

KARAKTERISTIK SOAL *HIGHER ORDER THINKING SKILLS* (HOTS) MATERI DAMPAK PENYALAHGUNAAN PSIKOTROPIKA UNTUK SMA

CHARACTERISTIC OF HIGHER ORDER THINKING SKILLS (HOTS) QUESTIONS IN IMPACT OF PSYCOTROPIC TOPIC FOR SENIOR HIGH SCHOOL

Afrida Ning'um Wardani

Pendidikan Biologi, FMIPA, Universitas Negeri Surabaya
Jalan Ketintang Gedung C3 Lt. 2 Surabaya 60231
E-mail : afriwardani@mhs.unesa.ac.id

Muslimin Ibrahim

Pendidikan Biologi, FMIPA, Universitas Negeri Surabaya
Jalan Ketintang Gedung C3 Lt. 2 Surabaya 60231
E-mail : musliminibrahim@unesa.ac.id

Abstrak

Higher order thinking skills (HOTS) merupakan proses berpikir pada level yang lebih tinggi, tidak sekedar menghafal fakta, tetapi melibatkan aktivitas mental dalam usaha mengeksplorasi pengalaman kompleks, reflektif, dan kreatif. Ranah HOTS mencakup tingkat kognitif dan dimensi ilmu. Penelitian ini bertujuan untuk mendeskripsikan karakteristik soal HOTS materi dampak penyalahgunaan psikotropika. Karakteristik soal HOTS meliputi validitas, reliabilitas, tingkat kesukaran, dan indeks daya beda. Soal HOTS ini merupakan hasil penelitian pengembangan yang disusun dengan cara menentukan tingkat kognitif dan menjabarkan menjadi indikator, menentukan bentuk tes dan jenis HOTS yang diujicobakan. Soal HOTS divalidasi oleh ahli untuk mendapatkan validitas soal HOTS. Soal HOTS diujicobakan untuk mendapatkan reliabilitas dan parameter soal. Teknik analisis data menggunakan teknik analisis secara deskriptif yang meliputi analisis hasil tes HOTS untuk mengetahui nilai reliabilitas dengan menggunakan rumus *Spearman-Brown*. Hasil penelitian menunjukkan bahwa instrumen penilaian HOTS yang dikembangkan dinyatakan valid dengan nilai 4 yang termasuk dalam kategori sangat valid, serta memiliki nilai reliabilitas sebesar 0,65. Indeks tingkat kesukaran yang diperoleh heterogen yaitu, 25% mudah, 50% sedang, dan 25% sukar. Indeks daya beda menyatakan bahwa 62% soal HOTS dapat membedakan kemampuan peserta didik sehingga soal HOTS materi dampak penyalahgunaan psikotropika dinyatakan valid dan reliabel.

Kata Kunci: Karakteristik Soal HOTS, Tingkat Kesukaran, Indeks Daya Beda, Materi Dampak Penyalahgunaan Psikotropika, Reliabilitas.

Abstract

Higher order thinking skills (HOTS) is a process of thinking at a higher level. It not only memorizing facts, but also involves mental activity in an effort to explore complex, reflective, and creative experiences. The HOTS realm covers the cognitive level and dimensions of science. This study aimed to describe the characteristics of HOTS material about the Impact of Psychotropic Abuse. Characteristics of HOTS item are validity, reliability, level of difficulty, and different power indexes. This HOTS item was the result of development research compiled by determining the cognitive level and describing it as an indicator, determining the form of the test and the type of HOTS than tested. Item they were validated by experts to get the validity of HOTS item. HOTS item then tested to get the reliability and parameters of the questions. Data analysis techniques used descriptive analysis techniques which included analysis of HOTS test results to determine the reliability value using the Spearman-Brown formula. The results showed that the HOTS assessment instrument developed was declared valid with a value of 4 which was included in the highly valid category, and had a reliability value of 0,65. Difficulty index were varies : could 25% easy, 50% moderate, and 25% difficult. The difference power index showed that 62% of HOTS item can distinguish students' abilities so that the matter of HOTS impact of psychotropic abuse was declared valid and reliable.

Keywords: Characteristics of HOTS item, Difficulty Index, Different Power Index, Impact of Psychotropic Topic, Reliability.

PENDAHULUAN

Penilaian merupakan proses pengumpulan data atau informasi untuk mengukur hasil belajar peserta didik. Penilaian pada aspek kognitif dapat dilakukan dengan cara tes tertulis, tes lisan, dan penugasan sesuai dengan kompetensi yang diharapkan. Tujuan penilaian selain untuk mengetahui atau mengukur kemampuan peserta didik adalah untuk memperbaiki proses pembelajaran peserta didik, serta menyusun laporan perkembangan hasil belajar baik harian, tengah semester atau kenaikan kelas (Permendikbud, 2016).

Instrumen penilaian HOTS merupakan instrumen yang digunakan untuk mengukur kemampuan berpikir tingkat tinggi peserta didik. Berpikir tingkat tinggi adalah berpikir pada level yang lebih tinggi, tidak sekedar menghafal fakta tetapi melibatkan aktivitas mental dalam usaha mengeksplorasi pengalaman kompleks, reflektif, dan kreatif yang dilakukan secara sadar untuk mencapai tujuan, yaitu memperoleh pengetahuan yang meliputi tingkat berpikir analitis, sintesis, dan evaluatif (Wardana, 2010).

Manfaat dari penilaian HOTS yaitu meningkatkan prestasi, meningkatkan motivasi belajar, dan meningkatkan sikap positif peserta didik (Nugroho, 2018). Menurut King *et.al.* (2010) dalam melatih keterampilan berpikir tingkat tinggi pada peserta didik dapat dibiasakan dengan memberikan masalah yang tidak biasa terjadi atau mengenai kasus yang sedang terjadi saat ini sehingga peserta didik dapat menyelesaikan masalah. Kemampuan peserta didik dapat diketahui berdasarkan hasil tes.

Tes merupakan instrumen atau alat yang digunakan dalam pengukuran dan pengambilan keputusan, baik melalui uji hipotesis maupun tidak (Murtono dan Evi, 2014). Dalam sebuah tes terdapat dua jenis soal, yaitu jenis tes dengan format objektif dan tes dengan format subjektif. Tes subjektif merupakan tes yang berbentuk esai dimana peserta didik dapat menguraikan jawaban dengan menggunakan bahasanya sendiri. Tes objektif merupakan bentuk tes yang terdiri dari jawaban benar-salah (*true false*), pilihan ganda (*multiple choice*), jawaban singkat (*short answer*) dan penjodohan (*matching*) (Suwandi, 2009).

Tes objektif memiliki beberapa keunggulan yaitu, memiliki objektivitas yang tinggi dalam pemeriksaan jawaban, dapat memuat banyak soal dengan waktu tes yang singkat sehingga komprehensivitas cakupan isi tes benar-benar dapat dicapai. Butir soal pada tes objektif dapat dibuat dalam berbagai tipe. Menurut Azwar (2016) terdapat berbagai tipe soal yang dapat digunakan dalam tes format objektif yaitu, pilihan ganda (*multiple choice*), jawaban pendek (*short-answer*), dan tes dengan pilihan

benar-salah (*true-false*). Menurut Arifin (2010) butir soal tipe pilihan ganda berguna untuk mengukur hasil belajar peserta didik yang bersifat kompleks dan berhubungan dengan aspek ingatan, pengertian, aplikasi, analisis, sintesis, serta evaluasi. Butir soal pada tes pilihan ganda tersusun dari suatu pertanyaan atau pernyataan yang kemudian diikuti dengan beberapa pilihan jawaban. Pertanyaan atau pernyataan yang terdapat dalam tes pilihan ganda dinamakan *stem*, sedangkan pilihan jawaban yang tersedia dinamakan *opsi*. Pada pilihan jawaban yang tersedia terdapat satu jawaban benar yang dinamakan *kunci* dan pilihan yang lainnya dinamakan distraktor atau pengecoh.

Guru dapat mengetahui kemampuan peserta didik dengan tepat apabila alat ukur yang digunakan merupakan alat ukur yang baik. Arikunto (2008) menyatakan bahwa suatu tes dikatakan baik jika memenuhi lima kriteria, yaitu : validitas, reliabilitas, objektivitas, praktis, dan ekonomis. Kriteria minimal suatu alat ukur yang baik adalah alat ukur tersebut harus valid dan reliabel. Validitas dapat diartikan bahwa hasil evaluasi harus sesuai dengan keadaan yang dievaluasi (Agustini, dkk., 2016).

Instrumen yang valid akan menghasilkan data yang valid pula atau dapat dikatakan bahwa jika data yang dihasilkan dari sebuah instrumen valid, maka instrumen tersebut juga valid (Widoyoko, 2009). Selain validitas kriteria tes yang baik adalah reliabilitas. Reliabilitas menunjukkan kejelasan angka suatu tes, jadi reliabilitas menunjukkan sejauh mana variasi hasil tes apabila dilaksanakan pada situasi yang berbeda (Ibrahim, 2005). Selain validitas dan reliabilitas suatu tes dikatakan baik apabila tingkat kesukaran dan indeks daya beda juga baik (Arikunto, 2008).

Tingkat kesukaran yaitu mengkaji suatu butir soal dari tingkat kesukarannya agar dapat diperoleh butir soal yang termasuk dalam kategori mudah, sedang, dan sukar. Pada umumnya suatu butir soal dinyatakan baik apabila butir soal tersebut tidak terlalu mudah dan tidak terlalu sukar. Oleh sebab itu apabila butir soal tidak dapat dijawab benar oleh seluruh peserta tes karena terlalu sukar maka butir soal tersebut dinyatakan tidak baik. Sebaliknya apabila butir soal dapat dijawab oleh seluruh peserta tes karena terlalu mudah maka soal tersebut juga dinyatakan tidak baik (Bagiyono, 2017). Menurut Widoyoko (2014) menjelaskan bahwa tingkat kesukaran yang baik pada suatu tes yaitu 25% sukar, 50% sedang, dan 25% mudah.

Indeks daya beda adalah kemampuan suatu butir soal untuk membedakan suatu kelompok, yaitu kelompok atas dan kelompok bawah. Artinya suatu butir soal dapat

membedakan antara peserta didik yang telah menguasai materi dengan peserta didik yang belum menguasai materi (Kusaeri dan Suprananto, 2012). Tingkat kesukaran butir soal dapat berpengaruh pada indeks daya beda. Apabila setiap peserta didik menjawab benar ($p=1$), atau apabila setiap peserta didik menjawab salah ($p=0$), maka soal tidak dapat digunakan untuk membedakan kemampuan peserta didik (Surapranata, 2004). Semakin tinggi nilai indeks daya beda artinya butir soal tersebut mampu membedakan peserta didik yang telah menguasai materi (kelompok atas) dengan peserta didik yang belum menguasai materi (kelompok bawah). Indeks daya beda suatu butir soal berkisar antara -1,00 hingga +1,00. Semakin tinggi indeks daya bedanya maka butir tes tersebut juga semakin baik. Apabila nilai indeks daya beda suatu butir soal negatif maka banyak kelompok bawah yang menjawab benar pada butir tes tersebut dibandingkan kelompok atas (Sudaryono, 2011).

Salah satu kompetensi yang harus dimiliki seorang guru adalah kemampuan menyusun dan mengembangkan butir soal sebagai instrumen tes dengan baik dan menganalisis hasil tes tersebut. Berdasarkan hasil wawancara didapatkan hasil bahwa 66,7% guru Biologi kelas XI belum mengembangkan instrumen penilaian HOTS pada KD. 3.11, sedangkan 33,3% guru lainnya sudah mengembangkan instrumen pada KD.3.11, namun didapatkan fakta bahwa guru yang telah mengembangkan instrumen penilaian HOTS tersebut tidak memperhatikan validitas, reliabilitas, dan karakteristik butir soal tes yang telah dibuat. Sebanyak 66,7% guru yang belum mengembangkan instrumen penilaian HOTS mengaku masih menggunakan butir soal tes dengan tingkat kognitif C-1 (mengingat) dan C-2 (memahami) dikarenakan susah membedakan materi yang bisa dan tidak bisa digunakan untuk mengembangkan instrumen penilaian HOTS dan tidak semua materi juga dapat dikembangkan menjadi instrumen penilaian HOTS.

Hal tersebut didukung dengan penelitian yang dilakukan oleh Masruroh, dkk. (2012), hasil analisis pada soal ulangan semester gasal mata pelajaran Biologi kelas X di Kecamatan Petanahan Kabupaten Kebumen termasuk soal dengan taraf kognitif C-1 (mengingat) dan C-2 (memahami). Pada penelitian lain yang dilakukan oleh Gentari (2016) bahwa soal ujian tengah semester dan ujian akhir semester mata pelajaran Biologi kelas X, XI, dan XII yang digunakan di SMA se-Kecamatan Gedong Tataan, Pesawaran Tahun Ajaran 2015/2016 hanya sampai pada tingkat kognitif C-1 (mengingat) dan C-2 (memahami).

Berdasarkan paparan di atas tujuan dari penelitian ini adalah untuk mendeskripsikan instrumen penilaian HOTS yang valid dan reliabel pada sub materi dampak penyalahgunaan psikotropika kelas XI SMA yang mencakup tingkat kesukaran dan indeks daya beda.

METODE

Jenis penelitian ini adalah penelitian pengembangan, sebab penelitian ini mengembangkan butir soal HOTS. Metode yang digunakan yaitu *Research and Development* (R&D) yang terdiri dari sepuluh tahap, namun pada penelitian ini hanya sampai pada tahap keenam karena keterbatasan waktu. Penelitian ini dilakukan pada bulan Januari-Juli 2019. Sasaran penelitian ini adalah butir soal HOTS yang telah divalidasi secara teoritis dan dinyatakan valid. Uji coba dilakukan terhadap 20 peserta didik kelas XI SMA Negeri 1 Menganti Gresik yang dianggap telah mewakili untuk mengetahui reliabilitas instrumen penilaian HOTS.

Metode pengumpulan data dalam penelitian ini yaitu menggunakan metode tes belah dua (*split-half*). Butir soal HOTS yang digunakan pada saat tes yaitu tipe pilihan ganda dan terdapat stimulus soal yang berisi hasil penelitian, data-data hasil penelitian, dan kasus. Penyusunan butir soal HOTS dilakukan dengan cara menguraikan kompetensi dasar menjadi beberapa indikator kemudian mencari sumber yang sesuai dengan indikator untuk dijadikan stimulus soal, kemudian membuat pertanyaan dan pilihan-pilihan jawaban berdasarkan konteks pada stimulus. Karakteristik butir soal HOTS yang baik yaitu bersifat valid dan reliabel. Valid berarti data sesuai dengan kenyataan dan reliabel berarti tingkat keajegan atau kestabilan hasil pengukuran.

Teknik analisis data menggunakan teknik analisis secara deskriptif yang meliputi analisis hasil tes HOTS untuk mengetahui nilai reliabilitas dengan menggunakan rumus persamaan *Spearman-Brown*, nilai tingkat kesukaran dengan cara menghitung jumlah peserta tes yang menjawab benar dibagi jumlah seluruh peserta tes dikali 100%, dan indeks daya beda dengan cara menghitung jumlah peserta tes yang menjawab benar pada kelompok atas dikurangi jumlah peserta tes yang menjawab benar pada kelompok bawah dibagi setengah dari jumlah seluruh peserta tes. Berikut merupakan spesifikasi butir soal HOTS pada materi dampak penyalahgunaan psikotropika yang disajikan pada **Tabel 1**.

Tabel 1. Spesifikasi Butir Soal HOTS pada Materi Dampak Penyalahgunaan Psikotropika

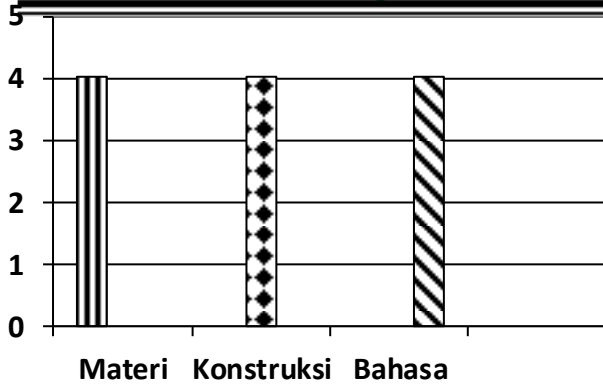
KOMPETENSI DASAR	INDIKATOR	BENTUK PENILAIAN	NO. SOAL	JENJANG KOGNITIF	JENIS HOTS
3.11 Mengevaluasi pemahaman diri tentang bahaya penggunaan senyawa psikotropika dan dampaknya terhadap kesehatan diri, lingkungan dan masyarakat.	Menganalisis faktor-faktor yang mempengaruhi penyalahgunaan narkotika dan psikotropika.	Soal pilihan ganda	1	C-4 (menganalisis)	Berpikir Kritis
		Soal pilihan ganda	8	C-4 (menganalisis)	Pengambilan Keputusan
	Menganalisis penyebab terjadinya berbagai gangguan yang terjadi pada sistem regulasi akibat penyalahgunaan narkotika dan psikotropika.	Soal pilihan ganda	3	C-4 (menganalisis)	Berpikir Kritis
		Soal pilihan ganda	10	C-4 (menganalisis)	Berpikir Kritis
	Mengaitkan pengaruh yang ditimbulkan dari berbagai bahan narkotika dan psikotropika untuk masa depan peserta didik.	Soal pilihan ganda	5	C-5 (mengevaluasi)	Berpikir Kritis
		Soal pilihan ganda	11	C-4 (menganalisis)	Berpikir Kritis
	Menganalisis dampak yang ditimbulkan akibat penyalahgunaan narkotika dan psikotropika yang ditinjau dari aspek jasmani.	Soal pilihan ganda	7	C-4 (menganalisis)	Berpikir Kritis
		Soal pilihan ganda	12	C-4 (menganalisis)	Berpikir Kritis
	Menganalisis dampak yang ditimbulkan akibat penyalahgunaan narkotika dan psikotropika yang ditinjau dari aspek rohani (kejiwaan).	Soal pilihan ganda	9	C-4 (menganalisis)	Berpikir Kritis
		Soal pilihan ganda	6	C-4 (menganalisis)	Berpikir Kritis
	Menganalisis dampak yang ditimbulkan akibat penyalahgunaan narkotika dan psikotropika yang ditinjau dari aspek lingkungan masyarakat.	Soal pilihan ganda	14	C-4 (menganalisis)	Berpikir Kritis
		Soal pilihan ganda	4	C-4 (menganalisis)	Berpikir Kritis
	Memilih sikap terhadap penggunaan senyawa narkotika dan psikotropika.	Soal pilihan ganda	13	C-5 (mengevaluasi)	<i>Problem Solving</i>
		Soal pilihan ganda	2	C-5 (mengevaluasi)	<i>Problem Solving</i>
		Soal pilihan ganda	16	C-5 (mengevaluasi)	<i>Problem Solving</i>
Soal pilihan ganda		15	C-5 (mengevaluasi)	Pengambilan Keputusan	

HASIL DAN PEMBAHASAN

Validitas yaitu dapat mengukur atribut sesuai dengan tujuannya. Sifat valid dapat dilihat dari akurasi dan kecermatan pada hasil ukur yang dilakukan oleh dosen ahli materi dan dosen ahli pendidikan dengan menggunakan instrumen validasi yang telah disediakan. Suatu tes dapat dikatakan tidak valid jika tidak mampu menghasilkan informasi yang akurat mengenai atribut yang diukurnya (Azwar, 2016). Menurut Agustini, dkk., (2016) valid atau sah diartikan bahwa hasil evaluasi harus sesuai dengan keadaan yang dievaluasi.

Validitas instrumen penilaian HOTS yang dikembangkan diperoleh dari hasil validasi oleh ahli materi dan ahli pendidikan. Dalam menyusun instrumen penilaian HOTS yang dikembangkan memperhatikan 3

aspek yaitu aspek materi, aspek konstruksi, dan aspek bahasa. Hasil validasi instrumen penilaian HOTS tersebut memperoleh modus keseluruhan sebesar 4. Modus yang dihasilkan menunjukkan bahwa instrumen penilaian HOTS valid secara teoritis dengan interpretasi sangat valid. Berikut merupakan rekapitulasi hasil validasi instrumen penilaian HOTS yang disajikan pada **Gambar 1**.



Gambar 1. Hasil Rekapitulasi Validasi Butir Soal HOTS pada Instrumen Penilaian HOTS Materi Dampak Penyalahgunaan Psikotropika

Aspek yang pertama yakni aspek materi. Pada aspek materi terdapat 3 kriteria yang divalidasi oleh para ahli yakni, 1) kesesuaian butir soal dengan indikator, 2) kesesuaian butir soal dengan kebenaran konsep, 3) terdapat satu jawaban benar pada setiap butir soal. Aspek tersebut memperoleh total skor modus dalam kategori **sangat valid**.

Aspek kedua yaitu aspek konstruksi yang terdiri dari 8 kategori yaitu 1) petunjuk soal mudah dipahami, 2) butir soal menggunakan stimulus yang menarik (mendorong peserta didik untuk membaca), 3) butir soal menggunakan stimulus yang kontekstual (gambar/grafik, teks, dll sesuai dunia nyata), 4) butir soal mengukur level kognitif penalaran (menganalisis dan mengevaluasi), 5) butir soal mengukur level kognitif HOTS (*problem solving*, pengambilan keputusan, dan berpikir kritis), 6) tidak mengandung jawaban ganda, 7) butir soal tidak bergantung pada jawaban soal sebelumnya, 8) pilihan jawaban tidak menggunakan “semua jawaban benar/salah” atau sejenisnya. Aspek tersebut memperoleh total skor modus dalam kategori **sangat valid**.

Aspek ketiga yaitu aspek bahasa yang terdiri dari 4 kategori yaitu 1) menggunakan bahasa Indonesia sesuai dengan kaidah, 2) menggunakan bahasa yang komunikatif dan mudah dimengerti, 3) tidak mengandung bahasa yang ambigu, 4) pilihan jawaban tidak mengulang kata/kelompok kata yang sama, kecuali merupakan satu kesatuan pengertian. Aspek tersebut memperoleh total skor modus dalam kategori **sangat valid**, akan tetapi pada kategori menggunakan bahasa yang komunikatif dan mudah dimengerti terdapat beberapa nomor soal dengan interpretasi kurang valid yaitu nomor soal 1, 4, 8, 11 dan 14.

Menurut Azwar (2016) suatu tes yang baik memiliki karakteristik valid dan reliabel. Valid dapat dilihat dari akurasi dan kecermatan pada hasil ukur yang dilakukan oleh seorang ahli materi dan ahli pendidikan dengan menggunakan instrumen validasi yang telah disediakan,

dengan kata lain, suatu tes yang valid akan berfungsi sebagai pengukur atribut sesuai tujuannya.

Reliabilitas merupakan keajegan angka suatu tes yang menunjukkan sejauh mana variasi hasil tes bila tes tersebut dilaksanakan pada situasi yang berbeda (Ibrahim, 2005). Hasil uji coba terbatas pada instrumen penilaian HOTS diperoleh data yang digunakan untuk menentukan nilai reliabilitas tes yang dihitung dengan menggunakan rumus persamaan *Spearman-Brown* sebab koefisien reliabilitas dihitung dengan mencari korelasi antara angka tes nomor soal ganjil dan angka tes nomor soal genap dan diperoleh nilai sebesar 0,65 dengan kategori tinggi.

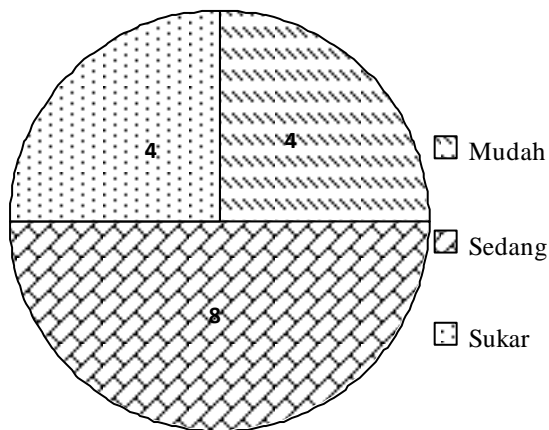
Reliabilitas diperoleh berdasarkan hasil tes peserta didik. Tes yang tidak dapat menghasilkan skor yang reliabel, tidak akan menghasilkan skor yang valid, dengan kata lain suatu tes yang valid akan berfungsi sebagai pengukur atribut sesuai dengan tujuannya (Azwar, 2016). Nilai reliabilitas yang diperoleh pada pengembangan instrumen penilaian HOTS sebesar 0,65 dengan interpretasi yaitu tinggi, sehingga dapat dikatakan bahwa instrumen penilaian HOTS yang dihasilkan memenuhi kriteria tes yang baik yaitu reliabel. Artinya suatu alat ukur dinyatakan memiliki koefisien reliabilitas yang tinggi apabila digunakan untuk mengukur suatu hal yang sama pada waktu yang berbeda hasilnya sama atau mendekati. Hal ini dapat diartikan bahwa reliabilitas merupakan sifat dari sekumpulan skor (Frisbie, 2005). Dalam dunia pendidikan, suatu alat ukur yang reliabel akan menghasilkan pengukuran yang sama informasinya walaupun penguji berbeda, korektor berbeda, atau butir soal berbeda tetapi mengukur hal yang sama dan memiliki karakteristik butir yang sama.

Perhitungan tingkat kesulitan berfungsi untuk mengetahui apakah butir soal tersebut terlalu sukar atau terlalu mudah bagi peserta didik yang ditentukan melalui perbandingan antara jawaban benar dengan banyak peserta didik yang mengikuti tes. Tingkat kesukaran butir soal tes ditentukan dari presentase peserta tes yang menjawab benar butir soal yaitu jumlah peserta didik yang menjawab benar dibagi jumlah semua peserta didik yang mengikuti tes dikali 100% (Maenani dan Oktova, 2015). Berikut merupakan tabel dan diagram hasil rekapitulasi tingkat kesukaran butir soal HOTS.

Tabel 2. Distribusi Tingkat Kesukaran Butir Soal HOTS Materi Dampak Penyalahgunaan Psikotropika

Nomor Soal	Tingkat Kesulitan	Kriteria
1.	0,65	Sedang
2.	0,55	Sedang
3.	0,6	Sedang
4.	0,65	Sedang
5.	0,45	Sedang
6.	0,3	Sukar

Nomor Soal	Tingkat Kesulitan	Kriteria
7.	0,3	Sukar
8.	0,95	Mudah
9.	0,3	Sukar
10.	0,85	Mudah
11.	0,5	Sedang
12.	0,45	Sedang
13.	0,65	Mudah
14.	0,85	Mudah
15.	0,75	Mudah
16.	0,30	Sukar



Gambar 2. Tingkat Kesukaran Soal HOTS Materi Dampak Penyalahgunaan Psikotropika

Perhitungan tingkat kesulitan berfungsi untuk mengetahui apakah butir soal tersebut terlalu sukar atau terlalu mudah bagi peserta didik yang ditentukan melalui perbandingan antara jawaban benar dengan banyak peserta didik yang mengikuti tes. Pada setiap instrumen penilaian HOTS yang dikembangkan memperoleh hasil yaitu masing-masing 25% mudah, 50% sedang, 25% sukar. Menurut Sudjana (2004) sebaiknya sebuah tes yang diberikan kepada siswa memiliki keseimbangan antara mudah: sukar : sedang dengan perbandingan 3 : 4 : 3. Berikut ini merupakan butir soal HOTS yang memiliki tingkat kesulitan yang sukar. Butir soal tersebut disajikan pada **Gambar 3**.

Berdasarkan butir soal HOTS di atas dapat dijawab benar oleh peserta didik sebanyak 6 dari 20 peserta didik yang mengikuti tes dan diperoleh nilai tingkat kesukaran sebesar 0,3 dengan interpretasi sukar. Berdasarkan indeks tingkat kesukaran apabila nilai yang diperoleh besar maka butir soal tersebut termasuk dalam kategori mudah yang banyak dikerjakan peserta didik dengan benar, apabila nilai yang diperoleh kecil dan hamr mendekati nol maka butir soal tersebut sukar dikerjakan peserta didik. Butir soal yang terlalu mudah tidak dapat mendorong peserta didik untuk menyelesaikan masalah (memilih jawaban yang tepat), sedangkan butir soal yang terlalu sukar dapat menyebabkan peserta didik putus asa karena tidak mampu menjawab butir soal tersebut (Arikunto, 2012).

Suatu butir soal termasuk dalam kategori sukar apabila butir soal tersebut memiliki 2 atau lebih jawaban benar, materi yang ditanyakan belum diajarkan atau belum tuntas pembelajarannya sehingga kompetensi minimum yang harus dikuasai peserta didik belum tercapai, materi yang diukur tidak cocok ditanyakan dengan menggunakan bentuk soal yang diberikan, dan pertanyaan atau kalimat terlalu panjang (Rokhyati, 2011).

Butir soal yang terlalu mudah atau terlalu sukar bukan berarti tidak boleh digunakan. Hal ini tergantung dari tujuan penggunaannya. Apabila peserta tes banyak padahal yang dikehendaki lulus hanya sedikit maka diambil peserta dengan nilai terbaik, untuk itu diberikan butir soal tes dengan kategori sukar. Demikian sebaliknya jika kekurangan peserta tes, maka diberikan butir soal tes dengan kategori mudah. Selain itu, butir soal yang sukar akan menambah motivasi belajar bagi peserta didik golongan atas, sedangkan butir soal yang mudah akan membangkitkan semangat peserta didik kelompok bawah (Arikunto, 2001).

6. Perhatikan kasus berikut, kemudian pilihlah jawaban yang tepat!

BALI EXPRESS, SINGARAJA - Motif penganiayaan terhadap Made Subakti, 64 yang dilakukan I Gede Ariawan, 38 akhirnya terkuak. Dari hasil penyelidikan, pelaku diduga dalam pengaruh obat terlarang. hingga nekat memukul korban dengan besi sepanjang satu meter Jumat (16/11) lalu.

"Dari kronologis memang tidak logis, pelaku tanpa sebab tiba-tiba saja memukul orang yang tidak dikenal tanpa ada masalah. Nah dari sana penyidik berinisiatif melakukan tes urine terhadap pelaku. Ternyata hasilnya positif," ujar AKP Wisnaya, Minggu (18/11) siang.

Indikasi kasus yang dialami pelaku sebut AKP Wisnaya, terjadi saat korban Made Subakti memanggil rekannya sambil berteriak di Banjar Peken, Desa Sangsit, Kecamatan Sawan. Kala itu korban memang mendatangi rumah temannya yang posisinya bertetangga dengan pelaku. Pelaku mengira seolah-olah orang tersebut akan menyerangnya sehingga dia mengambil besi berbentuk pipih sepanjang satu meter untuk memukul korban.

Berdasarkan kasus di atas bagaimana dampak yang ditimbulkan akibat dari penyalahgunaan narkoba dan psikotropika yang ditinjau dari aspek rohani (kejiwaan)?

- Mudah marah
- Mengalami paranoid
- Berpikir aneh dan dangkal
- Tidak dapat berpikir jernih
- Gangguan pada sistem syaraf

Gambar 3. Salah Satu Soal HOTS yang Memiliki Nilai Tingkat Kesukaran Tertinggi

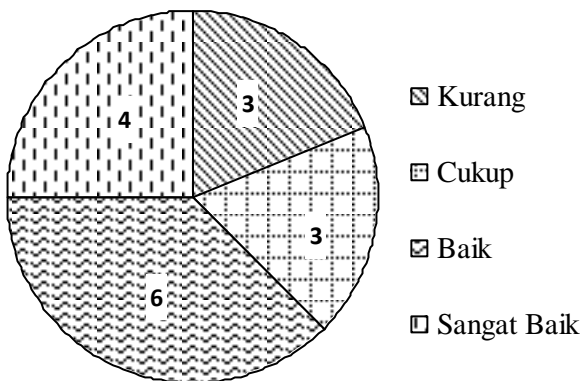
1. Indeks Daya Beda

Indeks daya beda merupakan kemampuan butir tes untuk membedakan antara peserta didik yang termasuk kelompok atas dengan peserta didik yang termasuk kelompok bawah (Agustini, dkk., 2016). Indeks daya beda

butir soal dapat diketahui dengan cara mengurutkan skor hasil tes seluruh peserta didik dari skor tertinggi sampai skor terendah. Kemudian dibagi dua sama besar, 50% peserta didik dengan skor tinggi dari seluruh peserta tes (kelompok atas) dan 50% peserta didik lainnya dengan skor rendah (kelompok bawah). Perhitungan nilai indeks daya beda yakni jumlah peserta didik yang menjawab benar dari kelompok atas dikurangi jumlah peserta didik yang menjawab benar dari kelompok bawah dibagi setengah dari jumlah peserta didik yang mengikuti tes (Ibrahim, 2005). Berikut merupakan grafik hasil rekapitulasi indeks daya beda butir soal HOTS.

Tabel 3. Distribusi Indeks Daya Beda Butir Soal HOTS Materi Dampak Penyalahgunaan Psikotropika

Nomor Soal	Indeks Daya Beda	Kriteria
1.	0,30	Baik
2.	0,20	Cukup
3.	0,40	Sedang
4.	0,10	Kurang
5.	0,30	Baik
6.	0,40	Sangat baik
7.	0,20	Cukup
8.	0,10	Kurang
9.	0,60	Sangat baik
10.	0,30	Baik
11.	0,20	Cukup
12.	0,30	Baik
13.	0,30	Baik
14.	0,10	Kurang
15.	0,30	Baik
16.	0,60	Sangat baik



Gambar 4. Indeks Daya Beda Soal HOTS Materi Dampak Penyalahgunaan Psikotropika

Perhitungan indeks daya beda berfungsi untuk membedakan kemampuan peserta didik yang memiliki kemampuan tinggi dan kemampuan rendah. Apabila butir soal dapat membedakan peserta didik yang memiliki

kemampuan tinggi dan rendah maka butir soal tersebut semakin baik atau layak (Sulistiyadi, 2016). Berdasarkan data yang diperoleh didapatkan hasil perhitungan nilai indeks daya beda pada instrumen penilaian HOTS yang dikembangkan yaitu 19% kurang, 19% cukup, 37% baik, dan 25% sangat baik. Indeks daya beda butir soal dengan interpretasi kurang artinya butir soal tersebut tidak mampu membedakan peserta didik yang memiliki kemampuan tinggi dan rendah. Indeks daya beda dengan interpretasi kurang sebanyak 3 butir yaitu nomor soal 4, 8, dan 14. Indeks daya beda dengan interpretasi cukup sebanyak 3 butir yaitu nomor soal 2, 7, dan 11. Indeks daya beda dengan interpretasi baik sebanyak 6 butir yaitu nomor soal 1, 5, 10, 11, 13, dan 15. Indeks daya beda dengan interpretasi sangat baik sebanyak 4 butir yaitu nomor soal 3, 6, 9, dan 16.

SIMPULAN

Berdasarkan hasil penelitian pengembangan instrumen penilaian HOTS pada materi dampak penyalahgunaan psikotropika dapat disimpulkan bahwa instrumen penilaian HOTS yang telah dinyatakan valid dan reliabel berdasarkan hasil validasi ahli pakar dapat disimpulkan bahwa instrumen penilaian HOTS dalam kategori sangat valid. Reliabilitas diperoleh berdasarkan hasil tes peserta didik dengan memperoleh nilai reliabilitas sebesar 0,65. Indeks tingkat kesukaran yang diperoleh heterogen yaitu, 25% mudah, 50% sedang, dan 25% sukar. Indeks daya beda yang diperoleh yaitu 19% kurang baik, 19% cukup, 37% baik, dan 25% sangat baik, sehingga dapat dinyatakan bahwa 62% soal HOTS dapat membedakan kemampuan peserta didik. Berdasarkan penelitian yang telah dilakukan dapat disimpulkan bahwa butir HOTS materi dampak penyalahgunaan psikotropika dinyatakan valid dan reliabel.

DAFTAR PUSTAKA

- Agustini, R., Nasrudin, H., Azizah, U., dan Muchlis. 2016. *Asesmen*. Yogyakarta : Absolute Media.
- Arifin, Z. 2010. *Evaluasi Pembelajaran Prinsip Teknik Prosedur*. Bandung : Remaja Rosdakarya.
- Arikunto, S. 2008. *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta : Rineka Karya.
- Arikunto, S. 2012. *Dasar-Dasar Evaluasi Pendidikan Edisi 2*. Jakarta : Bumi Aksara.
- Arikunto, S. 2001. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta : Bumi Aksara.
- Azwar, S. 2016. *Konstruksi Tes Kemampuan Kognitif*. Yogyakarta : Pustaka Belajar.
- Bagiyono. 2017. "Analisis Tingkat Kesukaran dan Daya Pembeda Butir Soal Ujian Pelatihan Radiografi Tingkat I". *Jurnal Widyanuklida Vol.16 No.1*.
- Frisbie, D. A. 2005. *Measurement 101 : Some fundamentals revisited*. Educational Measurement : Issues and Practice.
- Ibrahim, Muslimin. 2005. *Asesmen Berkelanjutan*. Surabaya : Unesa University Press.

- Kemendikbud. 2016. *Permendikbud No.23 tentang Standar Penilaian Pendidikan*. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- King, J. F., Ludwika G., dan Faranak R. 2010. *Higher Order Thinking Skills, Definition, Teaching Strategies, Assessment*. A Publication of The Educational Services Program.
- Kusaeri dan Supranoto. 2012. *Pengukuran dan Penilaian Pendidikan*. Yogyakarta : Graha Ilmu.
- Maenani, L. dan Oktova, R. 2015. "Analisis Butir Soal Fisika Ulangan Umum Kenaikan Kelas X". *Jurnal Berkala Fisika Indonesia Vol. 7 No.1*.
- Masruroh, Ely, R., dan Saiful, R. 2012. "Analisis Soal Ulangan Semester Gasal Biologi Kelas X di Kecamatan Petanahan Kebumen". *Unnes Journal of Biology Education Vol. 1 No. 2*.
- Murtono dan Evi, M. 2014. "Pengembangan Instrumen Evaluasi Dengan Teknik Simulasi Sebagai Asesmen Alternatif dalam Pembelajaran Fisika Materi Mekanika Fluida SMA Kelas XI". *Jurnal Inovasi dan Pembelajaran Fisika*.
- Nugroho, R. A. 2018. *HOTS (Kemampuan Berpikir Tingkat Tinggi : Konsep, Pembelajaran, Penilaian, dan Soal-soal)*. Jakarta : PT. Gramedia.
- Rokhyati. 2011. Karakteristik Secara Kualitatif dan Kuantitatif Soal Ulangan Akhir Semester Genap Bahasa Indonesia Kelas XII SMA Negeri di Kabupaten Purbalingga Tahun Pelajaran 2010/2011. *Skripsi*. Universitas Negeri Yogyakarta : diterbitkan di <https://eprints.uny.ac.id> (diakses pada tanggal 13 Juli 2019).
- Sudaryono. 2011. "Implementasi Teori Responsi Butir (*Item Response Theory*) pada Penilaian Hasil Belajar Akhir di Sekolah". *Jurnal Pendidikan dan Kebudayaan Vol. 17 No. 6*
- Sudjana, N. 2004. *Penilaian Hasil Proses Belajar Mengajar*. Bandung : Remaja Rosdakarya Offset.
- Sulistiyadi, D.C. 2018. Analisis Kualitas Butir Soal Pilihan Ganda pada Ulangan Akhir Semester Mata Pelajaran Penjasorkes Kelas X SMKN 1 Ngasem Kabupaten Kediri Tahun Ajaran 2017/2018. *Skripsi*. Universitas Negeri Yogyakarta : diterbitkan di <https://eprints.uny.ac.id> (diakses pada tanggal 13 Juli 2019).
- Supranata, S. 2009. *Analisis, Validitas, Reliabilitas, dan Interpretasi Hasil Tes : Implementasi Kurikulum 2004*. Bandung : PT. Remaja Rosdakarya.
- Suwandi, S. 2009. *Model Asesmen Dalam Pembelajaran*. Surakarta : Mata Padi Presindo.
- Wardana, N. 2010. "Pengaruh Model Pembelajaran Berbasis Masalah dan Ketahananmalangan Terhadap Kemampuan Berfikir Tingkat Tinggi dan Pemahaman Konsep Fisika". *Jurnal Ilmiah Pendidikan dan Pembelajaran IPA Indonesia Vol. 2 No.1*.
- Widoyoko, S.E.P. 2009. *Evaluasi Program Pembelajaran (Panduan praktis Bagi Pendidik dan Calon Pendidik)*. Yogyakarta : Pustaka Pelajar.
- Widoyoko, S.E.P. 2014. *Penilaian Hasil Pembelajaran di Sekolah*. Yogyakarta : Pust aka Belajar.