pelaksanaan Uji Coba

	Kelas	Banyaknya Siswa	Soal Paket	Waktu	Jam ke-
SMAN 4	X.E	28	A	Kamis, 28 Mei 2015	3 - 4
	X.F	25	A	Sabtu, 30 Mei 2015	7 - 8
	X.G	22	C	Rabu, 27 Mei 2015	3 - 4
	X.H	19	C	Kamis, 28 Mei 2015	7 - 8
SMAN 6	X.8	22	B	Rabu, 27 Mei 2015	3 - 4
	X.9	19	C	Sabtu, 30 Mei 2015	3 - 4
SMAN 10	X.A	18	A	Rabu, 27 Mei 2015	5 - 6
	X.D	16	B	Kamis, 28 Mei 2015	1 - 2

Analisi Data

a. Analisis Faktor Eksploratori

Setelah peneliti selesai melaksanakan penelitian di lapangan, yakni dengan uji coba tiga paket soal terhadap subjek coba sebanyak delapan kelas di tiga sekolah, peneliti melakukan

kegiatan *scoring* atau penskoran. Hal ini dilakukan untuk membuktikan validitas konstruk, estimasi reliabilitas instrumen, menganalisis tingkat kesukaran, menganalisis daya pembeda, keefektifan distraktor, dan untuk mengetahui kemampuan HOTS matematika siswa.

Setelah kegiatan penskoran/*scoring* selesai, langkah selanjutnya adalah membuktikan validitas konstruk dengan menggunakan analisis faktor eksploratori. Hasil analisis tersebut dapat dilihat pada Tabel berikut:

Tabel 4
Hasil Analisis Faktor Eksploratori

No.	Paket Soal	Persentase Varians yang dapat Dijelaskan
1.	Soal Paket A	73,7%
2.	Soal Paket B	54,6%
3.	Soal Paket C	50,6%

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

a. Reliabilitas Instrumen

Tujuan penelitian pengembangan ini, selain ingin menghasilkan instrumen yang valid, juga ingin menghasilkan instrumen yang reliabel. Berikut ini peneliti sajikan hasil kereliabelan masing-masing paket soal dalam bentuk Tabel.

Tabel 5
Reliabilitas Instrumen

No.	Paket Soal	Nilai <i>Cronbach's Alpha</i>	Kesimpulan
1.	Soal Paket A	0,738	Reliabel
2.	Soal Paket B	0,658	Reliabel
3.	Soal Paket C	0,488	Tidak reliabel

Berdasarkan Tabel di atas, dapat dilihat bahwa nilai koefisien *Cronbach's Alpha* adalah 0,738 yang berarti soal paket A reliabel. Soal paket B diperoleh nilai koefisien *Cronbach's Alpha* adalah 0,658 yang berarti soal paket B juga reliabel. Soal paket C diperoleh nilai koefisien *Cronbach's Alpha* adalah 0,488 yang berarti soal paket C tidak reliabel.

b. Tingkat Kesukaran

Langkah dalam kegiatan analisis selanjutnya adalah menentukan tingkat kesukaran untuk masing-masing butir soal pada setiap paket soal. Hasil analisis tingkat kesukaran dan keputusannya, peneliti sajikan dalam Tabel berikut ini.

Tabel 6
Tingkat Kesukaran Butir Soal Setiap Paket

Butir Soal	Tingkat Kesukaran					
	Soal Paket A	Kategori	Soal Paket B	Kategori	Soal Paket C	Kategori
1	0,45	Sedang	0,82	Mudah	0,14	Sulit
2	0,21	Sulit	0,10	Sulit	0,30	Sedang
3	0,38	Sedang	0,16	Sulit	0,01	Sulit
4	0,17	Sulit	0,03	Sulit	0,52	Sedang
5	0,10	Sulit	0,16	Sulit	0,3	Sedang
6	0,23	Sulit	0,04	Sulit	0,45	Sedang
7	0,45	Sedang	0,21	Sulit	0,69	Sedang
8	0,40	Sedang	0	Sulit	0,10	Sulit

Butir Soal	Tingkat Kesukaran					
	Soal Paket A	Kategori	Soal Paket B	Kategori	Soal Paket C	Kategori
9	0,04	Sulit	0,22	Sulit	0,10	Sulit
10	0,94	Mudah	0,07	Sulit	0,25	Sulit
11	0,15	Sulit	0,41	Sedang	0,03	Sulit
12	0,56	Sedang	0,02	Sulit	0,35	Sedang
13	0,40	Sedang	0,02	Sulit	0,43	Sedang
14	0,51	Sedang	0,22	Sulit	0,20	Sulit
15	0,23	Sulit	0,74	Mudah	0,30	Sedang
16			0,68	Sedang		

Berdasarkan Tabel di atas, diinformasikan bahwa pada paket A butir soal yang berkategori mudah sebesar 6,67%, butir soal yang berkategori sedang sebesar 46,67%, dan butir soal yang berkategori sulit sebesar 46,67%. Pada paket B, butir soal yang berkategori mudah sebesar 12,50%, butir soal yang berkategori sedang sebesar 12,50%, dan butir soal yang berkategori sulit sebesar 75%. Pada paket C, tidak ada butir soal yang berkategori mudah, butir soal yang berkategori sedang sebesar 53,33% dan butir soal yang berkategori sulit sebesar 46,67%.

c. Daya Pembeda

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.

Tabel 7
Daya Pembeda Butir Soal Setiap Paket

Butir Soal	Daya Pembeda					
	Soal Paket A	Keputusan	Soal Paket B	Keputusan	Soal Paket C	Keputusan
1	0,23	Direvisi	0,21	Direvisi	0,10	Direvisi
2	0,30	Diterima	0,10	Direvisi	0,10	Direvisi
3	0,31	Diterima	0,30	Diterima	0,02	Ditolak
4	0,34	Diterima	0,07	Ditolak	0,10	Direvisi
5	0,19	Direvisi	0,30	Diterima	0,10	Direvisi
6	-0,10	Ditolak	0,08	Ditolak	0,47	Diterima
7	0,46	Diterima	0,42	Diterima	0,10	Direvisi
8	0,59	Diterima	0	Ditolak	0,20	Direvisi
9	0,09	Ditolak	0,34	Diterima	-0,03	Ditolak
10	0,10	Direvisi	-0,01	Ditolak	0,2	Direvisi
11	0,24	Direvisi	0,38	Diterima	0	Ditolak
12	0,45	Diterima	0,05	Ditolak	0,23	Direvisi
13	0,42	Diterima	0,05	Ditolak	0,24	Direvisi
14	0,39	Diterima	0,3	Diterima	0,31	Diterima
15	0,02	Ditolak	-0,1	Ditolak	0,4	Diterima
16			0,21	Direvisi		

Pada soal paket A, butir soal yang diterima sebesar 53,33%, butir soal yang perlu direvisi sebesar 26,67%, dan butir soal yang ditolak sebesar 20%. Pada soal paket 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 soal paket C, butir soal yang diterima sebesar 20%, butir soal yang perlu direvisi sebesar 60%, dan butir soal yang ditolak sebesar 20%. Untuk melihat butir berapa saja yang diterima, direvisi, dan ditolak.

d. Analisis Pengecoh/Distraktor

Analisis pengecoh/distraktor ini dilakukan pada instrumen tes yang berupa pilihan ganda. Dari tiga paket soal yang dibuat, masing-masing terdapat tiga butir soal pilihan ganda. Kesimpulan secara umum diperoleh bahwa masih ada pengecoh yang belum baik. Artinya, pengecoh tersebut dipilih oleh siswa kurang dari 5% dari seluruh siswa, dan masih banyak dipilih oleh kelompok atas. Secara rincinya dijelaskan sebagai berikut.

1) Soal Paket A

Tabel 8
Pengecoh/Distraktor pada Soal Paket A

Option	Persentase terpilihnya Soal Butir (%)		
	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

Berdasarkan Tabel di atas, 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.

2) Soal Paket B

Tabel 9
Pengecoh/Distraktor pada Soal Paket B

Option	Persentase terpilihnya Soal Butir (%)		
	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

Berdasarkan Tabel di atas, 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.

3) Soal Paket C

Tabel 10
Pengecoh/Distraktor pada Soal Paket C

Option	Persentase terpilihnya Soal Butir (%)		
	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

Berdasarkan Tabel di atas, 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.

HASIL PENELITIAN DAN PEMBAHASAN

Pembahasan Hasil Penelitian

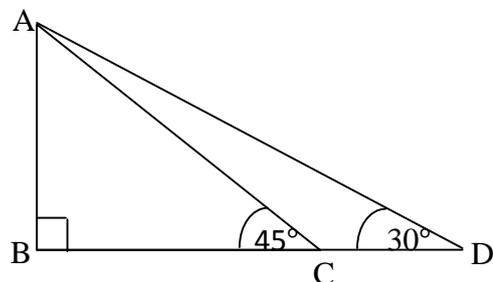
a. Perbaikan Instrumen

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. Berikut ini adalah salah satu contoh soal yang menjadi instrumen penelitian yang mengalami perbaikan.

Sebelum direvisi;

Amati gambar berikut ini.

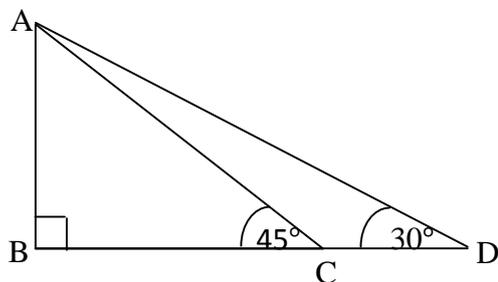


Pak Guru : “Perhatikan Adi. Jika $CD = 3$ cm, bagaimana cara mencari panjang AB ?”
 Adi : “Dengan perbandingan dua sudut tangen dapat dicari panjang AB Pak.”
 Raihan : “Untuk mencari panjang AB perlu membandingkan dua sudut sin.”

Dapatkan kamu menentukan panjang AB tersebut dengan menggunakan perbandingan dua sudut tangen. Jelaskan.

Setelah direvisi:

Amati gambar berikut ini.



Lalu cermati dialog berikut:

Pak Guru : “Perhatikan Adi. Jika $CD = 3$ cm, bagaimana cara mencari panjang AB?”
 Adi : “Dengan perbandingan dua sudut tangen dapat dicari panjang AB Pak.”
 Raihan : “Untuk mencari panjang AB perlu membandingkan dua sudut sin.”

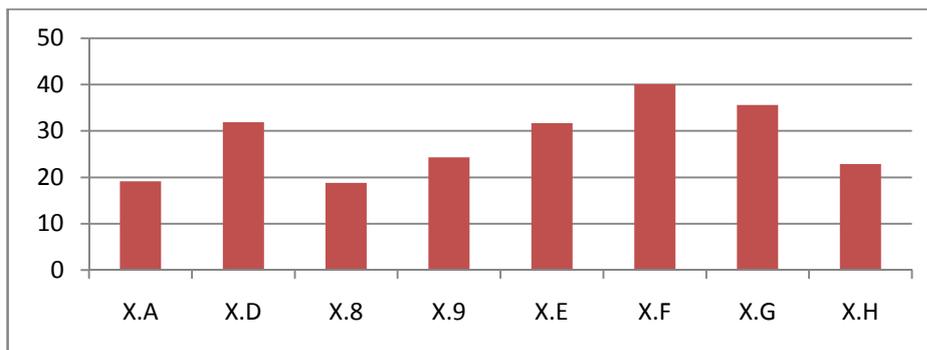
Berdasarkan dialog di atas, maka panjang AB adalah ... cm.

- | | |
|---|---|
| A. $\frac{\sqrt{4}}{\left(1-\frac{\sqrt{4}}{3}\right)}$ | D. $\frac{\sqrt{2}}{\left(1-\frac{\sqrt{2}}{2}\right)}$ |
| B. $\frac{\sqrt{2}}{\left(1-\frac{\sqrt{4}}{3}\right)}$ | E. $\frac{\sqrt{3}}{\left(1-\frac{\sqrt{3}}{3}\right)}$ |
| C. $\frac{\sqrt{3}}{\left(1-\frac{\sqrt{3}}{2}\right)}$ | |

b. Kemampuan *Higher Order Thinking* (HOT) Matematika Siswa

Tujuan analisis ini adalah untuk mengetahui kemampuan *higher order thinking* (HOT) matematika siswa. Jika siswa memperoleh nilai 65 maka *higher order thinking* (HOT) matematika siswa tersebut baik, sedangkan jika siswa memperoleh nilai < 65 maka HOTS matematika siswa kurang baik.

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 gambar di bawah ini.



Gambar 2
Nilai HOTS Siswa pada Masing-masing Kelas

Berdasarkan gambar di atas juga, diinformasikan bahwa siswa SMAN 4 Kota Yogyakarta memiliki kemampuan *higher order thinking* (HOT) yang lebih baik dibandingkan SMAN 10 dan SMAN 6, dan SMAN 10 memiliki kemampuan *higher order thinking* (HOT) yang lebih baik SMAN 6.

Pada soal paket 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.

Tabel 11
Perbaikan Instrumen pada Soal Paket A

No.	Nomor Butir	Kategori 2	Kategori 3	Alasan
1.	1	√		DB rendah
2.	2	√		TK mudah
3.	3	√		Pengecoh kurang baik
4.	4	√		TK mudah
5.	5		√	DB sangat rendah
6.	6	√		TK tinggi dan DB rendah
7.	9		√	TK sangat mudah dan DB sangat rendah
8.	10	√		TK tinggi dan DB rendah
9.	11	√		TK mudah dan DB rendah
10.	12	√		Pengecoh kurang baik
11.	14	√		Pengecoh kurang baik
12.	15		√	TK mudah dan DB sangat rendah

Keterangan:

TK : Tingkat Kesukaran

DP : Daya Pembeda

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. Lebih lengkapnya, peneliti sajikan pada Tabel berikut.

Tabel 12
Perbaikan Instrumen pada Soal Paket B

No.	Nomor Butir	Kategori 2	Kategori 3	Alasan
1.	1	√		TK tinggi dan DB rendah
2.	2	√		TK mudah dan DB rendah
3.	3	√		TK mudah
4.	4		√	TK sangat mudah dan DB sangat rendah
5.	5	√		TK mudah
6.	6		√	TK sangat mudah dan DB sangat rendah
7.	7	√		TK mudah
8.	8		√	TK sangat mudah dan DB sangat rendah
9.	9	√		TK mudah
10.	10		√	TK sangat mudah dan DB sangat rendah
11.	12		√	TK sangat mudah dan DB sangat rendah
12.	13		√	TK sangat mudah dan DB sangat rendah
13.	14	√		TK mudah
14.	15		√	TK tinggi dan DB sangat rendah
15.	16	√		DB rendah

Keterangan:

TK : Tingkat Kesukaran

DP : Daya Pembeda

Pada soal paket 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. Lebih detailnya dapat dilihat pada Tabel berikut.

Tabel 13
Perbaikan Instrumen pada Soal Paket C

No.	Nomor Butir	Kategori 2	Kategori 3	Alasan
1.	1	√		TK mudah dan DB rendah
2.	2	√		DB rendah
3.	3		√	TK sangat mudah dan DB sangat rendah
4.	4	√		DB rendah
5.	5	√		DB rendah
7.	7	√		DB rendah
8.	8	√		TK mudah dan DB rendah
9.	9		√	TK mudah dan DB sangat rendah
10.	10	√		TK mudah dan DB rendah
11.	11		√	TK sangat mudah dan DB sangat rendah
12.	12	√		DB rendah
13.	14	√		TK mudah

Keterangan:

TK : Tingkat Kesukaran

DP : Daya Pembeda

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil analisis yang telah peneliti lakukan, dapat diambil kesimpulan sebagai berikut:

1. Berdasarkan perhitungan menggunakan rumus Aiken, instrumen pengukur *higher order tinkering 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. Soal paket A berjumlah 15 butir, soal paket B berjumlah 16 butir, dan soal paket C berjumlah 15 butir. Soal paket A dan paket B menghasilkan soal paket yang reliabel, dengan masing-masing nilai *Cronbach's Alpha* sebesar 0,738 dan 0,658, sedangkan soal paket C tidak reliabel dengan nilai *Cronbach's Alpha* sebesar 0,488.
2. Hasil dari uji coba instrumen dapat disimpulkan bahwa kemampuan *higher order thinking* (HOT) 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.

Saran

Penelitian pengembangan ini menghasilkan sebuah produk instrumen pengukur kemampuan *higher order tinkering* (HOT) matematika siswa kelas X. Oleh karena itu, bagi guru matematika kelas X dapat menggunakannya seperti pada Ujian Tengah Semester maupun Ujian Kenaikan Kelas. Selain itu, peneliti lain juga dapat membuat dan mengembangkan instrumen pengukur HOTS matematika siswa kelas XI dan kelas XII, atau kelas VII, VIII, IX SMP.

Diseminasi dan Pengembangan Produk Lebih Lanjut

1. Hasil penelitian ini akan disebarakan ke sekolah-sekolah yang dijadikan tempat uji coba penelitian, yaitu SMAN 4 Kota Yogyakarta, SMAN 6 Kota Yogyakarta, dan SMAN 10 Kota Yogyakarta. Nantinya, instrumen ini diharapkan bisa digunakan atau dijadikan contoh oleh guru matematika kelas X.
2. Hasil penelitian akan diikuti pada kegiatan seminar dengan tema yang berkaitan.
3. Produk yang berupa instrumen pengukur HOTS matematika siswa kelas X ini, dapat dijadikan contoh atau acuan oleh guru matematika kelas XI dan XII dengan membuat instrumen yang materinya ada pada kelas XI atau XII.
4. Peneliti berharap akan ada yang mengembangkan produk ini lebih lanjut agar menghasilkan instrumen yang lebih baik.

DAFTAR PUSTAKA

- Anderson, L. W. & Krathwohl, D. R. (2010). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives. A bridge Edition*. (Terjemahan Agung Prihantoro). New York: Addison Wesley Longman, Inc. (Buku asli diterbitkan tahun 2001).
- Allen, M.J., & Yen, W.M. (1979). *Introduction to measurement theory*. Monterey, Mexico: Brooks/Cole Publishing Company.
- Azwar, S. (2011). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Belajar.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. New York: Routledge.

- 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.
- Depdikbud. (2008). *Kamus besar bahasa Indonesia*. Jakarta: Pusat Bahasa.
- Ebel, R.L. & Frisbie, D.A. (1991). *Essentials of educational measurement. Fifth edition*. New Delhi: Prentice Hall of India
- Gronlund, N.E., Linn, R.L., & Miller, M.D. (2009). *Measurement & evaluation in teaching. Tenth edition*. New York: Macmillan Publishing Co., Inc.
- James, G. & James, R.C. (1992). *Mathematics dictionary. Fifth edition*. New York: Van Nostrand Reinhold Company.
- 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.
- Mehrens, W.A. & Lehmann, I.J. (1973). *Measurement and evaluation in education and psychology*. New York : Hold, Rinehart and Wiston, Inc.
- 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.
- Reynolds, C.R, Livingston, R.B., & Willson, V. (2010). *Measurement and assessment in education. Second edition*. Upper Saddle River, New Jersey: Pearson Education.
- Saveedra, A. R. (2011). *Teaching and learning 21st century skills: Lessons from the learning sciences*. Asia Society. Partner for Global Learning. RAND Corporation.
- Setyosari, P. (2013). *Metode penelitian pendidikan & pengembangan*. Jakarta: Prenadamedia group.
- Thompson, T. (2012). Mathematics teachers' interpretation of higher-order thinking in bloom's taxonomy. *International Electronic Journal of Mathematics Education, Volume 3, Number 2, July 2008*.