

**PENGEMBANGAN SOAL *HIGHER ORDER THINKING SKILLS* MATEMATIKA
LINGKUP MATERI UJIAN NASIONAL SMP/MTS DI SMPN 1 MATARAM TAHUN
AJARAN 2019/2020**

Muhamad Syahidul Qirom*, Nyoman Sridana, Sudi Prayitno

Program Studi Pendidikan Matematika, Universitas Mataram

*Email: syahidulqirom36@gmail.com

Abstrak - Soal Matematika berbasis *Higher Order Thinking Skills* dapat membantu meningkatkan keterampilan abad 21 siswa sehingga Kementerian Pendidikan dan Kebudayaan mencanangkan pelaksanaan program penilaian berbasis HOTS di setiap jenjang pendidikan. Faktanya adalah di SMPN 1 Mataram program tersebut belum berjalan sesuai mestinya. Terdapat dua masalah yang menyebabkan hal tersebut terjadi, yaitu kurangnya kemampuan guru dalam membuat soal HOTS matematika dan ketersediaan soal berbasis HOTS pada buku pelajaran matematika yang digunakan di sekolah tersebut masih belum cukup. Penelitian ini bertujuan untuk menghasilkan soal HOTS matematika untuk membantu memenuhi ketersediaan referensi soal HOTS baik bagi guru maupun siswa di SMPN 1 Mataram. Metode yang digunakan dalam penelitian ini adalah model pengembangan Tessmer yang terdiri atas dua tahap yakni *preliminary* dan *formative evaluation*. Hasil dari penelitian ini adalah diperolehnya 20 soal HOTS matematika dengan nilai validitas 75,9375%, kepraktisan 80,41%, keefektifan 79%, kesukaran soal dalam kategori sedang, daya pembeda dalam kategori cukup, soal reliabel, dan menunjukkan hasil sesuai harapan yakni keterampilan berpikir tingkat tinggi siswa berada pada kategori cukup. Kesimpulan dalam penelitian ini adalah 1) telah dihasilkan soal HOTS matematika yang valid, praktis, dan efektif; 2) dalam mengembangkan soal HOTS stimulus harus menarik, baru, kontekstual, mendorong siswa untuk menelaah informasi yang diberikan, dan menggabungkan satu konsep dengan konsep lainnya.

Kata Kunci: Keterampilan 4C, Taksonomi Bloom Revisi, penalaran

Abstract - *Mathematics HOTS Question has trusted can improve student's skills in 21st century so Ministry of Education and Culture launched assessment program based on HOTS in every grade of school. The fact is at SMPN 1 Mataram the program does not run properly. There are two problems caused it happens, first the lack of teacher's ability to build math HOTS question and the availability of HOTS question in math textbooks that used at school are not enough. This study aims to produce math HOTS question to help providing more references about math HOTS question for teacher and student at SMPN 1 Mataram. The method in this study adapts development model by Tessmer which consists two main phases, namely preliminary and formative evaluation. The result of this study has obtained 20 math HOTS question with score of validity 75.9375%, practicality 80.41%, effectiveness 79%, the difficulty index of questions in moderate category with enough discrimination power, questions also are reliable and shows the result as expected which student's HOTS in enough category. In conclusion the study 1) has produced math HOTS questions with valid, practical, and effective category and 2) in order to develop HOTS questions, we need to consider about stimulus that are attractive, new, contextual, and it encourage student's reasoning, in addition the questions should be connect one concept with others.*

Keywords: 4C's Skills, Revised Bloom's Taxonomy, Reasoning

PENDAHULUAN

Mulainya era industri 4.0 serta berkembangnya dunia IPTEK, tentunya merupakan suatu tantangan bagi masyarakat abad 21 khususnya bangsa Indonesia. Dengan berbagai kemajuan yang ada, persaingan baik di kancah nasional maupun internasional akan semakin pesat. Sehingga setiap orang dituntut untuk menguasai keterampilan abad 21 dalam rangka mempersiapkan diri menghadapi persaingan yang terus meningkat. Keterampilan abad 21 sendiri menekankan pada empat hal utama

atau lebih dikenal dengan 4C, yakni *critical thinking, creativity, communication, and collaboration* (Bialik, 2015:1).

Dalam menghadapi tantangan tersebut salah satu upaya yang dilakukan pemerintah adalah melalui pengembangan dan inovasi kurikulum 2013. Inovasi terus dilakukan agar relevan dengan perkembangan zaman (Kunandar, 2013:16). Kurikulum 2013 dirancang sesuai dengan standar pendidikan internasional sehingga telah dilakukan berbagai penyempurnaan. Pada standar isi, dilakukan pengurangan materi yang tidak

relevan kemudian memberikan pendalaman dan perluasan materi yang relevan serta diperkaya dengan kebutuhan peserta didik untuk berpikir kritis dan analitis. Pada standar penilaian, diadaptasi secara bertahap model-model penilaian standar internasional. Melalui penilaian tersebut diharapkan dapat membantu peserta didik meningkatkan keterampilan berpikir tingkat tinggi (*Higher Order Thinking Skills/HOTS*), karena hal tersebut dapat mendorong mereka untuk berpikir secara luas dan mendalam tentang suatu materi pelajaran (Widana, dkk., 2019:1).

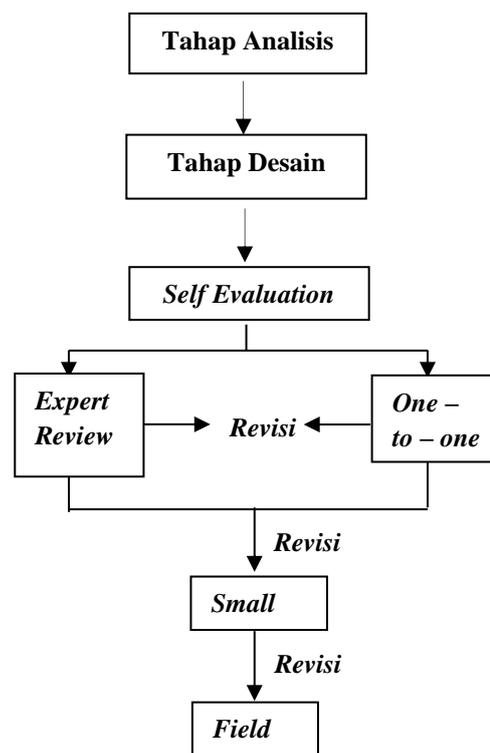
Penilaian berbasis HOTS hingga saat ini tampaknya masih belum maksimal dilakukan di semua sekolah. Salah satu sekolah yang mengalami kendala dalam melakukan penilaian berbasis HOTS adalah SMPN 1 Mataram. Hal ini disebabkan karena beberapa faktor seperti kurangnya kemampuan guru membuat soal, siswa yang jarang dibiasakan soal HOTS, hingga pada ketersediaan sumber soal-soal berbasis HOTS pada buku pelajaran di sekolah tersebut. Oleh karena itu, dalam rangka membantu ketersediaan sumber soal HOTS baik bagi guru maupun siswa di SMPN 1 Mataram maka peneliti tertarik untuk melakukan penelitian tentang Pengembangan Soal *Higher Order Thinking Skills* Matematika Lingkup Materi Ujian Nasional SMP/MTs di SMPN 1 Mataram Tahun Ajaran 2019/2020.

Tujuan dari penelitian ini adalah untuk 1) menghasilkan satu set soal *higher order thinking skills* matematika pada lingkup materi ujian nasional tingkat SMP/MTs yang valid, praktis, dan efektif, 2) mendeskripsikan pengembangan soal HOTS matematika lingkup materi Ujian Nasional SMP/MTs. Adapun manfaat dari penelitian ini adalah 1) bagi siswa, dapat digunakan sebagai latihan dalam menjawab soal bertipe HOTS, 2) bagi guru, dapat digunakan sebagai acuan dalam mengembangkan soal

HOTS dan sebagai referensi untuk soal latihan atau tes, 3) bagi peneliti, dapat memberikan pengalaman langsung dalam mengembangkan soal HOTS.

METODE PENELITIAN

Jenis penelitian yang digunakan adalah metode *Research and Development* (R&D) dengan model *Formative Evaluation* yang dikembangkan oleh Tessmer (1994:3). Dalam penelitian ini terdiri atas dua tahap utama, yakni tahap *preliminary* dan *Formative Evaluation*. Pada tahap *preliminary* terdiri atas tahap analisis dan tahap desain. Pada tahap *Formative Evaluation* terdiri atas *self evaluation*, *expert review*, *one – to – one*, *small group*, dan *field test*.



Gambar 1. Desain Formative Evaluation Tessmer (Zulkardi, 2006)

Subjek pada penelitian ini adalah siswa kelas IX – A di SMPN 1 Mataram tahun ajaran 2019/2020. Penentuan subjek penelitian berdasarkan pada saran guru mata pelajaran matematika kelas IX di SMPN 1 Mataram, karena pada kelas tersebut cenderung lebih heterogen daripada kelas

lainnya. Di kelas tersebut ada siswa yang memiliki kemampuan matematis yang sangat tinggi namun di kelas tersebut juga ada siswa yang memiliki kemampuan matematis yang sangat rendah. Objek pada penelitian ini adalah soal HOTS matematika yang disusun berdasarkan lingkup materi ujian nasional matematika tingkat SMP sederajat. HOTS dalam penelitian ini berdasarkan pada HOTS menurut Taksonomi Bloom Revisi. Prosedur penelitian ini dapat dilihat pada Gambar 1.

Tahap Preliminary

Tahap Analisis

Pada tahap ini terdapat tiga hal yang dianalisis, yakni analisis siswa, materi, dan kurikulum. Analisis siswa bertujuan untuk mengetahui tingkat kemampuan matematis siswa. Analisis kurikulum bertujuan untuk mengetahui kurikulum yang digunakan di sekolah tempat penelitian dilaksanakan. Analisis materi bertujuan untuk mengetahui lingkup materi yang diujikan dalam ujian nasional.

Tahap Desain

Pada tahap ini dikembangkan soal HOTS berdasarkan Taksonomi Bloom Revisi sebanyak 20 buah. Soal HOTS tersebut disesuaikan dengan lingkup materi yang diujikan dalam ujian nasional matematika tingkat SMP sederajat yakni materi bilangan, aljabar, geometri dan pengukuran, serta statistika dan peluang.

Tahap Formative Evaluation

Self evaluation

Pada tahap ini dilakukan evaluasi secara mandiri terhadap soal HOTS yang telah didesain, bertujuan untuk mengevaluasi kembali apakah soal yang telah dibuat sesuai dengan indikator HOTS. Soal yang telah didesain pada tahap ini disebut *prototype* I. Terdapat tiga karakteristik yang menjadi fokus prototipe

yang akan dikembangkan yakni materi, konstruksi, dan bahasa.

Expert Review

Pada tahap ini dilakukan evaluasi atau penilaian oleh ahli mengenai soal HOTS yang telah dibuat sekaligus penilaian terhadap keefektifan soal. Hasil komentar dan saran dari ahli tersebut akan digunakan sebagai acuan dalam merevisi soal HOTS.

One – To – One

Pada tahap ini dilakukan parallel dengan *Expert Review*. *Prototype* I diuji coba pada kepada tiga orang siswa (bukan subjek penelitian) yang memiliki kemampuan berbeda, satu siswa dengan kemampuan tinggi, satu siswa dengan kemampuan sedang, dan satu siswa dengan kemampuan rendah. Respon siswa melalui angket, serta hasil komentar, dan saran dari ahli pada tahap *Expert Review* akan digunakan sebagai acuan dalam merevisi *prototype* I. Hasil revisi pada tahap tersebut disebut *prototype* II.

Small Group

Pada tahap ini dilakukan uji coba terhadap *prototype* II kepada enam orang siswa dengan kemampuan berbeda (bukan subjek penelitian), dua siswa dengan kemampuan tinggi, dua siswa dengan kemampuan sedang, dan dua siswa dengan kemampuan tinggi. Respon siswa melalui angket akan digunakan sebagai acuan dalam merevisi soal yang akan digunakan pada tahap field test. Hasil revisi pada tahap ini disebut *prototype* III.

Field Test

Pada tahap ini *prototype* III diujikan pada siswa kelas IX – A di SMPN 1 Mataram yang merupakan subjek dari penelitian. Pada tahap ini soal HOTS yang telah dikembangkan dihitung nilai reabilitas, daya pembeda, dan indeks kesukarannya.

Tujuan dari tahap ini adalah untuk mendapatkan produk akhir berupa soal matematika pada lingkup materi ujian nasional SMP yang sesuai kriteria HOTS menurut Taksonomi Bloom Revisi.

Instrumen Penelitian dan Teknik Pengumpulan Data

Instrumen penelitian yang digunakan pada penelitian ini adalah lembar angket dan soal tes, namun sebelum digunakan instrument tersebut divalidasi oleh ahli dengan menggunakan lembar validasi. Sedangkan teknik pengumpulan data yang digunakan adalah teknik angket dan tes.

Teknik Analisis Data

Analisis Validitas

Dalam penelitian ini uji validitas yang digunakan adalah Validitas versi Aiken. Menurut Azwar (dalam Novaliah dan Kartowagiran, 2018:64), uji ini didasarkan pada hasil penilaian dari validator terhadap suatu item dari segi sejauh mana item tersebut mewakili sesuatu yang akan diukur. Untuk menghitung validitas berdasarkan Validitas Aiken digunakan rumus berikut:

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

Keterangan:

V = Indeks validitas Aiken

s = r - lo

r = Angka yang diberikan validator

lo = Angka Penilaian Validitas Terendah

c = Angka Penilaian Validitas Tertinggi

n = Banyak Validator

Setelah rata-rata skor validasi diperoleh selanjutnya skor tersebut diolah dan hasilnya dikonsultasikan pada tabel kriteria validitas berikut:

Tabel 1. Kriteria Validitas

Indeks Validitas	Kategori
$0.80 \leq V \leq 1.00$	Sangat Valid
$0.60 \leq V < 0.80$	Valid
$0.40 \leq V < 0.60$	Cukup Valid
$0.20 \leq V < 0.40$	Kurang Valid

Indeks Validitas	Kategori
$0.00 \leq V < 0.20$	Tidak Valid

(Arikunto, 2012)

Analisis Angket Respon Siswa (Kepraktisan)

Data untuk analisis angket ini diperoleh dari penilaian siswa terhadap soal melalui angket yang diberikan. Setelah siswa menjawab soal, mereka akan memberikan penilaian sangat setuju, setuju, tidak setuju, atau sangat tidak setuju terhadap beberapa pernyataan pada angket yang berkaitan dengan kepraktisan soal. Selanjutnya untuk masing-masing aspek pernyataan ditentukan persentase jawaban siswa untuk masing-masing butir pertanyaan dalam angket menggunakan rumus berikut (Widoyoko, 2012):

$$P = \frac{f}{n} \times 100\% \quad (2)$$

Keterangan:

P = persentase jawaban

f = Skor yang diperoleh

n = Skor ideal (skor maksimum)

Persentase yang diperoleh pada masing-masing butir pertanyaan, kemudian ditafsirkan berdasarkan kriterianya. Butir pernyataan diterima apabila memperoleh kriteria minimum baik. Berikut kriteria untuk angket respon siswa.

Tabel 2. Kriteria Persentase Angket Respon Siswa

Presentase Jawaban (%)	Kriteria
$81 \leq P \leq 100$	Sangat Praktis/Baik
$63 \leq P < 81$	Praktis/Baik
$44 \leq P < 63$	Kurang Praktis/Baik
$25 \leq P < 44$	Tidak Praktis/Baik

Terdapat dua bentuk pernyataan dalam angket respon siswa yakni pernyataan positif dan negatif, penilaian masing-masing pernyataan disajikan dalam tabel berikut:

Tabel 3. Skor Pernyataan Positif dan Negatif

Pilihan Jawaban	Pernyataan Positif	Pernyataan Negatif
Sangat Setuju (SS)	4	1
Setuju (S)	3	2
Tidak Setuju (TS)	2	3

Sangat Tidak Setuju (STS)	1	4
---------------------------	---	---

Analisis Butir Soal

Indeks Kesukaran

Dalam mengukur indeks kesukaran soal maka untuk setiap butir soal ditentukan terlebih dahulu banyak siswa yang menjawab benar pada butir-butir soal tersebut, kemudian dibandingkan dengan jumlah siswa yang mengikuti tes. Untuk mengukur tingkat atau indeks kesukaran untuk masing – masing item pilihan ganda, digunakan rumus sebagai berikut (Arikunto, 2012):

$$IK = \frac{B}{JS} \quad (3)$$

Keterangan

IK= Indeks Kesukaran

B = Banyak siswa yang menjawab benar

JS = Jumlah seluruh siswa peserta tes

Hasil dari masing-masing indeks kesukaran soal kemudian disesuaikan dengan kategori indeks kesukaran soal berikut.

Tabel 4. Kategori Indeks Kesukaran

Kriteria Indeks Kesukaran	Kategori
$0 \leq IK < 0,3$	Sukar
$0,3 \leq IK < 0,7$	Sedang
$0,7 \leq IK \leq 1,00$	Mudah

Daya Pembeda

Untuk mengukur daya pembeda maka terlebih dahulu siswa akan dibagi menjadi dua kelompok berdasarkan hasil tes pada tahap *field test*, yakni kelompok atas (50% siswa dengan nilai tertinggi) dan kelompok bawah (50% siswa dengan nilai terendah), selanjutnya untuk setiap butir soal akan dianalisis daya pembedanya dengan mengurangi perbandingan siswa kelompok atas yang menjawab benar dengan perbandingan siswa kelompok bawah yang menjawab benar. Untuk masing – masing item soal pilihan ganda, digunakan rumus sebagai berikut (Arikunto, 2013):

$$D = \frac{JAB}{BA} - \frac{JBB}{BB} \quad (4)$$

Keterangan:

D = Daya pembeda

BA = Banyak siswa pada kelompok atas (50% siswa dengan nilai tertinggi)

BB = Banyak siswa pada kelompok bawah (50% siswa dengan nilai terendah)

JAB = Banyak siswa pada kelompok atas yang menjawab benar

JBB = Banyak siswa pada kelompok bawah yang menjawab benar

Selanjutnya setelah diketahui nilai daya pembeda masing-masing soal, nilai tersebut ditafsirkan sesuai dengan kategori daya pembeda soal berikut.

Tabel 5. Kriteria Daya Pembeda

Kriteria Daya Pembeda	Keterangan
$0,70 \leq D \leq 1,00$	Sangat Baik
$0,40 \leq D < 0,70$	Baik
$0,20 \leq D < 0,40$	Cukup
$0 \leq D < 0,20$	Jelek
$D < 0$	Sangat Jelek

Penyebaran Pilihan Jawaban

Penyebaran pilihan jawaban adalah proporsi siswa yang memilih suatu jawaban tertentu. Melalui penyebaran pilihan jawaban dapat diketahui keefektifan dari suatu pengecoh. Untuk mengetahui penyebaran pilihan jawaban maka jawaban siswa pada setiap butir soal di tahap *field test* dianalisis dengan menjabarkan banyaknya siswa yang menjawab option tertentu. Selanjutnya Penyebaran pilihan jawaban dihitung menggunakan rumus (Zulaiha, 2008):

$$PPJ = \frac{BJ}{N} \quad (5)$$

Keterangan:

PPJ : Penyebaran pilihan jawaban

BJ : Banyak siswa yang memilih jawaban tertentu

N : Jumlah seluruh siswa

Apabila banyak siswa yang memilih opsi pengecoh $\geq 2,5\%$ maka pengecoh tersebut dikatakan efektif. Selain itu, ditentukan juga banyak siswa yang tidak menjawab suatu soal (omit). Adapun besar

omit yang dapat ditoleransi adalah 10% dari banyak siswa (Arikunto, 2013).

Reliabilitas Soal

Reliabilitas soal berkaitan erat dengan bagaimana tingkat ketetapan soal, artinya dimanapun dan kapanpun soal tersebut diujikan akan memberikan hasil yang sama. Realibilitas soal diperoleh melalui analisis terhadap hasil skor tes siswa pada tahap *field test*. Untuk mengukur reliabilitas soal digunakan rumus K-R. 20. Menurut Arikunto (2013), rumus realibilitas dengan K-R 20 sebagai berikut:

$$r_{11} = \left(\frac{n}{n-1}\right) \left(\frac{S^2 - \sum pq}{S^2}\right) \quad (6)$$

Keterangan:

- r_{11} : reliabilitas soal secara keseluruhan
- p : proporsi subjek yang menjawab item dengan benar
- q : proporsi subjek yang menjawab item dengan salah ($q = 1 - p$)
- $\sum pq$: jumlah hasil perkalian antara p dan q
- n : banyaknya item
- S : standar deviasi

Setelah diperoleh indeks reliabilitas (r_{hitung}), selanjutnya mengkonsultasikan indeks tersebut dengan tabel *r product moment* dengan jumlah n yang sama pada taraf signifikansi 5%. Apabila $r_{hitung} \geq r_{tabel}$ maka instrumen dianggap reliabel. Sebaliknya apabila $r_{hitung} < r_{tabel}$ maka instrumen dianggap tidak reliabel (Widoyoko, 2012:163).

Analisis Keefektifan

Keefektifan dari soal HOTS yang dimaksud dalam penelitian ini adalah soal efektif mengukur keterampilan berpikir tingkat tinggi siswa. Sehingga soal HOTS dikatakan efektif dalam penelitian ini apabila:

- a) Menurut penilaian ahli soal tersebut efektif untuk mengukur keterampilan berpikir tingkat tinggi siswa.

- b) Menunjukkan hasil yang sesuai dengan yang diharapkan. Pada penelitian ini diharapkan soal dapat menunjukkan rata-rata hasil tes keterampilan berpikir tingkat tinggi siswa pada kriteria maksimal cukup. Kriteria cukup ditetapkan berdasarkan pada kenyataan di lapangan yakni siswa jarang dibiasakan menjawab soal berbasis HOTS, sehingga keterampilan berpikir tingkat tinggi siswa kemungkinan pada kategori kurang atau cukup (Sagala & Andriani, 2019).

Berikut teknik analisis data untuk menentukan keefektifan soal sesuai kriteria di atas:

- a) Keefektifan Menurut Pendapat Ahli

Proses analisis keefektifan berdasarkan penilaian ahli dilakukan setelah ahli memberikan penilaian (skor 1-5) terhadap instrumen soal (tahap *expert review*), selanjutnya hasil penilaian dari masing-masing ahli disatukan kemudian dianalisis secara kuantitatif dengan menggunakan rumus:

$$PK = \frac{f}{n} \times 100\% \quad (7)$$

(Widoyoko, 2012)

Keterangan:

- PK = Persentase Keefektifan
- f = Skor yang diperoleh
- n = Skor ideal (skor maksimum)

Keefektifan dalam hal ini hanya merupakan suatu tanggapan, sehingga digunakan teknik persentase untuk melihat tingkat keefektifan menurut ahli. Jadi rumus yang digunakan sama dengan teknik analisis angket (non-tes) hanya saja menggunakan kriteria penafsiran yang berbeda. Setelah diketahui presentase keefektifan masing-masing soal, selanjutnya dihitung rata-ratanya dan nilai rata-rata tersebut ditafsirkan sesuai dengan kriteria berikut.

Tabel 6. Kriteria Tingkat Keefektifan Soal

Presentase Keefektifan (%)	Kriteria
$84 \leq P \leq 100$	Sangat Efektif

$68 \leq P < 84$	Efektif
$52 \leq P < 68$	Cukup Efektif
$36 < P < 52$	Kurang Efektif
$20 \leq P < 36$	Tidak Efektif

(Widoyoko, 2012:111)

b) Menunjukkan Hasil Sesuai dengan yang Diharapkan

Keefektifan soal pada bagian ini dilihat setelah subjek penelitian selesai mengerjakan instrumen soal tersebut (tahap *field test*). Dalam penelitian ini keterampilan berpikir tingkat tinggi siswa akan dilihat dari hasil rata-rata skor siswa. Berikut rumus yang digunakan untuk menentukan skor siswa (Arikunto, 2012):

$$S = B$$

Keterangan:

S = Skor siswa

B = Banyak soal yang dijawab benar oleh siswa

Kemudian untuk menentukan rata-rata skor siswa digunakan rumus berikut:

$$R = \frac{\sum_{i=1}^N Si}{N} \quad (8)$$

Keterangan:

Si = Skor siswa ke-i

R = rata-rata

i = 1, 2, 3, dst

N = Total siswa

Selanjutnya rata-rata skor yang diperoleh dikonsultasikan menurut tabel berikut:

Tabel 7. Keterampilan Bepikir Tingkat Tinggi Siswa

Skor	Kategori
$15 \leq \text{skor} \leq 20$	Sangat Baik
$10 \leq \text{skor} < 15$	Baik
$5 \leq \text{skor} < 10$	Cukup
$0 \leq \text{skor} < 5$	Kurang

(Adaptasi Lewy, et al., 2009)

HASIL DAN PEMBAHASAN

Kevalidan Instrumen Soal

Untuk menguji validitas soal digunakan Uji Validitas Aiken. Validitas soal dilakukan melalui dua tahapan, yakni (i) validitas soal secara umum untuk mengetahui kaidah penulisan dan kesesuaian soal dengan kompetensi dasar yang akan

diuji, aturan pembuatan soal berbentuk pilihan ganda, serta tata bahasa yang digunakan; dan (ii) validitas soal HOTS untuk mengetahui tingkat kemunculan indikator HOTS pada soal yang dikembangkan. Masing-masing tahapan dinilai atau divalidasi oleh dua orang dosen Pendidikan Matematika FKIP UNRAM. Secara umum berikut hasil validasi soal HOTS yang dikembangkan:

Tabel 8. Hasil Validasi Instrumen Soal

Jenis Validasi	Rata-Rata Validitas Aiken
Soal Umum	0,7875
Soal HOTS	0,73125
Total	1,51875
Rata-rata	0,759375

Berdasarkan Tabel 3.1 diperoleh bahwa instrumen soal yang dikembangkan memperoleh rata-rata validitas sebesar 0,759375 dengan kategori valid. Hasil ini diperoleh dari penilaian yang diberikan oleh ahli dengan memberikan skor dari 1 (Tidak Baik) sampai 5 (Sangat Baik). Selanjutnya skor tersebut dikalkulasikan untuk kemudian dianalisis lebih lanjut dengan menggunakan Uji Validitas Aiken untuk mengetahui tingkat validitas soal.

Kepraktisan Soal

Kepraktisan soal diukur dari hasil angket respon siswa yang diberikan setelah siswa menjawab soal yang diberikan. Siswa berdasarkan pengalamannya akan memberikan pendapat sangat setuju, setuju, tidak setuju, dan sangat setuju terhadap pernyataan yang diberikan pada angket. Pemberian angket diberikan pada tahap *one-to-one* dan *small group*. Bertujuan untuk mengetahui (i) karakteristik HOTS berupa kebaruan dan kemenarikan soal (menurut pengalaman siswa); dan (ii) kepraktisan soal berupa kebahasaan, alokasi waktu, serta kegrafikan.

Tabel 9. Hasil Kepraktisan Soal

Jenis Validasi	Rata-Rata Validitas Aiken
<i>one-to-one</i>	80
<i>small group</i>	80,417

Dari Tabel 3.2 menunjukkan tingkat kepraktisan soal sudah praktis. Hal ini dilihat dari hasil skor baik pada tahap *one-to-one* maupun *small group* menunjukkan hasil yang ≥ 80 .

Keefektifan Soal

Keefektifan dari soal HOTS yang dimaksud dalam penelitian ini adalah soal efektif mengukur keterampilan berpikir tingkat tinggi siswa. Sehingga soal HOTS dikatakan efektif dalam penelitian ini apabila:

Menurut penilaian ahli soal tersebut efektif untuk mengukur keterampilan berpikir tingkat tinggi siswa. Menunjukkan hasil yang sesuai dengan yang diharapkan. Pada penelitian ini diharapkan soal dapat menunjukkan rata-rata hasil tes keterampilan berpikir tingkat tinggi siswa pada kriteria maksimal cukup. Kriteria cukup ditetapkan berdasarkan pada kenyataan di lapangan yakni siswa jarang dibiasakan menjawab soal berbasis HOTS, sehingga keterampilan berpikir tingkat tinggi siswa kemungkinan pada kategori kurang atau cukup (Sagala & Andriani, 2019).

Penelitian keefektifan menurut ahli dilakukan pada tahap *expert review*. Berdasarkan pengalamannya ahli/validator akan menilai (memberikan tanggapan) apakah soal yang dikembangkan efektif dalam mengukur keterampilan berpikir tingkat tinggi siswa. Diperoleh keefektifan soal menurut ahli memperoleh skor sebesar 79% dengan kategori efektif. Selanjutnya, berdasarkan hasil analisis keterampilan berpikir tingkat tinggi diperoleh bahwa nilai rata-rata tingkat keterampilan berpikir

tingkat tinggi siswa sebesar 6,6 berada dalam kategori cukup. Hal ini menunjukkan bahwa soal yang dikembangkan sudah efektif karena menunjukkan hasil yang sesuai dengan yang diharapkan. Sesuai dengan pendapat Akker dalam Sagala & Andriani (2019) bahwa suatu soal dikatakan efektif apabila menurut ahli soal tersebut efektif dan soal tersebut memberikan hasil sesuai yang diharapkan. Selain itu dari hasil keterampilan berpikir tingkat tinggi tersebut mengindikasikan bahwa adanya efek potensial soal untuk membantu melatih keterampilan berpikir tingkat tinggi siswa.

Analisis Butir Soal

Efektifitas Option dan Omit

Berdasarkan hasil analisis butir soal diperoleh bahwa semua pengecoh (*distractor*) berfungsi dengan baik artinya tidak ada pilihan jawaban yang mengarahkan siswa pada jawaban yang benar atau menyesatkan dan pengecoh tersebut memiliki daya tarik yang besar bagi siswa yang kurang memahami konsep (Arikunto, 2012). Selain itu, semua tingkatan omit pada soal dapat diterima. Hal ini menunjukkan bahwa soal dapat dipahami dan dikerjakan oleh siswa dengan baik.

Menurut Jakwerth (1999) omit atau keadaan dimana siswa tidak merespon atau menjawab soal diakibatkan karena beberapa hal (i) siswa tidak memahami kalimat dalam soal baik kalimat pernyataan atau pertanyaan, (ii) soal dilewatkan oleh siswa tanpa sengaja, (iii) motivasi dalam menjawab soal yang kurang, (iv) waktu menjawab soal, (v) siswa kekurangan pengetahuan untuk menjawab soal tersebut. Jika diperhatikan kembali hasil angket respon siswa, jelas bahwa dari segi kalimat, gambar/grafik, waktu yang diberikan, dan ketertarikan/motivasi siswa dalam menjawab soal mendapat respon baik. Sehingga hasil semua omit diterima dan efektifitas pengecoh yang berfungsi sesuai

dengan respon siswa yang baik pula terhadap soal.

Indeks Kesukaran dan Daya Pembeda

Dari segi tingkat kesukaran diperoleh untuk tingkat kesukaran diperoleh soal dengan kategori mudah 0%, sedang 70%, dan sukar 30%. Hal ini menunjukkan bahwa sebagian besar soal berada pada kategori sedang yakni soal tidak terlalu mudah atau sulit. Selain itu menurut pendapat Arikunto (2012) soal yang baik adalah soal dengan indeks kesukaran dengan kategori sedang (0.30 – 0.70). Namun, tentunya ini kembali lagi pada tujuan pembuatan soal. Apabila soal digunakan untuk seleksi maka jelas soal yang baik adalah soal yang sulit sehingga diperoleh peserta dengan kompetensi yang tinggi. Jika soal dibuat untuk mengukur kecepatan dalam memproses informasi (*speed test*) soal dengan kategori mudah bisa menjadi solusi. Adapun jika soal digunakan untuk membedakan siswa dengan keterampilan atau kemampuan tinggi dan rendah, soal dengan kesulitan sedang lebih diprioritaskan sehingga dapat memetakan kemampuan siswa yang sebenarnya.

Pada penelitian ini, daya pembeda soal dengan kategori sangat baik 0%, baik 20%, cukup 55%, jelek 10%, dan sangat jelek 15%. Artinya bahwa sebagian besar soal cukup dapat membedakan siswa dengan kemampuan tinggi dan rendah. Sedangkan jika tentukan rata-ratanya akan diperoleh skor daya pembeda soal secara umum adalah 0.24 dengan kategori cukup mendekati jelek. Hal ini sejalan dengan hasil dari indeks/tingkat kesukaran soal yang berkategori sedang mendekati sulit. Jika melihat kembali daya pembeda per item/butir soal akan ditemukan soal dengan daya pembeda 0 bahkan negatif yang menunjukkan bahwa soal tersebut tidak memiliki daya pembeda bahkan tidak dapat membedakan siswa dengan kemampuan

tinggi dan rendah. Menurut Arikunto (2012), soal pilihan ganda dengan daya pembeda negatif berarti soal tersebut kemungkinan besar dapat dijawab dengan menebak. Sehingga pada umumnya soal dengan daya pembeda perlu dipertimbangkan untuk dibuang atau direvisi secara keseluruhan.

Hasil tingkat kesukaran soal yang menuju kearah sulit serta daya pembeda yang menuju kearah jelek disebabkan karena soal yang diberikan kepada siswa adalah soal dengan kategori HOTS semua. Apalagi siswa-siswa di sekolah tersebut belum terbiasa mengerjakan soal dengan level tersebut. Oleh karena itu, apabila guru akan membuat soal evaluasi maka soal tersebut sebaiknya tidak hanya soal yang HOTS (C4-C6) namun juga ada soal yang berbasis LOTS (C1-C3). Menurut Brookhart (2010:233) persentase soal LOTS pada suatu tes adalah 55% sedangkan HOTS 45%. Selain itu, sangat penting untuk menentukan proporsi tingkat kesukaran soal, menurut Sudjana (2017:136) perbandingan tingkat kesukaran soal yang mudah, sedang, dan sulit adalah 3:4:3.

Reliabilitas

Melalui skor siswa pada tahap uji lapangan (*Field Test*) diperoleh reliabilitas soal HOTS yang dikembangkan memiliki nilai 0,62 (Lampiran 11). Jika mengkonsultasikan pada *r*-tabel diperoleh *r*-hitung > *r*-tabel sehingga instrumen soal sudah reliabel. Tingkat reliabilitas soal menunjukkan daya ketetapan soal, artinya apabila siswa dites kembali dengan soal tersebut akan memberikan hasil yang tetap. Kata “tetap” menurut Arikunto (2012:100) tidak harus sama, namun lebih menggambarkan kondisi jika hasil tes tersebut berubah-ubah hal tersebut dapat dikatakan tidak berarti. Misalkan pada tes pertama skor A lebih tinggi dari B, maka apabila dites kembali skor A akan “tetap” lebih tinggi dari B bahkan jika skor A dan B

berubah (tidak sama dengan skor pada tes awal).

PENUTUP

Telah dihasilkan perangkat soal yang terdiri atas kisi-kisi soal, kunci jawaban dan pembahasan, serta 20 soal HOTS matematika lingkup materi UN SMP/MTs berbentuk pilihan ganda yang valid, praktis, dan efektif. Perangkat soal berdasarkan pada penilaian ahli/validator memperoleh rata-rata skor sebesar 0,759375 yang menunjukkan kategori valid. Perangkat soal berdasarkan analisis angket respon siswa memperoleh rata-rata skor sebesar 80,417% dengan kategori praktis. Perangkat soal berdasarkan penilaian ahli/validator dengan memperoleh rata-rata skor sebesar 79% dengan kategori efektif dan menunjukkan hasil yang sesuai dengan yang diharapkan yakni rata-rata keterampilan berpikir tingkat tinggi siswa berada dalam kategori cukup dengan skor 6,6.

Adapun dalam mengembangkan soal HOTS ada dua hal utama yang perlu diperhatikan yakni, 1) stimulus soal haruslah menarik, baru, kontekstual, dan mengarahkan siswa untuk melakukan penelaahan atau penalaran terhadap suatu informasi; 2) serta soal tersebut hendaknya menggabungkan satu konsep dengan konsep lainnya.

DAFTAR PUSTAKA

- Arikunto, S. (2016). *Dasar - Dasar Evaluasi Pendidikan Edisi 2*. Jakarta: Bumi Aksara.
- Bialik, M., & Fadel, C. (2015). *Mathematics for 21st Century: What Should Students Learn*. Boston: Central Curriculum Redesign.
- Kunandar. (2013). *Penilaian Autentik Kurikulum 2013*. Jakarta: PT Grafindo Persada.
- Jakwerth, P. M., Stancavage, F. B., Reed, E. D., & Research, A. I. (1999). *An Investigation of Why Students Do*

Not Respond to Questions. Washington, D.C.: NAEP Validity Studies Panel.

- Lewy, Zulkardi, & Aisyah, N. (2009). Pengembangan Soal untuk Mengukur Kemampuan Berpikir Tingkat Tinggi Pokok Bahasan Barisan dan deret Bilangan di Kelas IX Akselerasi SMP Xaverius Maria Palembang. *Jurnal Pendidikan Matematika* , 14-28.
- Sagala, P. N., & Andriani, A. (2019). Development of Higher-Order Thinking Skills (HOTS) Questions of Probability Theory Subject Based on Bloom's Taxonomy. *Journal of Physics*, doi:10.1088/1742-6596/1188/1/012025.
- Tessmer, M. (1994). Formative Evaluation Alternatives. *Performance Improvement Quarterly*, 7(1), 3-18.
- Widiyoko, E. P. (2012). *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta: Pustaka Belajar
- Zulaiha, R. (2008). *Analisis Soal Secara Manual*. Jakarta: PUSPENDIK.
- Zulkardi. (2006). *Formative Evaluation: What, why, when, and how*. Retrived Desember 4,2019, from <https://www.oocities.org/zulkardi/books.html>