

## PENGEMBANGAN INSTRUMEN TES FISIKA BERBASIS *HOTS* TINGKAT SMA DI KOTA MEDAN

Nurhikmah Weisdiyanti<sup>1</sup>, Rita Juliani<sup>2</sup>

<sup>1,2</sup>Jurusan Fisika FMIPA Universitas Negeri Medan

Email: nurhikmahweisdiyanti@gmail.com

### ABSTRAK

Telah dilakukan penelitian pengembangan instrumen tes berbasis *Higher Order Thinking Skills* yang bertujuan mengembangkan instrumen yang memenuhi aspek validitas, reliabilitas, daya pembeda, tingkat kesukaran dan efektivitas pengecoh. Jenis penelitian yang digunakan *Research and Development* model Borg & Gall dengan tahapan analisis masalah dan kebutuhan instrumen tes *HOTS*, perencanaan pengembangan dan pengujian instrumen, pengembangan instrumen tes berbasis *HOTS*, uji validasi ahli materi, konstruksi dan bahasa, revisi validasi ahli, uji skala kecil di SMAN 11 Medan, revisi instrumen skala kecil, uji skala besar di SMAN 11 Medan dan SMAS Budi Satrya, serta revisi instrumen skala besar. Hasil uji validasi ahli diperoleh instrumen tes valid. Hasil uji skala kecil diperoleh 93% item valid, sangat reliabel, 87% daya beda baik, 93% kesukaran sedang serta 67% distraktor baik. Hasil uji skala besar diperoleh 87% valid, reliabel, 73% daya beda baik, 67% kesukaran sedang, dan 93% distraktor baik sehingga instrumen tes layak mengukur keterampilan berpikir tingkat tinggi.

**Kata kunci:** Keterampilan berpikir tingkat tinggi, instrumen tes, penelitian dan pengembangan

### ABSTRACT

Research on the development of test instruments based on Higher Order Thinking Skills has been carried out which aims to develop instruments that meet the aspects of validity, reliability, differentiation, difficulty level and deceptive effectiveness. The type of research used is the Borg & Gall Research and Development model with the stages of analyzing the problems and needs of the *HOTS* test instrumen, planning development and testing of instruments, developing *HOTS*-based test instruments, validating material experts, construction and language, revision of expert validation, small-scale testing SMAN 11 Medan, small scale instrumen revision, large scale test at SMAN 11 Medan and SMAS Budi Satrya, and large scale instrumen revision. Expert validation test results obtained valid test instruments. Small scale test results obtained 93% valid items, very reliable, 87% good difference power, 93% moderate difficulty and 67% good distractors. Large scale test results obtained 87% valid, reliable, 73% good difference power, 97% moderate difficulty, and 93% good distractors so that the test instrumen is feasible to measure higher order thinking skills.

**Keywords:** *Higher Order Thinking Skills*, test instrument, research and development

### PENDAHULUAN

Salah satu tujuan mata pelajaran Fisika di SMA agar peserta didik mampu mengembangkan kemampuan bernalar dalam berpikir analisis induktif dan deduktif menggunakan konsep dan prinsip Fisika untuk menjelaskan berbagai peristiwa dan menyelesaikan masalah secara kualitatif maupun kuantitatif (BSNP, 2006). Alat ukur yang digunakan untuk penilaian penalaran tingkat tinggi pada standar internasional diantaranya *Programme for International Student Assessment (PISA)*, *Trends in International Mathematics and Science Study (TIMSS)*, dan *Progress in International Reading Literacy Study (PIRLS)* yang merupakan instrumen tes berbasis *Higher Order Thinking Skills (HOTS)*.

Hasil penelitian Simalango (2018) bahwa peserta tes merasa kesulitan dalam menyelesaikan instrumen tes level 3 (penalaran) dan menginterpretasikannya dalam situasi nyata. Rendahnya capaian peserta tes dalam *PISA* (OECD, 2019) dan *TIMSS* karena pembelajaran

di Indonesia umumnya belum berbasis *HOTS*. Perlu adanya pembiasaan dalam menyelesaikan instrumen tes berbasis *HOTS* sehingga peserta didik terbiasa berpikir kritis, kreatif dan memecahkan permasalahan yang berkaitan dengan kehidupan sehari-hari (Saidah, dkk., 2019).

Pembiasaan menyelesaikan tes fisika berbasis *HOTS* telah dilakukan pada SMA di SMAN 11 Medan sejak tahun 2018 dengan memberikan instrumen tes harian, ulangan dan ujian yang bersifat *HOTS*. Hasil wawancara terhadap guru di SMAN 11 Medan bahwa guru membutuhkan instrumen tes berbasis *HOTS*. Guru bidang studi fisika mengaku kesulitan mencari contoh instrumen tes Fisika berbasis *HOTS* meskipun sudah memiliki modul penyusunan instrumen tes *HOTS*. Kurang tersedianya instrumen tes *HOTS* Fisika di SMAN 11 Medan penting dilakukan pengembangan instrumen tes Fisika berbasis *HOTS*.

Bentuk instrumen tes yang diujikan pada tes latihan, ulangan, ujian tengah semester, dan ujian akhir semester di SMAN 11 Medan merupakan instrumen tes fisika pilihan ganda, isian singkat dan esai. Instrumen tes disusun guru secara mandiri. Hasil analisis tingkat kognitif bahwa instrumen tes yang disajikan guru berada pada tingkat kognitif C1 sampai C5. Instrumen tes cenderung pada tingkat C1, C2, C3. Instrumen tes ujian tengah semester yang ditunjukkan guru sebagai instrumen tes *HOTS* masih berada ditingkat kognitif C3. Instrumen tes dikategorikan *HOTS* apabila mengukur berpikir tingkat kognitif C4, C5, C6 (L3) peserta tes dan membangkitkan keterampilan berpikir kritis, berpikir kreatif, pemecahan masalah, dan membuat keputusan.

Hasil observasi bersama peserta tes kelas XI IPA 1 bahwa 73% responden menyatakan instrumen tes yang diberikan guru hanya memasukkan rumus yang ada, 33% peserta tes menyatakan jarang menjumpai instrumen tes yang membangun kemampuan analisis, mengembangkan ide peserta tes dan tidak ada instrumen tes yang memadukan dua materi sekaligus. Hasil analisis wawancara dan observasi didapat bahwa guru belum memahami instrumen tes berbasis *HOTS*, sehingga instrumen yang disusun belum tepat sebagai instrumen tes berbasis *HOTS*. Kurangnya ketersediaan dan belum tepatnya instrumen tes *HOTS* Fisika yang berada di SMAN 11 Medan berdampak pada peserta didik yang tidak terbiasa menghadapi instrumen tes *HOTS* sehingga kemampuan menyelesaikan instrumen tes *HOTS* peserta didik rendah. Rendahnya kemampuan menyelesaikan instrumen tes *HOTS* peserta tes SMAN 11 Medan terbukti pada rata-rata hasil Ujian Nasional Fisika tahun 2019 adalah 45,89 (Dokumen UN SMAN 11 Medan, 2019). Permasalahan dapat diatasi apabila ada contoh instrumen tes berbasis *HOTS* yang tepat, sehingga guru bisa menerapkan dan mengembangkan instrumen tes dari contoh yang diberikan.

Sani, dkk., (2019) menyatakan bahwa *HOTS* berbeda dengan berpikir tingkat tinggi (*higher order thinking/HOT*). *HOT* mengacu pada kemampuan kognitif dalam menganalisis (C4), mengevaluasi (C5) dan mengkreasi (C6), sedangkan *HOTS* berkaitan dengan kemampuan menyelesaikan masalah, berpikir kritis, berpikir kreatif dan membuat keputusan. Sugrue (1995) dalam Sani (2019) mengidentifikasi tiga format yang digunakan untuk mengukur *HOTS*, yaitu: memilih jawaban seperti instrumen tes pilihan ganda dan instrumen tes menjodohkan,

membangkitkan (instrumen tes dengan jawaban singkat, esai dan unjuk kerja, dan menjelaskan).

Penelitian “Pengembangan Instrumen Tes Fisika Berbasis *Higher Order Thinking Skills (HOTS)* Tingkat SMA di Kota Medan” penting dilakukan untuk meningkatkan keterampilan berpikir tingkat tinggi, dan membangun kemandirian peserta didik untuk menyelesaikan masalah. Instrumen tes dikembangkan menggunakan metode *research and development (R&D)* model Borg & Gall. Hasil penelitian Afriani, dkk (2019) mengenai pengembangan instrumen tes fisika berbasis *HOTS* menggunakan model Borg & Gall diperoleh bahwa instrumen tes layak digunakan sebagai alat ukur kemampuan berpikir tingkat tinggi peserta tes dengan validitas, reabilitas tinggi, tingkat kesukaran sedang dan tingkat keterbacaan tes sesuai dengan jenjang yang diukur. Hasil penelitian serupa dilakukan oleh Kistiono (2019) bahwa instrumen tes yang dikembangkan valid, mempunyai reliabilitas yang baik. Penelitian bertujuan mengembangkan instrumen tes fisika Berbasis *HOTS* yang memenuhi kualifikasi aspek validitas, reliabilitas, daya pembeda, tingkat kesukaran dan efektivitas pengecoh.

## METODE PENELITIAN

### Waktu dan Tempat Penelitian

Penelitian dilaksanakan di SMAN 11 Medan yang berlokasi di Jalan Pertiwi No. 93 Medan dan di SMAS Budi Satrya Medan di Jl. Letda Sudjono No 166 Bandar Selamat, Medan. Penelitian dilaksanakan pada tanggal 09 Maret – 06 Juni 2020 semester genap tahun ajaran 2019/2020.

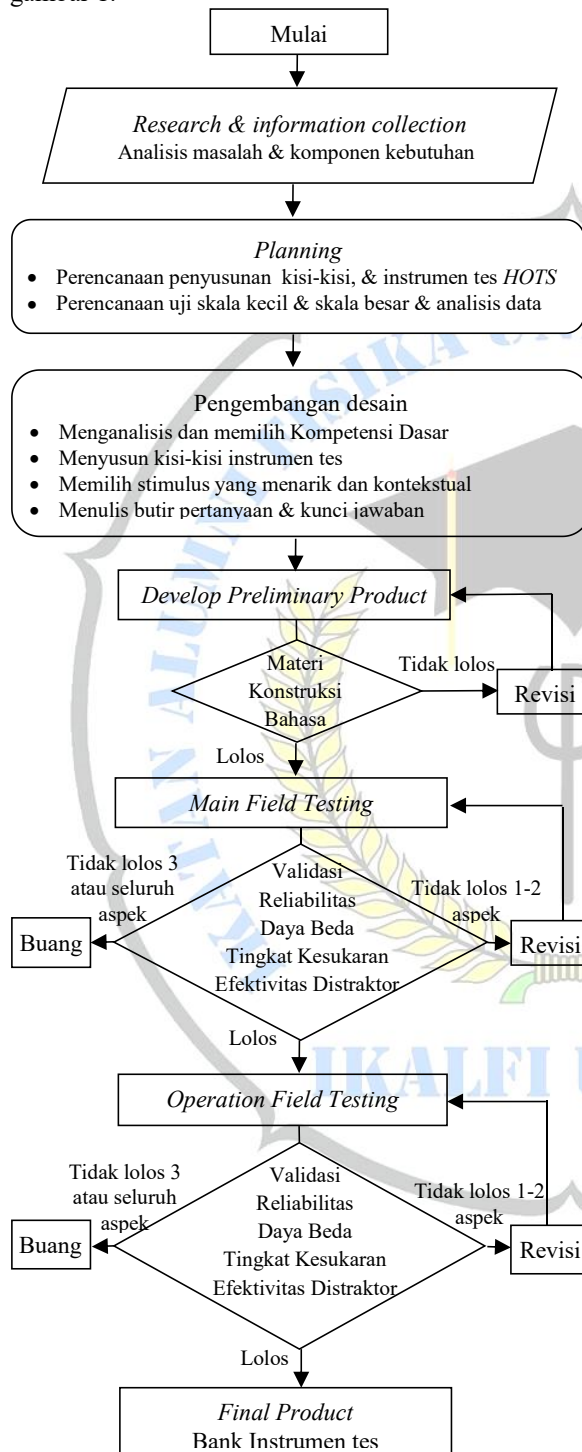
### Populasi dan Sampel

Populasi dalam penelitian adalah seluruh peserta tes di SMAN 11 Medan. Sampel dalam penelitian ada dua kelas yaitu 10 orang dari Kelas XI IPA 5 sebagai sampel kelas terbatas, dan 60 orang dari kelas XI IPA 5 SMAN 11 Medan dan kelas XI IPA 1 SMAS Budi Satrya untuk sampel kelas lebih luas. Variabel bebas dalam penelitian adalah instrumen tes objektif berbasis *HOTS*. Variabel terikat dalam penelitian berupa validitas, reliabilitas, tingkat kesukaran, daya pembeda dan efektivitas distraktor.

### Jenis Penelitian

Jenis penelitian yang digunakan *Research and Development (R&D)* model riset Borg & Gall. Penelitian menggunakan pendekatan *mixed method* yakni mengkombinasi bentuk kualitatif & kuantitatif (Creswell, 2013). Produk yang dikembangkan berupa instrumen tes fisika berbasis *HOTS* berdasarkan panduan

penulisan instrumen tes HOTS Fisika oleh Kementerian Pendidikan dan Kebudayaan (Widana, 2017) dan jurnal Scully (2017) terkait cara mengkonstruksi item tes HOTS bentuk pilihan berganda. Desain penelitian tertera pada gambar 1.



Gambar 1. Desain Penelitian pengembangan (R&D) model Borg & Gall (1989)

### Teknik Pengumpulan Data

Teknik pengumpulan data penelitian yaitu wawancara, dokumentasi dan studi pustaka. Peneliti mengumpulkan informasi awal melalui wawancara untuk mengetahui permasalahan dan kebutuhan peserta tes, guru maupun pihak sekolah. Metode dokumentasi digunakan untuk mendapatkan seperangkat instrumen tes, silabus, kisi-kisi instrumen tes, hasil Ujian Nasional 2020, serta respon jawaban peserta tes. Peneliti mengumpulkan informasi melalui kajian pustaka buku dan jurnal untuk memecah masalah dan acuan dalam mengembangkan dan melaksanakan penelitian.

### Teknik Analisis Data

Data penelitian dianalisis secara kualitatif dan kuantitatif.

#### 1) Analisis Kualitatif

Analisis kualitatif dilakukan menggunakan validasi isi yang ditentukan kesepakatan ahli materi, konstruksi tes dan bahasa. Para ahli mempertimbangkan tingkat kematangan berpikir peserta tes dengan menilai instrumen tes menggunakan skor skala likert yang terdiri atas 4 tingkatan, yaitu: 1, 2, 3 dan 4 dengan ketentuan 1 = tidak valid, 2 = kurang valid, 3 = valid, 4 = sangat valid. Indeks validitas isi item tes dihitung menggunakan rumus validitas menurut Aiken ditunjukkan pada persamaan 1.

$$V = \frac{\sum s}{n(c-1)} \quad (1)$$

Dengan  $s = r - l_0$  (2)

Keterangan:

$V$  = indeks kesepakatan rater mengenai validitas butir

$s$  = skor yang ditetapkan ahli dikurangi skor terendah dalam kategori penskoran

$r$  = skor kategori pilihan rater

$l_0$  = skor terendah dalam kategori penskoran

$n$  = banyaknya ahli (rater)

$c$  = banyaknya kategori yang dapat dipilih rater (Retnawati, 2016)

Tolak ukur untuk menginterpretasikan validitas isi dilakukan dengan membandingkan indeks validitas butir instrumen tes dengan nilai tabel kriteria

Tabel 3.2. Kriteria Validitas instrumen tes

Indeks Validitas	Kriteria Validitas
$0,8 \leq V \leq 1,0$	Validitas tinggi
$0,4 \leq V \leq 0,8$	Validitas sedang
$0,4 < V$	Validitas rendah

Sumber: Retnawati, 2016

#### 2) Analisis Kuantitatif

Karakteristik item tes dianalisis secara kuantitatif pada aspek validitas, reliabilitas, taraf kesukaran, daya pembeda, dan efektivitas

distraktor item tes dijelaskan secara rinci sebagai berikut.

a) *Validasi Konstruksi Item tes*

Setiap item tes dihitung validitas butir tes dengan membuat tabel skor dengan skor tes bentuk objektif untuk item biasa. Validitas hitung diperoleh menggunakan rumus korelasi *product moment* angka kasar yang ditunjukkan pada persamaan 3.

$$r_{XY} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} \quad (3)$$

Keterangan:

$r_{XY}$  = Koefisien korelasi variabel X dan Y

X = Skor item tes

Y = Skor total item tes (Widiyanto, 2018)

Tolak ukur untuk menginterpretasikan validitas instrumen dilakukan dengan membandingkan nilai koefisien korelasi hitung dengan koefisien korelasi tabel. Item tes dikatakan valid apabila  $r_{hitung} > r_{tabel}$ . Item tes yang tidak valid perlu revisi dengan meningkatkan penguasaan teknis pembuatan instrumen tes tes (Marthunis, dkk., 2015); (Windarto & Martubi, 2017).

b) *Reliabilitas*

Instrumen penilaian *HOTS* yang dikembangkan berupa instrumen tes bentuk pilihan ganda, sehingga untuk menghitung reliabilitas instrumen digunakan rumus *Kuder Richardson 20* (KR-20) dengan persamaan 4.

$$r_{11} = \frac{n}{n-1} \left[ \frac{s^2 - \sum pq}{s^2} \right] \quad (4)$$

Keterangan:

$r_{11}$  = reliabilitas tes secara keseluruhan

p = jumlah subjek yang menjawab benar

q = jumlah subjek yang menjawab salah  
 ( $q = 1 - p$ )

$\sum pq$  = jumlah hasil perkalian antara p dan q

n = banyaknya item

s = standar deviasi tes (Arikunto, 2008)

Tolak ukur untuk menginterpretasikan nilai korelasi ( $r_{11}$ ) yang diperoleh dengan membandingkan koefisien reliabilitas dengan nilai tabel kriteria reliabilitas instrumen tes (Kadir, 2015) tertera pada tabel 1.

Tabel 1. *Kriteria Reliabilitas Item Tes*

Koefisien Reliabilitas	Kriteria Reliabilitas
$0,91 \leq r_{11} \leq 1,00$	Sangat tinggi
$0,71 \leq r_{11} \leq 0,90$	Tinggi
$0,41 \leq r_{11} \leq 0,70$	Cukup
$0,21 \leq r_{11} \leq 0,40$	Rendah
$0,00 \leq r_{11} \leq 0,20$	Sangat rendah

Sumber: Direktorat Pembinaan SMA, 2010

c) *Daya Beda*

Daya pembeda item tes dianalisis menggunakan persamaan 3 (Widiyanto, 2018).

$$D = \frac{BA}{J_A} - \frac{BB}{J_B} \quad (5)$$

Keterangan:

D = Indeks daya pembeda

$B_A$  = Jumlah peserta tes yang menjawab benar pada kelompok atas

$B_B$  = Jumlah peserta tes yang menjawab benar pada kelompok bawah

$J_A$  = Jumlah peserta tes kelompok atas

$J_B$  = Jumlah peserta tes kelompok bawah

Tolak ukur untuk menginterpretasikan indeks daya pembeda dengan membandingkan koefisien korelasi daya pembeda hitung dengan nilai tabel kriteria indeks daya pembeda yang tertera pada tabel 2.

Tabel 2. *Kriteria Indeks Daya Pembeda*

Koefisien Daya Pembeda	Kriteria Daya Pembeda
$0,71 \leq DP \leq 1,00$	Sangat baik
$0,41 \leq DP \leq 0,70$	Baik
$0,21 \leq DP \leq 0,40$	Cukup
$0,00 \leq DP \leq 0,20$	Tidak baik
$0 < DP$	Sangat tidak baik

Sumber: Widiyanto, 2018

Item tes dengan daya beda cukup diterima dengan revisi, item tes dengan daya pembeda rendah memiliki dua kemungkinan tindak lanjut, yaitu direvisi atau dibuang, dan item tes dengan daya pembeda negatif perlu dibuang (Windarto & Martubi, 2017).

d) *Tingkat Kesukaran*

Tingkat kesukaran diberi simbol "P" dengan singkatan dari proporsi. Analisis tingkat kesukaran item tes menggunakan persamaan 5.

$$P = \frac{B}{JS} \quad (5)$$

Dimana:

P = Indeks kesukaran

B = Jumlah peserta tes menjawab benar

JS = Jumlah seluruh peserta tes

(Arikunto, 2008)

Indeks kesukaran diklasifikasikan dengan tabel indikator tingkat kesukaran yang tertera pada tabel 3.

Tabel 3. *Indikator Tingkat Kesukaran Item tes*

Koefisien Korelasi	Indikator Tingkat Kesukaran
$0,71 \leq P \leq 1,00$	Sukar
$0,31 \leq P \leq 0,70$	Sedang
$0,00 \leq P \leq 0,30$	Mudah

Sumber: Arikunto, 2008

Instrumen tes yang baik adalah instrumen tes dengan kategori sedang, namun instrumen tes yang terlalu mudah dan sukar tidak berarti

tidak boleh digunakan, karena tergantung kegunaan tiap instrumen tes (Widiyanto, 2018). Proses berpikir tingkat tinggi peserta didik dapat menunjukkan pemahaman informasi dan bernalar (L3) sehingga instrumen tes dapat diterima tanpa revisi pada tingkat kesukaran sedang dan sukar.

e) *Efektivitas Distraktor*

Pengecoh dianggap baik apabila jumlah peserta didik yang memilih pengecoh sama atau mendekati jumlah ideal. Efektivitas distraktor dihitung menggunakan persamaan:

$$IP = \frac{P}{N} \times 100\% \quad (5)$$

Dimana:

*IP* = Indeks efektivitas pengecoh

*P* = Jumlah peserta tes yang memilih pengecoh

*N* = Jumlah seluruh peserta tes

Distraktor berfungsi dengan baik jika sudah dipilih lebih 5% pengikut tes. Distraktor tidak berfungsi dengan baik jika kurang dari 5% (Sudijono, 2008).

f) *Pemilihan Kualitas Item tes*

Kriteria yang digunakan untuk menentukan kualitas item tes diadaptasi dari skala *Likert* (Sugiyono, 2010) tertera pada tabel 4.

Tabel 4. Kriteria Kualitas Item tes

Kriteria yang Terpenuhi	Kualitas Item tes	Keterangan Revisi	Masuk Bank Instrumen tes
4	Sangat Baik	Tidak ada	Ya
3	Baik	Minor	Belum
2	Sedang	Mayor	Belum
1	Jelek	Buang	Tidak
0	Sangat Jelek	Buang	Tidak

**HASIL DAN PEMBAHASAN**

**Deskripsi Instrumen Tes Fisika Berbasis HOTS**

Produk yang dikembangkan berupa instrumen tes Fisika berbasis *HOTS* dari 10 item tes pilihan berganda UN Fisika tahun 2016-2019 menjadi 15 item tes dengan bentuk pilihan berganda pada materi usaha dan energi. Scully (2017) menyatakan bahwa instrumen tes bentuk pilihan berganda memiliki kapasitas untuk menilai keterampilan tingkat tinggi.

Instrumen tes dirancang menggunakan stimulus yang menampilkan konsep, visualisasi, analogi, dan kesimpulan (Schraw & Robinson, 2011) sehingga bisa membangkitkan keterampilan berpikir kritis, keterampilan memecahkan masalah. Instrumen tes dikemas

kedalam bank instrumen tes dengan 10 item tes berpikir kritis dan 5 item tes memecahkan masalah. Persebaran instrumen tes *HOTS* dan tingkat kognitif item tes yang telah dikembangkan tertera pada tabel 5.

Tabel 5. Kategori dan Tingkat Kognitif Instrumen tes berbasis *HOTS* yang telah dikembangkan

Jenis Instrumen tes <i>HOTS</i>	Tingkat Kognitif	Nomor Instrumen tes	Persentase
Berpikir Kritis ( <i>Critical Thinking</i> )	C4	1, 2, 5, 9, 10, 12	40%
	C5	3, 7, 8, 11	26,7%
Memecahkan Masalah ( <i>Problem Solving</i> )	C6	4, 6, 13, 14, 15	33,3%

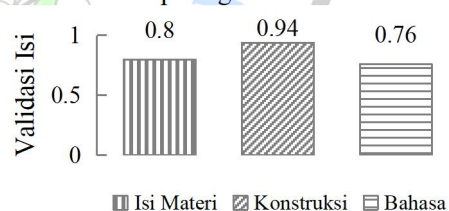
**Hasil Penelitian**

1) *Validasi Item Tes*

Instrumen tes diuji secara kualitatif melalui uji validasi isi dan secara kuantitatif melalui uji validasi konstruksi pada skala kecil dan skala besar.

a. *Validasi Isi*

Validasi isi item tes oleh ahli rata-rata diperoleh nilai rata-rata 0,848 dengan kriteria sangat valid dan valid. Persentase hasil uji skala kecil tertera pada gambar 2.

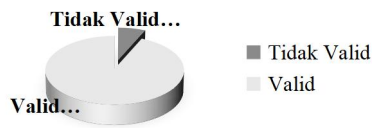


Gambar 2. Hasil Uji Validasi Isi Item Tes

Gambar 2 menunjukkan hasil uji validasi item tes oleh ahli bahwa instrumen tes sangat valid pada aspek konstruksi, valid pada aspek isi materi dan bahasa. Seluruh butir tes layak diuji pada skala kecil dengan 14 dari 15 butir tes perlu revisi pada aspek materi dan bahasa.

b. *Validasi Konstruksi Butir Tes*

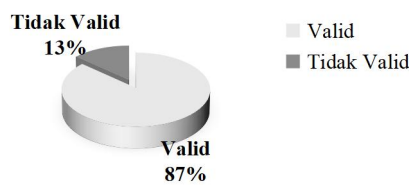
Validasi konstruksi instrumen tes pada skala kecil rata-rata sebesar 0,693 dengan indeks *V* berkisar 0,2193 – 0,9278 pada taraf signifikan 5%. Persentase hasil uji skala kecil tertera pada gambar 3.



Gambar 3. Persentase Hasil Uji Validasi Item Tes Skala kecil

Gambar 3 menunjukkan persentase hasil uji validasi item tes pada kelas skala kecil bahwa 14 item tes terbukti valid ( $r_{XY \text{ hitung}} > 0,4973$ ) dan 1 item tes tidak valid.

Validasi item instrumen tes pada skala besar rata-rata 0,496 yang berkisar antara 0,084 – 0,838. Persentase hasil uji skala kecil tertera pada gambar 4.



Gambar 4. Hasil Uji Validasi Item Tes Skala besar

Gambar 4 menunjukkan bahwa 13 item tes teruji valid ( $r_{XY \text{ hitung}} > 0,2108$ ) dan 2 item tes tidak valid.

#### 2) Reliabilitas Item Tes

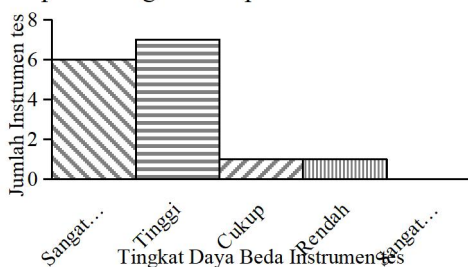
Hasil uji reliabilitas pada skala kecil diperoleh sebesar 0,931 dan pada lapangan lebih luas diperoleh sebesar 0,855 dengan taraf signifikan 5% (Tabel 6). Data menunjukkan bahwa instrumen tes sangat reliabel pada uji skala kecil. Data uji skala besar menunjukkan bahwa setelah direvisi, instrumen tes memiliki reliabel tinggi.

Tabel 6. Reliabilitas Intrumen Tes pada Uji

Skala	Jumlah Instrumen tes	Reliabilitas	Keterangan
Kecil	15	0,931	Sangat Reliabel
Besar		0,855	Reliabel

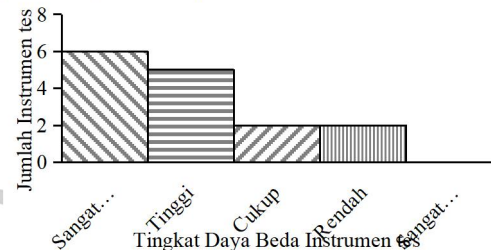
#### 3) Daya Beda Item Tes

Daya beda instrumen tes pada skala kecil rata-rata sebesar 0,57 (Gambar 5). Data pada gambar 5 menunjukkan bahwa 87 % butir tes pada kategori sangat tinggi dan tinggi serta 13% pada kategori cukup dan tidak baik.



Gambar 5. Hasil Uji Daya Beda Item Tes Skala kecil

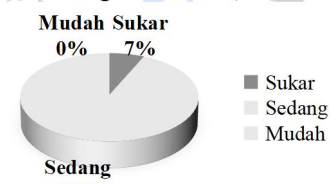
Daya beda instrumen tes pada skala besar rata-rata 0,53 (Gambar 6). Data pada gambar 6 menunjukkan bahwa 73 % butir tes pada kategori sangat tinggi dan tinggi serta 27% pada kategori cukup dan tidak baik.



Gambar 6. Hasil Uji Daya Beda Item Tes Skala besar

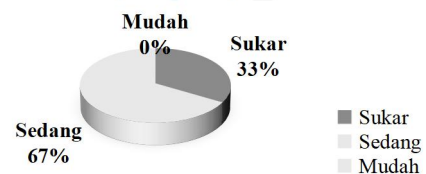
#### 4) Tingkat Kesukaran

Hasil analisis terhadap 15 item tes yang diujikan pada skala kecil bahwa terdapat 14 item tes tergolong sedang dan 1 item tes tergolong sukar serta tidak ada instrumen tes yang termasuk kategori mudah (Gambar 7).



Gambar 7. Persentase Tingkat Kesukaran Item tes Skala kecil

Hasil analisis terhadap 15 item tes yang diujikan, terdapat 10 item tes (67%) tergolong sedang dan 5 item tes (33%) tergolong sukar serta tidak ada instrumen tes yang termasuk kategori mudah (Gambar 8).



Gambar 8. Tingkat Kesukaran Item tes Skala besar

#### 5) Efektivitas Distraktor Item Tes

Efektivitas distraktor item tes dianalisis secara manual menggunakan Excel. Hasil uji efektivitas distraktor item tes tertera pada tabel 7.

Tabel 7. Efektivitas Distraktor Item tes pada Skala kecil

No.	Efektivitas Distraktor	Nomor Instrumen tes	Kriteria
1.	4 berfungsi	14	Sangat Baik
2.	3 berfungsi	1,4,7,8,9,11,12,13,15	Baik
3.	2 berfungsi	2, 3, 5, 10	Cukup

4.	1 berfungsi	6	Jelek
5.	Tidak ada berfungsi	-	Sangat Jelek

Data tabel 7 menunjukkan bahwa 66,67% item tes memiliki distraktor baik, 26,66% cukup dan 1 instrumen tes 6,67% tidak baik.

Hasil uji efektivitas distractor skala besar didapat bahwa 93,33% item tes memiliki distractor baik dan 6,67% cukup. Efektivitas distraktor item tes tertera pada tabel 8.

Tabel 8. Efektivitas Distraktor Item tes pada Skala besar

No.	Efektivitas Distraktor	Nomor Instrumen tes	Kriteria
1.	4 berfungsi	6,7, 10	Sangat Baik
2.	3 berfungsi	1-5,8,9,12-15	Baik
3.	2 berfungsi	11	Cukup
4.	1 berfungsi	-	Jelek
5.	Tidak ada berfungsi	-	Sangat Jelek

#### 6) Pemilihan Kualitas Item tes

Kualitas instrumen tes ditinjau pada aspek validitas, reliabilitas, daya beda, tingkat kesukaran dan efektivitas distraktor instrumen tes. Item tes yang telah diuji pada skala kecil memiliki kualitas yang diinterpretasikan pada tabel 9.

Tabel 9. Kualitas Item Tes pada Uji Skala Kecil

Instrumen tes Nomor	Jumlah kriteria terpenuhi	Kualitas Item tes	Keterangan Revisi	Lanjut Uji Skala besar
1,4,8,9, 11,13-15	8	Sangat Baik	Tidak ada	Ya
2, 3, 5, 6, 10, 12	6	Baik	Minor	Belum
7	1	Sedang	Mayor	Belum
-	0	Jelek	Buang	Tidak
-	0	Sangat Jelek	Buang	Tidak

Data tabel 9 menunjukkan bahwa 6 item tes perlu direvisi secara minor, 1 item perlu revisi mayor dan tidak ada item yang dibuang.

Instrumen tes yang telah direvisi selanjutnya diuji pada aspek validitas, reliabilitas, daya beda, tingkat kesukaran dan efektivitas distraktor pada skala besar. Kualitas instrumen tes dapat diinterpretasikan pada tabel 10.

Tabel 10. Kualitas Item Tes pada Uji Skala Besar

Instrumen tes Nomor	Jumlah kriteria terpenuhi	Kualitas Item tes	Keterangan Revisi	Masuk Bank instrumen tes
1,2,8,9 12,13, 14, 15	8	Sangat Baik	Tidak ada	Ya
3,7,11	3	Baik	Minor	Belum
4,5,6,10	4	Sedang	Mayor	Belum
-	0	Jelek	Buang	Tidak
-	0	Sangat Jelek	Buang	Tidak

Data tabel 10 menunjukkan bahwa 3 item tes perlu direvisi minor, 4 item tes perlu revisi mayor dan tidak ada instrumen tes yang dibuang. Instrumen tes yang telah layak kemudian dimasukkan pada bank instrumen tes Fisika berbasis HOTS.

## Pembahasan

### 1) Validasi Item tes

Diagram batang pada gambar 2 menunjukkan bahwa instrumen tes valid dengan revisi pada aspek materi dan Bahasa. Butir tes yang belum memenuhi aspek materi disebabkan oleh (1) Kesesuaian antara indikator instrumen tes, tingkat kognitif yang ingin dicapai dengan pertanyaan dan pilihan jawaban yang diberikan yang belum tepat. (2) Stimulus pada instrumen tes dan pilihan jawaban belum kontekstual. Butir instrumen tes yang belum memenuhi aspek Bahasa disebabkan karena kalimat yang digunakan belum komunikatif dan kata-kata pada pilihan jawaban yang berulang.

Data pada gambar 3 dan gambar 4 menunjukkan bahwa instrumen tes fisika berbasis HOTS teruji valid, dalam artian bahwa instrumen tes dapat mengukur keterampilan berpikir kritis peserta tes SMAN 11 Medan dan SMAS Budi Satrya. Item tes yang valid mencerminkan instrumen tes telah memiliki kehandalan dan tidak diragukan ketepatan instrumen tes dalam mengukur kemampuan peserta didik (Sudijono, 2012). Instrumen tes valid karena konstruksi dan materi mencakup keseluruhan yang ingin diukur.

Validasi instrumen tes pada skala kecil lebih tinggi dibandingkan pada skala besar. Perbedaan disebabkan karena data nilai dan jawaban peserta tes pada skala kecil lebih variatif dibandingkan peserta tes skala besar. Instrumen tes semakin valid jika skor dan jawaban peserta tes lebih variatif.

Hasil uji validasi instrumen tes tidak sejalan dengan penelitian Afriani, dkk. (2018) dan Sanjaya (2013). Instrumen tes akan semakin valid apabila jumlah peserta tes semakin banyak. Semakin banyak peserta tes, semakin bervariasi jawaban yang didapat, maka instrumen menjadi semakin valid (Afriani, dkk., 2018).

Validitas instrumen tes pada skala kecil lebih tinggi dibandingkan skala besar disebabkan karena nilai dan jawaban peserta tes lebih bervariasi dan rata-rata nilai peserta tes skala kecil lebih tinggi dibandingkan skala besar. Nilai peserta tes skala besar lebih rendah disebabkan karena dua kemungkinan penyebab, yaitu kemampuan berpikir tingkat tinggi peserta tes lebih rendah dibanding skala kecil, dan instrumen tes yang semakin sukar. Instrumen tes

yang telah uji skala kecil direvisi sehingga pokok pertanyaan instrumen tes lebih dalam, pilihan jawaban yang lebih analitis serta tingkat kognitif instrumen tes yang lebih tinggi sehingga instrumen tes skala besar menjadi lebih sukar dibanding instrumen skala kecil.

Nilai dan jawaban peserta tes skala kecil lebih bervariasi dibandingkan skala besar disebabkan karena konstruksi dan materi item tes skala kecil lebih mencakup keseluruhan yang ingin diukur. Pokok pernyataan, pokok pertanyaan dan pilihan jawaban skala kecil lebih tepat mengukur keterampilan berpikir tingkat tinggi peserta tes dengan tingkat kesukaran item tes sedang, sehingga perlu perbaikan pada aspek konstruksi dengan tingkat kesukaran instrumen tes yang sedang pada revisi hasil uji skala besar sebelum dimasukkan pada bank instrumen tes.

### 2) Reliabilitas

Data tabel 6 menunjukkan bahwa kedua instrumen tes reliabel ( $r_{11} \geq 0,70$ ), namun hasil uji skala kecil lebih tinggi dibandingkan pada skala besar. Perbedaan disebabkan nilai dan distribusi jawaban peserta tes skala kecil lebih bervariasi dibandingkan peserta tes skala besar. Reliabilitas instrumen tes mengacu pada kekonsistenan atau kestabilan hasil penilaian (Reynold, 2006); (Arifin, 2017).

Instrumen yang memiliki reliabilitas tinggi menunjukkan konsistensi instrumen dalam mengukur keterampilan berpikir tingkat tinggi peserta tes yang lebih baik, taraf kepercayaan evaluator dalam menempatkan instrumen tes sebagai hasil evaluasi dan faktor penting pertimbangan hasil interpretasi instrumen tes dapat dioperasionalkan (Sukardi, 2008). Instrumen yang reliabel akan memperoleh hasil yang tidak jauh berbeda bila digunakan di sekolah lain (Marwan, dkk., 2020). Kekonsistenan instrumen tes mengacu pada presisi nilai dan jawaban peserta tes, baik di SMAN 11 Medan maupun SMAS Budi Satrya. Data pada hasil uji skala besar pada uji skala besar di SMAN 11 Medan dan SMAS Budi Satrya tidak presisi sehingga instrumen tes yang telah direvisi dan diuji kembali pada skala besar tidak konsisten mengukur keterampilan berpikir tingkat tinggi peserta tes. Ketidakkonsistenan instrumen tes disebabkan karena keterampilan berpikir tingkat tinggi peserta tes di SMAN 11 Medan dan SMAS Budi Satrya memiliki perbedaan yang signifikan.

### 3) Daya Beda

Data pada gambar 5 dan 6 menunjukkan bahwa daya beda instrumen tes dominan baik, beberapa masih perlu revisi serta tidak ada instrumen tes yang dibuang (Budiman & Jailani, 2014). Hal ini menunjukkan bahwa keseluruhan

instrumen tes mampu membedakan peserta tes yang memiliki keterampilan berpikir kritis dengan peserta tes yang belum memiliki keterampilan berpikir kritis (Yunita, 2012).

Hasil uji lapangan menunjukkan bahwa instrumen tes pada skala besar memiliki daya beda rendah yang lebih dominan dibandingkan daya beda pada skala kecil. Daya beda instrumen tes yang rendah disebabkan karena instrumen tes yang mengandung bias dan terlalu sulit (Sarea, dkk., 2015). Hasil uji ini tidak sejalan dengan penelitian Budiman & Jailani (2014), karena semakin banyak subjek, maka daya beda item tes akan semakin baik.

### 4) Tingkat Kesukaran

Data pada gambar 7 dan gambar 8 menunjukkan bahwa instrumen tes lebih sukar pada skala besar dibandingkan skala kecil. Instrumen tes yang baik adalah instrumen tes dengan tingkat kesukaran sedang (Widiyanto, 2018); (Arikunto, 2008). Instrumen tes skala besar lebih sukar dibandingkan skala kecil disebabkan karena instrumen tes skala kecil direvisi pada pokok pertanyaan instrumen tes dengan tingkat kognitif yang lebih tinggi dan pilihan jawaban yang lebih analitis. Pokok pertanyaan pada instrumen tes skala besar menjadi lebih tinggi dikarenakan item direvisi dengan memanipulasi kata kerja operasional hingga item level kognitif item ditingkatkan bernalar (L3). Distraktor menjadi lebih analitis dari sebelumnya karena distraktor direvisi dengan semua distraktor yang diberikan secara teoritis masuk akal, tetapi kuncinya adalah jawaban 'terbaik' (Scully, 2017).

### 5) Efektivitas Distraktor

Analisis data pada tabel 7 dan tabel 8 menunjukkan bahwa distraktor instrumen tes pada skala besar lebih baik dari skala kecil dan kedua uji telah berfungsi sebagai pengecoh dengan beberapa item yang perlu revisi. Hal ini sejalan dengan penelitian Sari dkk (2018) mengenai pengembangan instrumen tes berbasis HOTS menggunakan metode R&D. Distraktor yang baik untuk instrumen tes berbasis HOTS adalah distraktor yang mirip dengan kunci item dan menuntut penilaian diskriminatif tingkat tinggi (Scully, 2017).

Hasil uji skala kecil diperoleh 1 item tes yang memiliki distraktor jelek. Distraktor yang jelek disebabkan kalimat pada distraktor yang rancu (Windarto & Martubi, 2017). Distraktor yang terlalu rancu kalimatnya perlu diperbaiki pada aspek bahasa sehingga distraktor dapat berfungsi dengan baik. Distraktor yang dapat



berfungsi dengan baik menjadikan kualitas item tes baik (Sari, dkk., 2018).

Hasil uji lapangan skala kecil dan skala besar menunjukkan bahwa kualitas instrumen tes pada lapangan skala kecil lebih baik dibandingkan skala besar berdasarkan pada aspek validasi, reliabilitas, daya beda dan tingkat kesukaran. Sedangkan aspek efektivitas distraktor item tes lebih baik pada skala besar dibandingkan skala kecil. Kualitas item tes pada skala kecil secara keseluruhan lebih baik daripada skala besar. Hasil penelitian tidak sejalan dengan penelitian Harahap (2019) tentang tentang pengembangan tes objektif *HOTS* Fisika di SMA/MA menggunakan model Borg & Gall dengan hasil instrumen pada kategori layak berupa aspek validasi, reliabilitas, daya beda, tingkat kesukaran dan analisis pengecoh dengan hasil uji skala besar lebih baik dibandingkan uji skala kecil. Hal serupa pada penelitian Sanjaya (2013) yang menyatakan bahwa semakin banyak subjek penelitian maka instrumen yang dihasilkan akan semakin baik. Kualitas item tes pada skala besar menjadi kurang dibandingkan sebelum revisi skala kecil disebabkan instrumen tes pada skala kecil yang direvisi tidak sesuai dengan aspek yang perlu direvisi sehingga konstruksi instrumen tes menjadi kurang harmoni dan tidak tepat dalam merevisi permasalahan.

Hasil uji item tes pada aspek validasi konstruksi, reliabilitas, daya beda dan tingkat kesukaran menunjukkan adanya hubungan yang saling terkait. Tinggi rendahnya validitas butir soal dapat mempengaruhi tinggi rendahnya reliabilitas (Aisah, 2020). Semakin baik daya beda dan tingkat kesukaran item tes maka item tes akan semakin valid dan reliabel, begitu pula sebaliknya. Validasi yang diuji pada tiap item tes berupa validasi konstruksi. Konstruksi item tes menjadi baik apabila kalimat pada pokok pernyataan, pokok pertanyaan dan pilihan jawaban jelas, ringkas dan tidak mengandung kalimat yang rancu. Konstruksi item tes yang jelas ringkas dan tidak rancu ini menjadi factor penentu kualitas daya beda dan tingkat kesukaran item tes.

Pilihan jawaban item tes skala besar menjadi semakin analitis namun lebih rancu dari item tes pada skala kecil, sehingga pengecoh lebih berfungsi pada skala besar.

Pengecoh yang terlalu mengecoh membuat item tes menjadi kurang valid dan reliabel, lebih sukar dan daya beda kurang baik dari sebelumnya. Meskipun demikian, perbedaan tidak terlalu signifikan, sehingga meskipun kualitas butir instrumen tes skala besar menurun dibanding skala kecil, rata-rata kualitas instrumen tes valid, reliabel, daya beda baik, tingkat kesukaran sedang dan efektivitas distraktor baik sehingga instrumen tes layak untuk mengukur keterampilan berpikir kritis tingkat tinggi di Kota Medan dengan revisi item tes pada aspek validasi, reliabilitas, daya beda dan tingkat kesukaran.

## KESIMPULAN DAN SARAN

### Kesimpulan

Instrumen tes Fisika berbasis *Higher Order Thinking Skills (HOTS)* Tingkat SMA di Kota Medan telah layak digunakan dengan karakteristik:

- (1) Valid oleh ahli dengan rata-rata 0,8 dan telah mendapatkan bukti empiris melalui validasi konstruksi dengan 93% item tes valid ( $r_{hitung} > r_{tabel}$ ) pada hasil uji skala kecil dan 87% item tes valid pada hasil uji skala besar.
- (2) Reliabel dengan nilai sebesar 0,931 pada kategori sangat tinggi ( $r \geq 0,70$ .) untuk hasil uji skala kecil dan 0,787 dengan kategori tinggi ( $r \geq 0,70$ ) pada hasil uji skala besar.
- (3) Daya beda item tes rata-rata 0,57 dengan 87% item kategori baik pada uji skala kecil dan daya beda rata-rata 0,53 dengan 73% item kategori baik pada uji skala besar.
- (4) Tingkat kesukaran item tes kategori sedang sebanyak 93 % pada uji skala kecil dan 67% item tes sedang pada uji skala besar.
- (5) Efektivitas distraktor item tes yang sangat baik 67%, baik pada hasil uji skala kecil dan 93% distraktor item tes pada hasil uji skala besar.

### DAFTAR PUSTAKA

- Afriani, E., Maria, H.T. & Oktaviany, E. (2019). Pengembangan Tes Higher Order Thinking Skills (HOTS) Materi Gerak Lurus Berubah Beraturan untuk SMA. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa*, Vol 8, No 3: 1-12

- Aisah, S. (2020). Pengembangan Instrumen Penilaian Higher Order Thinking Skills (HOTS) Pada Mata Pelajaran Korespondensi Kelas X OTP di SMK Negeri 1 Jombang. *Jurnal Pendidikan Administrasi Perkantoran (JPAP)*, Vol 8, No 1:146-156. <https://journal.unesa.ac.id/index.php/jpap>
- Arifin, Z. (2017). Kriteria Instrumen dalam suatu Penelitian. *Jurnal THEOREMS (The Original Research of Mathematics)*. Vol 2, No 1: 28-36
- Arikunto, S. (2008). *Dasar - Dasar Evaluasi Pendidikan (Edisi Revisi)*. Bumi Aksara, Jakarta
- Badan Standar Nasional Pendidikan. (2006). *Standar Kompetensi dan Kompetensi Dasar Mata Pelajaran Fisika Untuk SMA dan MA*. BSNP-Depdiknas, Jakarta
- Borg, W. R. and Gall, M.D. (1989). *Educational research: an introduction 5th*. Longman, New York
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD.
- Budiman, A., Jailani. (2014). Pengembangan Instrumen Asesmen Higher Order Thinking Skills (HOTS) Pada Mata Pelajaran Matematika SMP Kelas VIII Semester 1. *Jurnal Riset Pendidikan Matematika*, Vol 1 No 2 : 139-141
- Creswell, J. W. (2013). *Steps in conducting a scholarly mixed methods study*. University of Linkoln, Lincoln
- Harahap, W. (2019). *Pengembangan Tes Objektif Higher Order Thinking Skill (HOTS) Fisika Materi Suhu dan Kalor di SMA/MA. (Studi Kasus di MAN 2 Model Medan)*. Skripsi, Pendidikan Fisika, FMIPA, Universitas Negeri Medan
- Istiyono, E., Mardapi, D. & Suparno. (2014). Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PysTHOTS) Peserta Didik SMA. *Jurnal Penelitian dan Evaluasi Pendidikan*, Vol 18, No 1: 1-12
- Kepala Pusat Penilaian Pendidikan Balitbang Kemdikbud. (2017). *Panduan Penulisan Instrumen tes SMA/MA-SMK Tahun 2017*. Pusat Penilaian Pendidikan Badan Penelitian Dan Pengembangan Kementerian Pendidikan dan Kebudayaan, Jakarta
- Kadir, A. (2015). Menyusun dan Menganalisis Tes Hasil Belajar. *Jurnal Al: Ta'dib*, 70-81
- Kistiono. (2019). Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika SMA. *Jurnal Inovasi dan Pembelajaran Fisika*. Vol 06, No 1, 70-81
- Marthunis, Khaldun, I & Zulfadli. (2015). Analisis Kualitas Item tes Ujian Semester Genap Mata Pelajaran Kimia Kelas X MAN Model Banda Aceh Tahun Pelajaran 2014/2015 Menggunakan Program Proanaltes. *Jurnal Ilmiah Mahapeserta tes Pendidikan Kimia (JIMPK)*. Vol 1, No 4 :70-78.
- Marwan, M., Khaeruddin, Amin, B.D. (2020). Pengembangan Instrumen Asesmen Higher Order Thinking Skills (HOTS) Pada Bidang Studi Fisika. *Prosiding Seminar Nasional Fisika PPs UNM*, Vol 2: 116 - 119. Website: <http://ojs.unm.ac.id/semnastfisika>
- Organisation for Economic Cooperation and Development. (2019). *PISA 2018 Results Combined Executive Summaries Volume I, II & III*, Boston College, Boston
- Saidah, N., Yulistianti, H.D., & Megawati, E. (2019). Analisis Instrumen Tes Higher Order Thinking Matematika SMP. *Jurnal Pendidikan Matematika*, Vol 13 No 1: 41-54
- Sani, R.A. (2019). *Pembelajaran Berbasis HOTS (Higher Order Thinking Skills)*. Tira Smart, Tangerang
- Sani, R. A., Rahmatsyah, & Bunawan, W. (2019). *Instrumen tes Fisika HOTS*. Bumi Aksara, Jakarta
- Sarea, M., Sayahrul & Hadi, S. (2015). Analisis Kualitas Instrumen tes Ujian Akhir Semester Mata Pelajaran Kimia SMA di Kabupaten Gowa. *Jurnal Evaluasi Pendidikan*. Vol 3, No 1:35-43.
- Schraw, G., & Robinson, D. H. (2011). *Assessment of higher order thinking skills*. Information Age Publishing, Inc
- Scully, D. (2017). Constructing Multiple-Choice Items to Measure Higher-Order Thinking. *Practical Assessment, Research & Evaluation*, Vol 22, No 4: 1-13 Available online: <http://pareonline.net/getvn.asp?v=22&n=4>
- Simalango, M. M. (2018). Kesulitan peserta tes dalam menyelesaikan instrumen tes-instrumen tes PISA pada konten change

- and relationship level 4, 5, dan 6 di SMP N 1 Indralaya. *Journal Pendidikan Matematika*, **Vol 12, No 1**: 43–58
- Sugiyono, (2017). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Alfabeta, Bandung
- Sukardi. (2008). *Evaluasi Pendidikan Prinsip dan Operasional*. Bumi Aksara, Jakarta
- Wadana, I W. (2017). *Modul Penyusunan Instrumen tes Higher Order Thinking Skills (HOTS)*. Direktorat Pembinaan SMA, Jakarta
- Widiyanto, J. (2018). *Evaluasi Pembelajaran (Sesuai dengan Kurikulum 2013) Konsep, Prinsip & Prosedur*. UNIPMA Press, Madiun
- Windarto, F., Martubi. (2017). Analisis Item tes Ujian Akhir Semester Genap Mata Diklat Dasar-Dasar Mesin (Test Item Analysis of Even Semester Final Exam in The Basic of Engine Subject). *Jurnal Pendidikan Teknik Otomotif*. **Vol IX, No 2**: 132-143

