

Analisis Instrumen Penilaian Berbasis *Higher Order Thinking Skills* pada Mata Kuliah Matematika Lanjut

Sari Saraswati^{1*}, Iesyah Rodliyah², Novia Dwi Rahmawati³

^{1,2,3}Pendidikan Matematika, Universitas Hasyim Asy'ari Tebuireng, Indonesia
*sarisaraswati7@gmail.com

Abstrak

Penelitian ini bertujuan untuk mengetahui kualitas instrumen penilaian berbasis *Higher Order Thinking Skills* (HOTS) berdasarkan analisis kualitatif dan kuantitatif. Adapun jenis penelitian merupakan deskriptif kuantitatif. Penelitian ini melibatkan 32 mahasiswa Pendidikan Matematika. Instrumen penelitian berupa lembar telaah soal dan 16 item soal bentuk pilihan ganda berbasis HOTS berdasarkan pada taksonomi bloom. Telaah secara kualitatif meliputi aspek isi/materi, konstruksi, serta bahasa. Sedangkan analisis secara kuantitatif berdasarkan pendekatan Teori Tes Klasik dengan bantuan program ANATES versi 4.0.9. Aspek yang di telaah secara kuantitatif meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan fungsi pengecoh. Hasil penelitian menunjukkan bahwa 1) instrumen mencapai kriteria valid dari hasil telaah kualitatif dan kuantitatif, 2) reliabilitas butir soal mencapai 0,87 termasuk kriteria sangat tinggi, 3) daya pembeda minimal sebesar 22,22%, 4) tingkat kesukaran soal meliputi 25% mudah, 31,25% sedang, sedangkan 43,75% sukar, serta 5) semua pengecoh berfungsi dengan baik. Instrumen penilaian berbasis HOTS ini dapat diterapkan sebagai alat evaluasi dalam membiasakan mahasiswa berpikir tingkat tinggi.

Kata kunci: HOTS, Instrumen, Matematika Lanjut, Soal.

Abstract

This study aims to determine the quality of the assessment instrument based on Higher Order Thinking Skills (HOTS) based on qualitative and quantitative analysis. The type of research is descriptive quantitative. This study involved 32 Mathematics Education students. The research instrument was in the form of a question review sheet and 16 multiple-choice items based on HOTS based on Bloom's taxonomy. The qualitative study covers aspects of content/material, construction, and language. While the quantitative analysis is based on the Classical Test Theory approach with the help of the ANATES version 4.0.9 program. Aspects that are analyzed quantitatively include validity, reliability, level of difficulty, discriminatory power, and distractor functions. The results showed that 1) the instrument reached the valid criteria from the results of qualitative and quantitative studies, 2) the reliability of the items reached 0.87 including very high criteria, 3) the discriminatory power of at least 22.22%, 4) the level of difficulty of the questions included 25% easy, 31.25% moderate, while 43.75% difficult, and 5) all distractors work well. This HOTS-based assessment instrument can be applied as an evaluation tool in familiarizing students with higher-order thinking.

Keywords: HOTS, Instrument, Advanced Mathematics, Assessment.

Received: Mei 17, 2021/ Accepted: Juni 20, 2021/ Published Online: Juli 29, 2021

PENDAHULUAN

Pendidikan abad 21 menuntut dosen untuk selalu melakukan inovasi dalam berbagai aspek yang terkait dengan pembelajaran. Berbagai upaya dilakukan untuk mencetak lulusan yang berkualitas diantaranya melalui inovasi metode perkuliahan, model pembelajaran, serta evaluasi pembelajaran. Sedangkan kemampuan yang harus dimiliki mahasiswa sesuai abad ini adalah keterampilan 4C (*Communication, Collaboration, Critical thinking, and Problem Solving*) yang juga dikenal dengan kemampuan *Higher Order Thinking Skills* (HOTS). Hal ini sejalan dengan Suherman et al. (2020) yang mengungkapkan bahwa HOTS merupakan aspek penting dalam pembelajaran. Maka dari itu, hal penting yang perlu dipersiapkan untuk generasi muda saat ini salah satunya yaitu membekali dengan kemampuan untuk berpikir kritis, kreatif, dan terampil mengambil keputusan dalam penyelesaian masalah.

Keterampilan berpikir yang termasuk dalam tingkat tinggi bukan saja pada level mengingat namun lebih diperlukan suatu cara berpikir tingkat tinggi. Berdasarkan Taksonomi Bloom hasil revisi Anderson & Krathwohl (2001) bahwa *Higher Order Thinking Skills* (HOTS) meliputi tingkatan menganalisis (C4), mengevaluasi (C5), dan mencipta (C6). Oleh sebab itu, keterampilan berpikir tingkat tinggi cukup penting dibiasakan untuk peserta didik agar mempunyai kemampuan berpikir yang baik dan dapat memecahkan masalah serta menemukan solusi dari permasalahan sehari-hari (Oktarina & Armariena, 2020).

Kemampuan berpikir kritis mahasiswa masih sangat rendah. Kurniati et al. (2015) menjelaskan bahwa faktanya kemampuan berpikir matematis dan kritis tergolong rendah. Oleh karena itu, upaya yang dapat diterapkan dalam melatih mahasiswa berpikir kritis serta kreatif diantaranya dengan mengembangkan alat evaluasi pembelajaran berbasis HOTS. Alat evaluasi yang mengacu pada metode berpikir yang tinggi dapat mendukung peserta didik dalam menguasai keterampilan ini (Shidiq et al., 2015). Evaluasi memegang peranan yang cukup penting dalam pembelajaran. Tujuan evaluasi diantaranya adalah memberikan informasi tentang kemajuan belajar, kemampuan peserta didik, serta mengetahui sejauh mana tingkat ketercapaian hasil belajar. Evaluasi dalam pembelajaran dapat dilakukan dengan tes atau non tes. Instrumen penilaian yang sering digunakan dalam pembelajaran adalah tes. Sa'idah et al. (2019) memaparkan bahwa keterampilan berpikir tingkat tinggi dapat ditumbuhkan dengan tes.

Amelia (2016) menjelaskan bahwa instrumen penilaian sebagai suatu alat yang digunakan dalam evaluasi pembelajaran. Dalam melakukan evaluasi pembelajaran dibutuhkan instrumen penilaian yang memiliki kualitas valid dan dapat mengukur sejauh mana kapasitas peserta didik dengan tepat. Instrumen penilaian yang disebut juga dengan tes harus dirancang

dengan berkualitas dan sesuai dengan fungsinya. Sejalan dengan Nuryadi & Khuzaini (2016) yang menjelaskan bahwa prinsip pengembangan instrumen penilaian diantaranya harus dirancang sesuai dengan fungsinya agar diperoleh hasil sesuai dengan yang diinginkan dan memiliki reliabilitas yang dapat diandalkan. Konstruksi tes yang tidak memperhatikan kualitasnya kemungkinan tidak dapat digunakan untuk mengungkapkan dengan tepat kemampuan peserta didik (Amelia, 2016). Salah satu bentuk instrumen penilaian atau tes yang dapat dikembangkan adalah pilihan ganda.

Mengacu pada Mardapi (2008) bahwa tes yang berbentuk *multiple choice* (pilihan ganda) merupakan tes yang cara menjawabnya dengan menentukan satu jawaban dari beberapa alternatif jawaban yang ada. Dalam tes yang bentuknya pilihan ganda, konstruksi soal memuat pernyataan atau pokok soal, serta opsi jawaban yang meliputi pengecoh dan kunci jawaban. Kelebihan soal pilihan ganda adalah dapat mencakup materi lebih luas dan durasi penilaian yang tidak lama. Akibatnya ketidaktepatan hasil penilaian dapat dikurangi (Warju et al., 2020); (Tangianu et al., 2018); (Brown, 2004).

Soal berbentuk pilihan ganda juga bisa dipakai dalam mengukur kemampuan berpikir tingkat tinggi. Meskipun merancang instrumen penilaian HOTS lebih mudah disajikan sebagai bentuk pertanyaan terbuka atau tugas, contohnya soal esai dan tugas kinerja, namun soal pilihan ganda yang memiliki opsi kompleks bisa dikembangkan sebagai bentuk instrumen penilaian HOTS sepanjang pilihan kompleks tersebut berisi stimulus yang berasal dari situasi/permasalahan kontekstual (Kemendikbud, 2019).

Instrumen penilaian harus mempunyai kualitas yang layak untuk difungsikan dalam mengukur kompetensi peserta didik (Sa'idah et al., 2019). Kevalidan soal pilihan ganda pada umumnya memuat butir soal yang berkualitas. Kualitas item soal bisa diketahui dari hasil uji validitas, uji reliabilitas, bagaimana tingkat kesukaran soal, besarnya daya pembeda, serta bagaimana efektivitas pengecohnya (Friatma & Anhar, 2019); (Jannah et al., 2021). Untuk mengetahui kualitas tersebut maka perlu melakukan analisis pada soal yang diujikan. Berdasarkan analisis butir soal, penyusun soal dapat memperoleh umpan balik terhadap soal yang diujikan. Apabila soal yang diujikan termasuk kriteria layak maka dapat diterima atau bisa digunakan, jika kelayakannya kurang maka dapat dilakukan perbaikan, dan yang tidak layak dapat dibuang atau dihilangkan. Oleh karena itu, menganalisis item soal sangat penting karena bertujuan untuk mengetahui kehandalan dan validitas dari tes yang telah disusun (Quaigrain & Arhin, 2017).

Instrumen penilaian atau soal yang baik ditentukan dari: 1) validitas butir soal, menunjukkan bagaimana kecermatan dan ketepatan sebuah instrumen penilaian atau tes berfungsi sesuai alat ukurnya (Azwar, [1988](#)). Menurut Masriyah ([1999](#)); Arikunto, ([2013](#)) bahwa item soal dikategorikan valid jika memperoleh bantuan yang besar terhadap total skornya. Tinggi atau rendahnya skor total dipengaruhi oleh skor tiap tes. 2) Reliabilitas, yang dimaksud adalah ukuran suatu tes atau instrumen yang dinyatakan sebagai besarnya tingkat keajegan atau kekonsistenan (Arifin, [2013](#)); (Jihad & Haris, [2012](#)). 3) Tingkat kesukaran, berhubungan dengan sebaran soal yang mudah, sedang, dan sukar. Brown ([2004](#)) menjelaskan bahwa tingkat kesulitan mengacu pada tingkat kemudahan yang dimiliki item pilihan ganda. Memastikan bahwa suatu butir tes memiliki tingkat kesulitan yang sesuai adalah penting sebab bisa mempengaruhi kemampuan tes yang signifikan dalam mengukur kemampuan sebenarnya dari peserta tes. Sedangkan 4) daya beda, ukuran suatu alat evaluasi mampu memisahkan kelompok peserta tes yang kemampuannya tinggi dan rendah (Arikunto, [2013](#)). Semakin besar indeks daya pembeda, berarti butir soal tersebut makin dapat diketahui perbedaan antara peserta didik yang memahami serta kurang memahami materi/kompetensi. 5) Efektivitas pengecoh, pengecoh dibuat dengan tujuan untuk menyesatkan peserta tes yang belum paham dengan materi yang diujikan. Arikunto ([2013](#)) mengungkapkan bahwa suatu pengecoh disebut berfungsi baik apabila mampu menarik peserta tes yang kurang menguasai materi untuk memilih pengecoh tersebut.

Kriteria keberhasilan analisis butir soal dapat diperhatikan dari hasil penelitian terdahulu. Hasil penelitian Arifin & Retnawati ([2015](#)) diperoleh bahwa instrumen HOTS matematika kelas X yang memuat 60 butir soal dikategorikan valid. Kedua paket soal A dan B memperoleh hasil yang reliabel. Soal paket A termasuk dalam kategori soal mudah sebesar 6,67%, sedang 46,67%, dan sulit 46,67%. Sedangkan pada paket B, kategori soal mudah sebesar 12,50%, sebesar 12,50% sedang, dan 75% sulit. Sedangkan daya beda paket A menunjukkan 53,33% soal diterima, 26,67% direvisi, dan 20% ditolak. Pada soal paket B, soal yang dapat dipakai sebesar 37,50%, sebesar 18,75% direvisi, dan ditolak sebesar 43,75%. Penelitian lain dilakukan Sa'idah et al. ([2019](#)) bahwa instrumen penilaian yang diuji oleh ahli dari aspek kualitatif menghasilkan kriteria valid. Sedangkan telaah butir item *higher order thinking* menghasilkan reliabilitas kurang baik sebesar 0,596. Analisis daya beda, tingkat kesulitan, dan efektifitas distraktor menghasilkan kategori baik sebesar 50%.

Berdasarkan paparan diatas maka analisis butir soal sangat dibutuhkan untuk mendapatkan instrumen penilaian yang berkualitas. Oleh karena itu penelitian yang dilakukan ini bermaksud untuk menentukan bagaimana kualitas instrumen penilaian berbasis HOTS

pada mata kuliah matematika lanjut secara kualitatif dan kuantitatif agar instrumen yang disusun dapat diterapkan dalam mengukur kemampuan berpikir mahasiswa secara tepat.

METODE

Metode penelitian yang diterapkan merupakan deskriptif kuantitatif yang bertujuan untuk mengetahui kualitas instrumen penilaian berbasis HOTS berbentuk pilihan ganda yang ditelaah secara kualitatif dan kuantitatif. Sedangkan prosedur dalam penelitian ini yaitu *design research* tipe *development study* yang menekankan pada evaluasi formatif (Plomp & Nieveen, [2013](#)). Subjek penelitian sebanyak 32 mahasiswa S-1 Prodi Pendidikan Matematika Unhasy yang telah menempuh mata kuliah matematika lanjut. Teknik pengambilan data menggunakan instrumen lembar telaah soal HOTS dan soal berbentuk pilihan ganda. Instrumen penilaian berbasis HOTS yang diujikan sebanyak 16 butir.

Data dianalisis secara kualitatif serta kuantitatif terhadap instrumen penilaian berbasis HOTS. Analisis secara kualitatif dilakukan oleh *expert judgment* dengan cara menelaah setiap item soal berdasarkan kaidah/aturan penulisan dan konstruksi butir soal. Analisis kualitatif terhadap item soal HOTS bentuk pilihan ganda meliputi aspek isi, konstruksi, serta bahasa. Sedangkan telaah kuantitatif mengacu pada pendekatan teori tes klasik dimana analisis butir soal berdasarkan hasil jawaban peserta tes. Analisis kuantitatif dari instrumen penilaian memakai aplikasi ANATES versi 4.0.9. Arif ([2014](#)) mengemukakan bahwa program Anates efektif untuk membantu menganalisis butir soal dalam segi pengoperasian, waktu perhitungan, keakuratan hasil perhitungan dan memahami *output* yang dihasilkan. Aspek yang ditelaah secara kuantitatif meliputi validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas distraktor.

Suatu validitas dari soal mempunyai kualitas tinggi apabila skor pada item soal tersebut sejajar dengan skor totalnya. Kesejajaran dimaksud sebagai korelasi, oleh karena itu untuk menentukan besar validitas terhadap item soal maka diterapkan rumus korelasi *product moment pearson*. Ini diperoleh dengan menentukan korelasi antara skor yang diperoleh mahasiswa pada setiap item dan total skor yang diraih. Nilai validitas item berdasarkan pada Jihad & Haris ([2012](#)) dapat dilihat pada [Tabel 1](#):

Tabel 1. Kriteria nilai validitas soal

Nilai r_{xy}	Interpretasi
$r_{xy} \leq 0,20$	Sangat rendah
$0,20 < r_{xy} \leq 0,40$	Rendah
$0,40 < r_{xy} \leq 0,60$	Sedang
$0,60 < r_{xy} \leq 0,80$	Tinggi
$0,80 < r_{xy} \leq 1,00$	Sangat tinggi

Dalam mengukur tingkat keajegan atau reliabilitas soal ini dihitung menggunakan rumus *Alpha Cronbach*. Adapun kriteria indeks reliabilitas butir soal mengacu pada pendapat Guilford yang dilihat pada [Tabel 2](#) berikut.

Tabel 2. Kriteria nilai reliabilitas soal

Nilai r	Interpretasi nilai r
$r \leq 0,20$	Sangat rendah
$0,20 < r \leq 0,40$	Rendah
$0,40 < r \leq 0,60$	Sedang
$0,60 < r \leq 0,80$	Tinggi
$0,80 < r \leq 1,00$	Sangat tinggi

(Jihad & Haris, [2012](#))

Daya pembeda ditentukan untuk mengetahui kelompok mahasiswa kemampuan yang atas serta bawah. Daya beda dari setiap item soal dihitung dengan rumus, yaitu:

$$DP = \frac{(WL-WH)}{n}$$

Dengan:

DP : Daya beda

WL : Total peserta tes yang tidak menjawab benar dari kelompok bawah

WH : Total peserta tes yang menjawab salah dari kelompok atas

n : $27\% \times$ jumlah peserta tes

(Arifin, [2013](#))

Sedangkan untuk mendefinisikan koefisien daya pembeda tersebut, maka diterapkan kriteria dari hasil pengembangan oleh Ebel pada [Tabel 3](#), yaitu:

Tabel 3. Kriteria daya beda tes

Daya Pembeda (DP)	Interpretasi nilai DP
$DP \leq 0,19$	Item jelek, ditolak atau diperbaiki dengan revisi
$0,20 - 0,29$	Minimun, biasanya perlu diperbaiki
$0,30 - 0,39$	Cukup baik, tapi mungkin perlu diperbaiki
$DP \geq 0,40$	Sangat baik

(Arifin, [2013](#))

Perhitungan tingkat kesukaran soal merupakan ukuran besarnya derajat kesukaran butir tes. Indeks kesukaran diartikan sebagai taraf kesukaran atau kemudahan suatu soal. Interpretasi indeks kesukaran mengacu pada [Tabel 4](#) berikut (Arikunto, 2013):

Tabel 4. Kriteria tingkat kesukaran butir soal

Tingkat Kesukaran (TK)	Interpretasi nilai TK
0,00 – 0,30	Sukar
0,31 – 0,70	Sedang
0,71 – 1,00	Mudah

Pilihan jawaban dari instrumen penilaian HOTS yang diujikan terdiri dari 4 opsi. Fungsi distraktor dapat digunakan dengan baik apabila minimal 5% peserta tes memilihnya (Purwanto, 2014). Apabila distraktor tidak berfungsi maka perlu dilakukan revisi atau mengganti pengecoh tersebut.

HASIL PENELITIAN

Instrumen penilaian berbasis *Higher Order Thinking Skills* (HOTS) yang dikaji pada penelitian ini adalah hasil suatu pengembangan soal HOTS pada mata kuliah matematika lanjut. Materi yang dipelajari mahasiswa dalam mata kuliah ini adalah konsep matematika sekolah serta mereka dituntut dapat menyelesaikan soal-soal tingkat SMA. Hal ini yang mendasari bahwa soal yang diberikan kepada mahasiswa mengacu pada materi matematika sekolah sesuai kompetensi dasar yang digunakan pada jenjang SMA.

Pada tahap awal, instrumen penilaian berbasis HOTS dianalisis secara kualitatif oleh *expert judges* yaitu seorang dosen ahli pada bidang Pendidikan Matematika dengan memperhatikan aspek materi, konstruksi, dan bahasa. Dosen ahli melakukan telaah pada kisi-kisi, item soal, rubrik penskoran, dan kartu soal. Kegiatan validasi ini dilakukan dengan memberikan lembar telaah agar diperoleh saran dan masukan terkait instrumen penilaian HOTS yang akan diujikan kepada mahasiswa. Saran dan masukan tersebut dijadikan sebagai acuan untuk merevisi soal dan menyatakan bahwa instrumen penilaian HOTS valid. Pada [Tabel 5](#) berikut adalah hasil telaah yang dilakukan dosen ahli.

Tabel 5. Rekap hasil telaah kualitatif oleh *expert judgment*

Aspek	Valid	Tidak Valid	Nomor Soal yang perlu Direvisi
Materi	99,3 %	1 item	3
Konstruksi	97,2 %	4 item	6, 10, 11, 16
Bahasa	96,9 %	2 item	4, 8

Setelah instrumen penilaian HOTS dinyatakan valid oleh *expert judgment* dari segi materi, konstruksi, dan bahasa maka selanjutnya instrumen penilaian HOTS tersebut diujikan kepada 32 mahasiswa untuk memperoleh hasil jawaban yang selanjutnya hasil jawaban tersebut dianalisis dengan bantuan aplikasi ANATES versi 4.0.9. Dalam segi kuantitatif dilakukan analisis terhadap validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan kualitas pengecoh.

Tingkat keajegan instrumen penilaian berbasis HOTS yang berbentuk pilihan ganda dilihat dari perhitungan nilai reliabilitas butir soal. Adapun rekap reliabilitas butir soal bisa diperhatikan pada [Tabel 6](#) berikut.

Tabel 6. Rekap nilai reliabilitas tes

Kriteria	Nilai
Jumlah item	16
Skor rata-rata	3,88
Simpangan Baku	6,53
Korelasi XY	0,77
Reliabilitas Tes	0,87

Sedangkan kevalidan tiap item soal dan kemampuannya dalam menunjukkan perbedaan antara mahasiswa berkemampuan tinggi serta rendah dapat dilihat dari hasil perhitungan menggunakan Anates. Adapun hasil validitas serta daya beda dari 16 item soal ditunjukkan pada [Tabel 7](#) berikut.

Tabel 7. Rekap hasil validitas dan daya pembeda soal HOTS

Nomor Soal	Level Soal	Korelasi	Kriteria Validitas	Daya Pembeda (%)	Kriteria
1	C4	0,748	Tinggi	88,89	Sangat Baik
2	C6	0,770	Tinggi	88,89	Sangat Baik
3	C6	0,658	Tinggi	88,89	Sangat Baik
4	C5	0,715	Tinggi	44,44	Sangat Baik
5	C5	0,477	Sedang	44,44	Sangat Baik
6	C5	0,310	Sedang	22,22	Minimum
7	C5	0,513	Sedang	66,67	Sangat Baik
8	C5	0,577	Sedang	66,67	Sangat Baik
9	C5	0,566	Sedang	33,33	Cukup Baik
10	C6	0,548	Sedang	55,56	Sangat Baik
11	C4	0,516	Sedang	33,33	Cukup Baik
12	C4	0,639	Tinggi	44,44	Sangat Baik
13	C5	0,692	Tinggi	100,00	Sangat Baik
14	C5	0,582	Sedang	55,56	Sangat Baik
15	C5	0,525	Sedang	44,44	Sangat Baik
16	C6	0,514	Sedang	55,56	Sangat Baik

Selain validitas dan daya pembeda, berdasarkan hasil analisis butir soal menggunakan Anates diperoleh hasil tingkat kesukaran dan fungsi pengecoh dari 16 item soal HOTS. Tingkat kesukaran terbagi dalam kategori soal mudah, sedang, dan sukar tersebut disajikan pada [Tabel 8](#) sedangkan [Tabel 9](#) menunjukkan hasil analisis fungsi pengecoh.

Tabel 8. Hasil analisis tingkat kesukaran

Kategori	No. Soal	Persentase
Mudah	5, 6, 7, 8	25 %
Sedang	1, 2, 3, 13, 16	31,25 %
Sukar	4, 9, 10, 11, 12, 14, 15	43,75 %

Tabel 9. Hasil analisis fungsi pengecoh

Nomor Soal	Pilihan Jawaban (%)			
	A	B	C	D
1	15,6	25	40,6**	18,8
2	15,6	18,8	34,4	31,2**
3	12,5	12,5	15,6	59,4**
4	28,1	12,5**	37,5	21,9
5	9,4	9,4	71,9**	9,4
6	0	0	6,3	93,8**
7	6,3	71,9**	9,4	12,5
8	6,3	81,3**	6,3	6,3
9	12,5**	34,4	34,4	18,8
10	25	34,4	15,6**	25
11	34,4	28,1	25	12,5**
12	28,1	34,4	15,6**	21,3
13	12,5	18,8	6,3	62,5**
14	18,8**	40,7	18,8	21,3
15	31,3	15,6**	46,9	6,3
16	25	21,3	37,5**	15,6

Keterangan: Kunci jawaban (**)

Menurut hasil dari analisis kuantitatif pada item soal maka instrumen penilaian berbasis HOTS yang terdiri 16 item soal pilihan ganda dirangkum sesuai [Tabel 10](#) berikut:

Tabel 10. Rekap hasil analisis butir soal

Kategori	Nomor Soal	Jumlah
Diterima	1, 2, 3, 4, 5, 7, 8, 10, 12, 13, 14, 15, 16	81,25 %
Direvisi	6, 9, 11	18,75 %
Ditolak	-	0 %

PEMBAHASAN

Berdasarkan hasil telaah secara kualitatif oleh *expert judgment* diperoleh bahwa 16 butir soal telah memenuhi kriteria valid. Kevalidan dari aspek materi mencapai 100%, aspek konstruksi sebesar 97,2%, dan 96,9% dari segi bahasa. Berdasarkan masukan dari *expert judgment* terdapat beberapa soal yang perlu direvisi dari segi bahasa yaitu nomor 4 dan 8. Sedangkan revisi dari segi konstruksi dilakukan pada nomor 6, 10, 11, dan 16. Sedangkan pada soal nomor 3 perlu dilakukan revisi terhadap aspek materi. Soal nomor 3 merupakan soal dengan level C6 (mencipta) dengan konteks peluruhan zat radioaktif. Berdasarkan saran masukan *expert judgment* ditambahkan stimulus terutama pada tetapan zat radioaktif. Kemendikbud (2019) mengungkapkan bahwa salah satu karakteristik soal HOTS adalah berbasis permasalahan kontekstual dimana kontekstual ini meliputi *relating, experiencing, applying, communicating, dan trasnfering*. Pada konteks peluruhan zat radioaktif ini, mahasiswa dituntut untuk dapat mengkomunikasikan model matematika berdasarkan konteks soal yang diberikan. Selain itu, revisi juga dilakukan dengan menambahkan petunjuk pengerjaan soal dengan jelas untuk memilih salah satu jawaban, karena dalam soal HOTS ini mahasiswa hanya dapat memilih salah satu jawaban saja. Selanjutnya, soal yang telah valid kemudian diujikan kepada mahasiswa untuk mendapatkan hasil jawaban.

Tingkat kesukaran terhadap 16 butir soal berbasis HOTS menunjukkan bahwa distribusi soal tidak proporsional. Soal dikatakan baik apabila banyaknya item terbagi dalam perbandingan tingkat kesukaran yang seimbang/proporsional (Arifin, 2013:266). Berdasarkan tabel 8 diperoleh bahwa dari 16 soal berbasis HOTS, 25% merupakan soal mudah, 31,25 % soal sedang, dan 43,75 % soal sukar. Sudjana (2017) mengungkapkan bahwa proporsi tingkat kesukaran dapat menggunakan perbandingan 3-4-3 yang berrati bahwa 30% mudah, sedang sebanyak 40%, dan sulit 30%. Oleh karena itu, hasil analisis tingkat kesukaran instrumen penilaian berbasis HOTS ini tidak proporsional dimana soal lebih banyak yang sukar dan sedikit yang mudah. Hasil analisis diperoleh bahwa soal sukar lebih banyak dibandingkan soal sedang.

Telaah terhadap daya pembeda menghasilkan indeks dengan kriteria sangat baik sebanyak 13 item, 2 item cukup baik, dan 1 soal mencapai kriteria minimum (lihat tabel 7). Berdasarkan pada tabel 3, maka soal yang mencapai kriteria minimum perlu diperbaiki. Soal-soal yang telah mencapai kriteria cukup baik atau sangat baik berarti telah mampu mengklasifikasikan mahasiswa antara kemampuan dari kelompok atas dan bawah. Sedangkan pada butir soal yang mencapai kategori minimum, maka berarti soal tersebut tidak dapat mengkategorikan antara mahasiswa yang kemampuannya tinggi dan rendah. Hal ini sejalan

dengan Warju et al. (2020) bahwa apabila kriteria penyusunan soal tidak terpenuhi berarti soal tersebut belum bisa mengelompokkan peserta didik dalam kelompok atas dan bawah dimana hasil penelitiannya memperoleh nilai daya beda sebesar 10% baik sekali, 42% baik, 14 % baik, 28% cukup baik, dan 6% buruk. Indeks daya pembeda dengan kategori minimum ini dicapai dari hasil analisis instrumen berbasis HOTS pada item soal nomor 6. Adapun soal tersebut dapat diperhatikan berdasarkan [Gambar 1](#) berikut.

Butir Soal 6

Suara paling keras yang dapat di dengar oleh orang sehat tanpa merusak gendang telinga memiliki intensitas 1 triliun (1.000.000.000.000) kali lebih kuat daripada suara paling rendah yang bisa didengar. Berikut ini adalah tabel intensitas bunyi.

Besar Intensitas Bunyi (W/m^2)	Sumber Bunyi
$\pm 1,0 \times 10^{-12}$	Ambang batas bawah pendengaran manusia
$\pm 5,2 \times 10^{-10}$	Suara bisik-bisik
$\pm 3,2 \times 10^{-6}$	Percakapan normal
$\pm 0,5 \times 10^{-4}$	Lalu lintas padat
$\pm 8,3 \times 10^{-2}$	Pesawat jet lepas landas

Tiga mahasiswa Universitas Diponegoro berhasil mengubah kebisingan di bandara menjadi energi listrik. Cara kerja alat tersebut adalah kebisingan ditangkap oleh reflektor parabola kemudian difokuskan pada sistem transduser yang kemudian memicu terjadinya induksi elektromagnetik. Dari induksi elektromagnetik inilah dihasilkan energi listrik. Alat ini mampu menghasilkan tegangan mencapai 11,14 volt pada intensitas suara dari kebisingan sebesar 108,4 dB. Apabila ingin menghasilkan tegangan sebesar $\pm 9,18$ volt, maka sumber suara yang dapat diubah adalah ...

A. Batas bawah pendengana manusia
 B. Suara bisik-bisik
 C. Percakapan normal
 D. Lalu lintas padat



Gambar 1. Item soal no.6

Permasalahan nomor 6 dengan level C5 termasuk soal yang mudah. Mahasiswa dapat menjawab soal dengan menalar informasi yang ada dalam soal. Namun, mereka tidak dapat menunjukkan cara yang tepat dalam menemukan jawabannya. Disamping itu, berdasarkan pilihan jawaban A, B, dan C dari soal ini menunjukkan kriteria buruk. Berdasarkan tabel 9 terlihat bahwa opsi C hanya dipilih oleh 2 peserta tes yaitu sebesar 6,3 %, sedangkan pilihan A dan B tidak ada yang memilih atau 0%. Oleh karena itu, pada butir soal ini perlu dilakukan revisi. Perbaikan dapat dilakukan terhadap konstruksi butir soal atau mengubah konstruksi pilihan jawaban yang ada agar kualitas pengecoh baik. Adapun fungsi pengecoh dari masing-masing item bisa diperhatikan dari tabel 9. Selain itu, pada soal nomor 15 distraktor tidak bekerja secara signifikan. Hal ini disebabkan karena perbedaan distraktor lainnya lebih mencolok sehingga menimbulkan ketidaktertarikan dengan pengecoh lainnya yang hanya mencapai 6,3%.

Hasil analisis terhadap fungsi pengecoh menunjukkan bahwa semua pengecoh telah berfungsi sebab dipilih melebihi 5 % oleh peserta tes, kecuali pada butir soal 6. Fungsi pengecoh dalam soal pilihan ganda digunakan untuk mengecoh peserta tes sehingga merasa ragu-ragu atau bimbang dalam memilih pengecoh sebagai jawaban benar. Pengecoh yang tidak dipilih menunjukkan soal tersebut jelek dan perlu diperbaiki atau dirombak. Fungsi pengecoh ini memberikan dampak pada tingkat kesukaran butir soal. Hal ini sejalan dengan

penelitian Iskandar & Rizal (2017), jika ada satu atau dua distraktor yang tidak berfungsi maka tingkat kesukaran menurun karena kesempatan peserta tes menjawab benar semakin meningkat.

Berdasarkan telaah terhadap butir soal secara kuantitatif, maka 16 butir instrumen penilaian berbasis HOTS dinyatakan valid dan reliabel. Berdasarkan tabel 10, dari 16 item ini tidak ada butir soal yang ditolak, akan tetapi soal perlu dilakukan revisi sebanyak 18,75% sedangkan sisanya soal tersebut diterima. Revisi pada butir soal lebih banyak pada aspek konstruksi dan bahasa agar instrumen penilaian ini menjadi lebih berkualitas.

SIMPULAN

Berdasarkan analisis butir soal berbasis HOTS sebanyak 16 butir diperoleh bahwa soal mencapai kriteria valid secara kualitatif. Sedangkan analisis secara kuantitatif diperoleh bahwa soal dengan kriteria kevalidan tinggi sebanyak 6 item, dan 10 item mencapai kategori sedang. Daya pembeda soal mencapai kategori sangat baik sebanyak 13 item, cukup baik 2 item, dan minimum sebanyak 1 item. Distribusi tingkat kesukaran dari item yang mudah, sedang, serta sulit masing-masing 25%, 31,25%, dan 43,75%. Sedangkan masing-masing pilihan jawaban rata-rata berfungsi dengan baik karena mencapai kategori diatas 5%. Dari 16 item instrumen penilaian berbasis HOTS disimpulkan bahwa 81,25% diterima dan 18,75% direvisi.

REFERENSI

- Amelia, M. A. (2016). Analisis Soal Tes Hasil Belajar High Order Thinking Skills (HOTS) Matematika Materi Pecahan Untuk Kelas 5 Sekolah Dasar. *Jurnal Penelitian (Edisi Khusus PGSD)*, 20(2), 123–131.
- Anderson, L. W., & Krathwohl, D. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Addison Wesley Longman, Inc.
- Arif, M. (2014). Penerapan Aplikasi Anates Bentuk Soal Pilihan Ganda. *Jurnal Ilmiah Educativ*, 1(1), 1–9. <https://doi.org/10.21107/edutic.v1i1.398>
- Arifin, Zaenal, & Retnawati, H. (2015). Analisis Instrumen Pengukur Higher Order Thinking Skills (HOTS) Matematika Siswa SMA. *Seminar Nasional Matematika Dan Pendidikan Matematika UNY*, 20, 783–790.
- Arifin, Zainal. (2013). *Evaluasi pembelajaran: Prinsip, Teknik, Prosedur*. Bandung: PT Remaja Rosdakarya.
- Arikunto, S. (2013). *Dasar-dasar Evaluasi Pembelajaran*. Jakarta: Bumi Aksara.
- Azwar, S. (1988). *Sikap Manusia Teori dan Pengukurannya*. Yogyakarta: Liberty.
- Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. California:

Longman.

- Friatma, A., & Anhar, A. (2019). Analysis of Validity, Reliability, Discrimination, Difficulty and Distraction Effectiveness in Learning Assessment. *Journal of Physics: Conference Series*, 1387(1), 1–6. <https://doi.org/10.1088/1742-6596/1387/1/012063>
- Iskandar, A., & Rizal, M. (2017). Analisis Kualitas Soal di Perguruan Tinggi Berbasis Aplikasi TAP. *Jurnal Penelitian dan Evaluasi Pendidikan*, 21(2), 12–23. <http://dx.doi.org/10.21831/pep.v22i1.15609>
- Jannah, R., Hidayat, D. N., Husna, N., & Khasbani, I. (2021). An item analysis on multiple-choice questions: a case of a junior high school English try-out test in Indonesia. *Leksika: Jurnal Bahasa, Sastra dan Pengajarannya*, 15(1), 9–17. <https://doi.org/10.30595/lks.v15i1.8768>
- Jihad, A., & Haris, A. (2012). *Evaluasi Pembelajaran*. Yogyakarta: Multi Pressindo.
- Kemendikbud. (2019). *Buku Penilaian Berorientasi Higher Order Thinking Skills*. Jakarta: Direktorat Jenderal Guru dan Tenaga Kependidikan Kementerian Pendidikan dan Kebudayaan.
- Kurniati, K., Kusumah, Y. S., Sabandar, J., & Herman, T. (2015). Mathematical Critical Thinking Ability Through. *IndoMS. J.M.E*, 6(1), 53–62.
- Mardapi, D. (2008). *Teknik Penyusunan Instrumen Tes dan Non Tes*. Yogyakarta: Mitra Cendikia Press.
- Masriyah, M. (1999). *Validitas dan Reliabilitas*. Surabaya: Unesa University Press.
- Nuryadi, & Khuzaini, N. (2016). *Evaluasi Hasil Dan Proses Pembelajaran Matematika*. Yogyakarta: PT Leutika Nouvalitera.
- Oktarina, N., & Armariena, D. N. (2020). Analisis Soal Tipe Higher Order Thinking Skills (HOTS) pada Tes Objektif Pilihan Ganda USBN Bahasa Indonesia SD/MI. *Wahana Didaktika : Jurnal Ilmu Kependidikan*, 18(2), 146–156. <https://doi.org/10.31851/wahanadidaktika.v18i2.4374>
- Plomp, T. (SLO), & Nieveen, N. (SLO). (2013). Educational Design Research. In *Educational Design Research*. Netherlands: SLO.
- Purwanto, P. (2014). *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Belajar.
- Quaigrain, K., & Arhin, A. K. (2017). Using Reliability and Item Analysis to Evaluate a Teacher-Developed Test in Educational Measurement and Evaluation. *Cogent Education*, 4(1). <https://doi.org/10.1080/2331186X.2017.1301013>
- Sa'idah, N., Yulistianti, H. D., & Megawati, E. (2019). Analisis Instrumen Tes Higher Order Thinking Matematika SMP. *Jurnal Pendidikan Matematika*, 13(1), 41–54. <https://doi.org/10.22342/jpm.13.1.6619.41-54>
- Shidiq, A. S., Masykuri, M., & Susanti, E. (2015). Analisis Higher Order Thinking Skills (HOTS) Menggunakan Instrumen Two-Tier Multiple Choice Pada Materi Kelarutan Dan Hasil Kali Kelarutan Untuk Siswa Kelas Xi Sma N 1 Surakarta. *Prosiding Seminar Nasional Pendidikan Sains, November*, 159–166.
- Sudjana, N. (2017). *Penilaian Hasil Proses Belajar Mengajar* (2 ed.). Bandung: Remaja Rosdakarya.

- Suherman, Prananda, M. R., Proboningrum, D. I., Pratama, E. R., Laksono, P., & Amiruddin. (2020). Improving Higher Order Thinking Skills (HOTS) with Project Based Learning (PjBL) Model Assisted by Geogebra. *Journal of Physics: Conference Series*, 1467(1). <https://doi.org/10.1088/1742-6596/1467/1/012027>
- Tangianu, F., Mazzone, A., Berti, F., Pinna, G., Bortolotti, I., Colombo, F., Nozzoli, C., La Regina, M., Greco, A., Filannino, C., Silingardi, M., & Nardi, R. (2018). Are multiple-choice questions a good tool for the assessment of clinical competence in Internal Medicine? *Italian Journal of Medicine*, 12(2), 88–96. <https://doi.org/10.4081/itjm.2018.980>
- Warju, W., Ariyanto, S. R., Soeryanto, S., & Trisna, R. A. (2020). Analisis Kualitas Butir Soal Tipe HOTS pada Kompetensi Sistem Rem Di Sekolah Menengah Kejuruan. *Jurnal Pendidikan Teknologi dan Kejuruan*, 17(1), 95–104. <https://doi.org/10.23887/jptk-undiksha.v17i1.22914>