

DAYA SERAP MATERI PELAJARAN DAN RELIABILITAS TES MATEMATIKA BENTUK SOAL URAIAN

Absorbent Power Power Subject Matter and Reliability Math Test Form Item Essay

Fahmi

Pusat Asesmen dan Pembelajaran, Kementerian Pendidikan dan Kebudayaan
Jalan Gunung Sahari Raya No. 4 Jakarta Pusat
Email: ffahmi6@gmail.com

Naskah Diterima Tanggal 27 Maret 2020 – Direvisi Akhir Tanggal 5 Mei 2020 –
Disetujui Tanggal 26 Mei 2020 – Publikasi Online: 28 Juni 2020
DOI: <https://doi.org/10.26499/ijeajournal.v3i1.58>

Abstract. This study aims to determine the proportion correct of mathematics subject matter and compare the reliability of essay tests that are scored with 3 different rubric models. The method used in this research is an experiment using design treatments by subjects, an experiment using one group (one group experiment) which is also an experimental group and a control group in different experimental periods. Schools that were sampled were 6 public and private junior high schools and students were sampled as many as 452 students. The research sample was determined by two-stage stratified random sampling. The objects in this study consisted of 22 essay questions in Mathematics for junior high school compiled from the same test specifications. Scoring uses 3 scoring rubrics models, namely scoring rubric Model 1 (0,1), Model 2 (0,1,2), and Model 3 (keywords). The results obtained from this study indicate the reliability of the mathematical essay test on the scoring rubric of model 1 is 0.6859, model 2 is 0.7025, and model 3 is 0.6986. Test reliability of the three scoring rubrics is high or reliable. Students' competence in completing mathematical essay questions is still low, which is under 50%. Only mathematical essay questions calculate volume at the cognitive level of knowledge and understanding above 70%. Student competencies in solving questions that measure reasoning is an average of 12.20%. In other words, the proportion correct of subject matter in two-dimensional figure and space less than 50%, this shows that students still do not understand the concept of the subject matter, especially the material of the central angular relationship, arc length, sector area, which can only be answered correctly by 17.58% of students.

Keywords: tests, proportion correct, item difficulty, reliability, scoring rubric

PENDAHULUAN

Dalam kehidupan sehari-hari, kita sering melakukan kegiatan yang berkaitan dengan pengukuran, misalnya mengukur tinggi sebuah benda. Kita dapat melakukannya dengan mudah karena objek yang diukur merupakan objek kasat mata. Hasil pengukurannya pun tidak banyak berubah atau stabil walaupun dilakukan berulang kali. Akan berbeda hasilnya jika objek yang diukur merupakan benda abstrak seperti prestasi, sikap, kejujuran, dan lain sebagainya. Untuk melakukan pengukurannya diperlukan keahlian tertentu. Instrumen yang dapat digunakan di antaranya: tes, angket, atau alat ukur lainnya.

Respons yang diperoleh dari peserta tes biasanya berbentuk skor perolehan/tampak. Skor tampak tersebut merupakan pendekatan dari atribut yang diukur sehingga sangat penting memerhatikan tingkat konsistensi dari hasil pengukuran. Tes di berbagai lembaga, perusahaan, atau institusi sekolah dapat dilakukan untuk tujuan seleksi pegawai, mengetahui kompetensi guru, peserta didik, maupun untuk mendiagnosis kelemahan peserta didik. Bentuk soal yang sering digunakan di institusi sekolah adalah pilihan ganda dan uraian. Hal tersebut diketahui dari hasil penelitian yang dilakukan oleh Umar dan Hayat (2000) terhadap bentuk soal yang digunakan guru

dalam pemberian ulangan harian, antara lain sebanyak 81,0% guru menggunakan soal uraian dan sebanyak 56,5% guru masih menggunakan soal pilihan ganda.

Salah satu kelemahan dari soal pilihan ganda ialah adanya peluang untuk menebak, sedangkan kelemahan bentuk soal uraian adalah sulit membuat rubrik penskorannya. Rubrik penskoran dibuat untuk memperkecil faktor subjektivitas penskor. Faktor subjektivitas penskor dapat diperkecil dengan cara setiap butir soal uraian dilengkapi dengan pedoman atau rubrik penskoran. Dengan menggunakan rubrik penskoran diharapkan hasil penskoran relatif stabil dan konsisten.

Adanya beberapa bentuk rubrik penskoran yang digunakan untuk menskor butir soal uraian dikhawatirkan menyebabkan hasil penskoran menjadi berbeda sehingga dapat memengaruhi karakteristik butir soal seperti tingkat kesukaran, daya beda, dan reliabilitas tes. Namun dari hasil penelitian yang dilakukan Fahmi dan Idris (2019) diketahui bahwa tidak terdapat perbedaan tingkat kesukaran soal uraian yang diskor menggunakan rubrik penskoran berbeda. Tentunya perlu dilakukan penelitian lebih lanjut tentang reliabilitas tes yang diskor menggunakan rubrik penskoran berbeda serta daya serap materi yang diujikan dari tes tersebut.

Permasalahan

Permasalahan dalam penelitian ini adalah seberapa besar daya serap materi pelajaran bangun datar dan bangun ruang, serta reliabilitas tes soal uraian matematika yang diskor dengan menggunakan tiga rubrik penskoran berbeda.

Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengetahui seberapa besar daya serap materi pelajaran bangun datar dan bangun ruang, serta reliabilitas tes soal uraian matematika yang diskor dengan menggunakan tiga rubrik penskoran berbeda.

METODE PENELITIAN

Penelitian ini merupakan penelitian eksperimen dengan menggunakan *design treatments by subjects* yaitu eksperimen yang menggunakan satu kelompok (*one group experiment*) yang sekaligus menjadi kelompok eksperimen dan kelompok kontrol pada periode eksperimen berlainan. Dalam penelitian ini, hanya terdapat satu kelompok peserta didik sebagai kelompok kontrol yang dikenai tes bentuk soal uraian dan sekaligus menjadi kelompok eksperimen pada saat dilakukan penskoran dengan menggunakan rubrik penskoran berbeda. Objek dari penelitian ini adalah 22 butir soal uraian yang dibagi ke dalam 2 klaster, setiap klaster terdiri atas sebelas soal uraian. Populasi

penelitian ini adalah seluruh siswa SMP negeri dan swasta di DKI Jakarta, sedangkan penentuan sampel dilakukan dengan cara *two-stage random sampling and stratified random sampling*, yaitu pada tahap pertama dilakukan random sampling wilayah dan kecamatan menggunakan teknik undian. Setelah diundi, dari lima wilayah, terpilih sampel Wilayah Kota Jakarta Selatan dan dari 10 kecamatan di Jakarta Selatan, terpilih sampel Kecamatan Pasar Minggu. Sekolah yang menjadi lokasi uji coba dan sampel penelitian ialah sebanyak 6 sekolah SMP negeri dan swasta yang dipilih dari 24 SMP yang terdapat di Kecamatan Pasar Minggu. Penentuan sekolah sampel dalam penelitian ini dilakukan dengan mengelompokkan seluruh SMP negeri dan swasta yang ada di Kecamatan Pasar Minggu ke dalam tiga kategori, yaitu SMP dengan mutu sekolah baik, sedang, dan kurang. Dari tiap kelompok dipilih secara random 2 sekolah menggunakan teknik undian. Hasilnya, kelas sampel uji coba adalah kelas VIII dan dari tiap sekolah diambil 2 (dua) kelas utuh yang dipilih secara random menggunakan teknik undian. Jumlah siswa yang menjadi sampel sebanyak 452 siswa. Data peserta didik yang telah diskor kemudian dipakai untuk menghitung reliabilitas dari tes matematika tersebut.

KAJIAN TEORI

Seyogianya, pembelajaran berdampak positif terhadap peserta didik dalam kehidupan nyata. Apa yang diperoleh di kelas diharapkan dapat diterapkan dalam kehidupan nyata. Oleh karena itu, guru diharapkan dapat merancang pembelajaran di kelas sedemikian rupa sehingga proses belajar mengajar dapat berjalan dengan baik. Keberhasilan pembelajaran dapat diketahui salah satunya melalui tes. Hasil tes dapat digunakan untuk mengetahui prestasi belajar siswa dan memperbaiki mutu pendidikan. Hasil tes tersebut dapat berupa nilai atau daya serap (dalam persen) materi pelajaran. Menurut Depdikbud (2012), daya serap adalah seberapa besar secara persentase peserta didik dapat menjawab pertanyaan dalam tes. Daya serap, atau penguasaan materi ini, diperoleh dari hasil perhitungan jawaban peserta didik dari setiap kemampuan yang diujikan di setiap jenjang. Charles D. Hopkins dan Richard L. Antes (1979), tingkat kesukaran soal diukur dari jumlah peserta tes yang menjawab butir soal dengan benar, dan diukur dari jumlah peserta tes yang menjawab salah, sedangkan Suryabrata (1987), mengatakan tingkat kesukaran butir soal adalah proporsi subyek yang menjawab soal itu dengan betul. Jadi pada dasarnya daya serap materi pelajaran merupakan persentase dari tingkat kesukaran butir soal. Misal tingkat kesukaran butir soal adalah 0,87, hal tersebut berarti 87% siswa menjawab benar soal tersebut atau daya serap materi yang diujikan sebesar 87%.

Tes sebagai alat ukur harus memenuhi prasyarat sebagai alat ukur yang baik. Untuk mendapatkan tes yang baik, tes harus dirancang sedemikian rupa dan sistematis. Hal tersebut sesuai dengan pendapat Cronbach (1960), Linn & Gronlund (1990), dan Crocker dan Algina (1986), tes sebagai prosedur yang sistematis atau baku untuk mengamati atau memperoleh sampel perilaku peserta didik serta dari suatu ranah tertentu. Dari pendapat ahli tersebut dapat disimpulkan bahwa tes merupakan sebuah alat ukur standar yang dapat digunakan untuk memperoleh informasi tentang kemampuan yang dimiliki oleh peserta tes. Untuk tes prestasi belajar, ranah yang diukur berkaitan dengan level kognitif. Menurut Kemdikbud (2019), pada dasarnya ada tiga level kognitif untuk mendapatkan gambaran secara spesifik tingkat kemampuan peserta didik, yaitu: pengetahuan dan pemahaman, aplikasi, dan penalaran. Peserta didik pada level pengetahuan dan pemahaman memiliki kemampuan standar minimum dalam menguasai pelajaran; peserta didik pada level aplikasi memiliki kemampuan memperlihatkan pengetahuan dan pemahaman terhadap materi pelajaran dan dapat mengaplikasikan gagasan-gagasan dan konsep-konsep dalam konteks tertentu; sedangkan peserta didik pada level penalaran memiliki kemampuan menganalisis, mensintesis, mengevaluasi, menginterpretasi, memecahkan masalah, dan mendemonstrasikan pemikiran-pemikiran yang original.

Menurut M. Chabib Thoha (1996), tes dibagi menjadi tes kepribadian dan tes hasil belajar, sedangkan Arikunto (1997) membagi tes hasil belajar menjadi dua bentuk yaitu: (1) tes subjektif, pada umumnya berbentuk uraian dan (2) tes objektif, merupakan tes yang penskorannya dapat dilakukan secara objektif. Selanjutnya, tes objektif dibedakan menjadi: (1) tes benar-salah, (2) tes pilihan ganda, (3) tes menjodohkan, dan (4) tes isian atau melengkapi. Dari hasil penelitian yang dilakukan oleh Gronlund dan Linn (1990), kedua bentuk soal pilihan ganda dan soal uraian masing-masing mempunyai keunggulan dan kelemahan.

Tes yang standar pembuatannya lebih sulit dibandingkan dengan membuat alat ukur penggaris atau meteran. Hasil pengukuran menggunakan tes harus dapat dipercaya. Menurut Suryabrata (1987), syarat utama tes yang baik adalah valid dan reliabel. Syarat yang lain yaitu tes harus objektif, tidak diskriminatif, komprehensif, dan mudah digunakan. Valid artinya setiap butir tes mengukur atribut sesuai dengan apa yang ingin diukur, sedangkan menurut Allen dan Yen (1979), tes dikatakan reliabel jika skor amatan mempunyai korelasi yang tinggi dengan skor yang sebenarnya. Tes sebagai alat ukur yang standar harus dirancang sedemikian rupa dan sistematis. Menurut Azwar (1987), mutu informasi yang didapat dari hasil pengujian ditentukan oleh mutu tes, sedangkan

mutu tes ditentukan oleh mutu butir soal yang dirakit dalam tes tersebut. Pengujian mutu setiap butir soal dilakukan melalui analisis butir soal, baik secara kualitatif maupun secara kuantitatif. Analisis butir secara kuantitatif dilakukan untuk melihat karakteristik butir soal seperti daya beda soal, tingkat kesukaran, dan reliabilitas tes tersebut.

Menurut Brenan (2006), validitas tes atau kesahihan suatu tes adalah kemampuan suatu tes untuk mengukur karakteristik atau dimensi sesuai dengan apa yang ingin diukur, sedangkan reliabilitas atau keandalan suatu tes mengacu pada derajat suatu tes yang mampu mengukur berbagai atribut secara konsisten. Konstruksi tes yang baik harus memenuhi kedua syarat tersebut, sehingga tes itu mampu memberikan gambaran yang sebenarnya terhadap kondisi *testee* (orang) yang diuji. Menurut Mehrens & Lehmann (1991), reliabilitas merupakan derajat kejajegan (*consistency*) di antara dua buah hasil pengukuran pada objek yang sama. Sedangkan menurut Anastasi dan Urbana (1988), reliabilitas tes merupakan kestabilan skor yang diperoleh orang yang sama ketika mereka diuji ulang dengan tes yang sama pada saat yang berbeda, dengan perangkat soal-soal ekuivalen yang berbeda, atau pada kondisi pengukuran yang berbeda.

Reliabilitas memiliki dua kejajegan yaitu kejajegan internal dan kejajegan eksternal. Kejajegan internal yaitu tingkat sejauhmana butir soal tersebut homogen, baik dari segi tingkat kesukaran maupun bentuk soalnya. Sedangkan kejajegan eksternal yaitu sejauhmana skor yang dihasilkan tetap sama sepanjang kemampuan orang yang diukur belum berubah. Reliabilitas tes dapat dipengaruhi oleh teknik penskoran. Semakin objektif dalam memberikan skor, semakin tinggi tingkat kejajegannya. Dalam arti yang lebih luas, reliabilitas tes dapat diartikan sebagai proporsi varians skor perolehan yang merupakan varians skor murni dan dinyatakan dalam koefisien korelasi.

Allen dan Yen (1979) ada tiga pendekatan untuk mengestimasi reliabilitas tes, yaitu metode tes ulang (*test-retest*), metode bentuk paralel (*parallel forms*), dan metode konsistensi internal (*internal consistency*). Untuk menentukan reliabilitas tes menggunakan metode *test-retest*, satu perangkat tes dilakukan tes dua kali terhadap kelompok siswa yang sama dalam waktu yang berbeda. Kedua, hasil tes dikorelasikan. Pendekatan dengan tes paralel dilakukan dengan cara membuat dua tes paralel, kemudian kedua tes tersebut di teskan kepada sekelompok subjek, lalu hasilnya dikorelasikan, sedangkan metode konsistensi internal hanya menggunakan satu tes dan pengujian juga dilakukan satu kali. Untuk menentukan koefisien reliabilitas dengan metode ini, instrumen dibelah menjadi belahan-belahan yang paralel atau setara. Untuk menghitung

reliabilitas tes dengan metode konsistensi internal dapat dilakukan dengan berbagai cara atau rumus seperti Spearman Brown, Alpha Cronbach, Kuder Richadson-20 dan lain sebagainya.

Pada umumnya, metode yang sering digunakan adalah metode konsistensi internal karena kepraktisan penggunaannya dibandingkan dengan metode tes ulang dan metode bentuk paralel. Menurut Khumaedi (2012), kesulitan utama menggunakan metode bentuk paralel adalah dalam rangka mendapatkan instrumen yang benar-benar paralel secara keseluruhan. Sementara Azwar (2000) menyatakan untuk menentukan reliabilitas tes yang itemnya diberi skor dikotomi ataupun skor bukan dikotomi dapat menggunakan teknik analisis varians Hoyt yang dirumuskan sebagai berikut:

$$r_{xx'} = 1 - \frac{s_e^2}{s_x^2}$$

dimana s_e^2 adalah varians eror, s_x^2 adalah varians antarsubjek. Konsep teknik analisis varians Hoyt adalah memandang distribusi item keseluruhan subjek sebagai data pada suatu desain eksperimen faktorial dua-jalan tanpa replikasi yang dikenal sebagai *item by subject design*. Setiap item dianggap seakan suatu *treatment* atau perlakuan yang berbeda sehingga setiap kali subjek dihadapkan pada suatu item seakan-akan ia berada pada suatu perlakuan yang berbeda. Dalam hal ini, banyaknya item merupakan banyaknya perlakuan. Menurut Arikunto (2003), sebuah perangkat tes dikatakan mempunyai reliabilitas tinggi jika nilai koefisien korelasinya 0,600–0,799, dan dikatakan sangat tinggi jika koefisien korelasi di atas 0,800. Sedangkan menurut Santoso dalam Triton (2006), tes dikatakan reliabel bila nilai koefisien korelasi alpha 0,600–0,800 dan dikatakan sangat reliabel mempunyai koefisien alpha lebih dari 0,800. Demikian juga Litwin (1995) mengatakan koefisien reliabilitas 0,7 atau lebih dapat diterima sebagai reliabilitas yang baik.

Skoring Data

Setelah dilakukan pengetesan soal, langkah selanjutnya adalah melakukan penskoran terhadap jawaban peserta didik. Menurut Suryabrata (1987), penskoran adalah proses pendefinisian hasil tes ke dalam skala tertentu. Nitko (1996) mengatakan ada dua metode penskoran yang digunakan untuk menskor soal uraian, yaitu metode analitik (*analytic*), yakni pemberian skor berdasarkan kata kunci, nilai/rentang nilai, atau trait (ciri-ciri yang akan diukur) dan metode holistik, yaitu pemberian skor dilakukan secara global/umum, *sorting* (berurutan) atau rating (skala).

Menurut Bond & Fox (2007) dan DeMars (2010), pada penyekoran dikotomis ada dua kategori skor jawaban, yaitu jawaban betul diberi skor 1 (satu) dan jawaban salah diberi skor 0 (nol). Sementara itu Depdikbud(1993) menyatakan pedoman penskoran adalah pedoman yang memuat kemungkinan-kemungkinan jawaban benar atau kata-kata kunci serta besarnya skor yang ditetapkan untuk setiap kata kunci jawaban. Selanjutnya dalam penskoran soal uraian objektif, skor hanya dimungkinkan dengan dua kategori, yaitu “Benar” atau “Salah”. Untuk setiap kata kunci yang benar diberi skor 1 (satu) dan salah diberi skor 0 (nol). Satu rumusan jawaban dapat memuat lebih dari satu kata kunci, sehingga skor maksimum suatu jawaban adalah 1 (satu) atau lebih. Setiap kata kunci yang tidak dijawab atau dijawab “Salah” diberi skor 0 (nol). Sementara itu, menurut tim Puspendik (2019), pemberian skor sebaiknya per nomor soal yang sama untuk semua jawaban peserta didik agar konsisten dalam penskoran. Dalam survei *TIMSS* (2010), rubriks yang menggunakan dua digit, digit pertama untuk menentukan derajat kebenaran jawaban peserta didik dan digit kedua digunakan untuk mengklasifikasi metode yang digunakan peserta didik dalam menyelesaikan masalah.

Dari beberapa pendapat di atas dapat disimpulkan bahwa pemberian skor (*scoring*) adalah proses pemberian angka-angka dari suatu jawaban atau respons peserta didik terhadap suatu soal. Penskoran yang dilakukan dalam penelitian ini menggunakan tiga model rubriks penskoran, yaitu rubriks penskoran model 1 (0,1), artinya setiap rumusan jawaban peserta didik yang dijawab benar diberi skor 1 dan salah atau tidak dijawab diberi skor 0 (nol), rubriks penskoran model 2 (0,1,2), skor maksimum adalah 2 (dua) bila jawaban yang diberikan tanpa ada kesalahan (bernilai penuh), skor 1 (satu) untuk jawaban benar sebagian, dan skor 0 (nol) untuk jawaban salah atau tidak dijawab. Rubriks penskoran model 3 (berdasarkan kata kunci), setiap kata kunci dari jawaban peserta didik diberi skor.

HASIL DAN PEMBAHASAN

Dimensi yang diukur dalam penelitian terdiri atas dimensi konten dan dimensi kognitif. Dimensi konten meliputi bangun datar dan bangun ruang, sedangkan dimensi kognitif adalah pengetahuan, aplikasi, dan penalaran. Distribusi persentase dimensi konten dan kognitif disajikan pada Tabel 1.

Tabel 1. Distribusi persentase dimensi konten, level kognitif, dan materi yang diujikan

Dimensi Konten		Dimensi Konten	%	Materi yang Diujikan
41%	Bangun datar	Bangun datar	13%	Keliling dan luas
59%	Bangun ruang		14%	Garis singgung persekutuan dua lingkaran
Level Kognitif		Bangun ruang	14%	Hubungan sudut pusat, panjang busur, luas juring
63%	Pengetahuan Pemahaman		59%	Luas permukaan dan volum
14%	Aplikasi			
23%	Penalaran			

Pada Tabel 1 dapat dilihat bahwa instrumen yang digunakan dalam penelitian terdiri atas bangun datar dan bangun ruang, dengan jumlah butir soal yang digunakan materi bangun ruang lebih banyak daripada bangun datar. Butir soal

yang digunakan lebih mengukur level pengetahuan dan pemahaman dibandingkan mengukur level aplikasi dan penalaran. Persentase daya serap materi pelajaran bangun datar dan bangun ruang disajikan pada tabel 2 berikut.

Tabel 2. Persentase daya serap materi pelajaran bangun datar dan bangun ruang

Dimensi Konten	Materi	Daya Serap (%)
Bangun datar	Keliling dan luas.	37,35
	Garis singgung persekutuan dua lingkaran	41,47
	Hubungan sudut pusat, panjang busur, dan luas juring	17,58
Bangun ruang	Luas permukaan	39,61
	Volum	47,06
	Balok kerangka	23,24

Dari tabel 2, daya serap materi pelajaran bangun datar dan bangun ruang kurang dari 50%. Hal ini menunjukkan bahwa peserta didik masih kurang memahami konsep materi pelajaran tersebut terutama materi hubungan sudut pusat, panjang busur, dan luas juring, yang hanya dapat

dijawab benar oleh peserta didik sebanyak 17,58%. Kemampuan peserta didik dalam menyelesaikan soal-soal bangun datar dan bangun ruang berdasarkan level kognitif disajikan pada tabel 3 berikut.

Tabel 3. Persentase kemampuan peserta didik menyelesaikan soal berdasarkan level, kognitif

Dimensi Konten	Materi	Level Kognitif		
		Pengetahuan Pemahaman	Aplikasi	Penalaran
Bangun datar	Keliling dan luas	-	48,27%	15,51%
	Garis singgung persekutuan dua lingkaran.	36,69%	-	13,38%
	Hubungan sudut pusat, panjang busur, luas juring dan tembereng	-	36,26%	7,64%
Bangun ruang	Luas permukaan	44,98%	20,30%	-
	Volum	70,03%	40,49%	25,66%
	Balok kerangka		23,24%	
	Rata-rata	50,57%	33,71%	12,20%

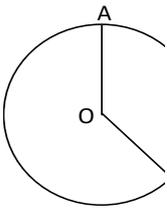
Dari tabel 3, kemampuan peserta didik dalam menyelesaikan soal-soal bangun datar dan bangun ruang masih rendah terutama dalam menyelesaikan soal-soal penalaran. Kemampuan peserta didik tertinggi dalam menyelesaikan soal level pengetahuan dan pemahaman materi volum sebesar 70,03%. Contoh soal, pedoman penskoran, dan persentase daya serap materi pelajaran

berdasarkan level kognitif, berdasarkan domain konten disajikan pada tabel 4 berikut.

Bangun Datar

Berikut disajikan contoh soal, pedoman penskoran, dan persentase daya serap materi berdasarkan level kognitif.

Soal 1:

	<p>Perhatikan gambar di samping! O adalah pusat lingkaran. Jika panjang OA = 21 cm dan panjang busur kecil AB = 44 cm, berapakah luas juring kecil AB?</p>	<p>Domain konten : Geometri Materi : Luas busur Level Kognitif : Pengetahuan dan Pemahaman Daya Serap : 36,26%</p>
---	--	---

Soal tersebut mengukur kemampuan peserta didik dalam menghitung luas juring, jika diketahui panjang jari-jari dan panjang busur kecil AB. Tingkat kesukaran soal ini termasuk kategori

sedang. Sebanyak 36,26% peserta didik dapat menjawab benar soal tersebut. Berikut rubrik yang digunakan untuk menskor jawaban peserta didik.

Tabel 5. Rubrik penskoran model 1

No. Soal	Kriteria/Kunci Jawaban	Skor
	Luas lingkaran = $\frac{22}{7} \times 21^2 = 1.386 \text{ cm}^2$ Keliling lingkaran = $2 \times \frac{22}{7} \times 21 = 132 \text{ cm}$ Luas juring kecil AB = $\frac{44}{132} \times 1.386 = 462 \text{ cm}^2$	1
Skor Maksimum		1

Soal yang diskor menggunakan rubrik penskoran model 1 hanya mempunyai dua kategori yaitu 0 dan 1. Jika jawaban peserta didik sesuai dengan rubrik penskoran diberi skor 1 dan jika tidak sesuai diberi skor 0. Berdasarkan penskoran

yang dilakukan menggunakan rubrik penskoran model 1, rata-rata peserta didik menjawab benar soal tersebut sebesar 38,50%. Berikut rubrik penskoran model 2 yang digunakan untuk menskor soal yang sama.

Tabel 6. Rubrik penskoran model 2

No. Soal	Skor	Kriteria/Kunci Jawaban
	2	Jawaban benar 462 cm ² beserta langkah penyelesaian Luas lingkaran = $\frac{22}{7} \times 21^2 = 1.386 \text{ cm}^2$ Keliling lingkaran = $2 \times \frac{22}{7} \times 21 = 132 \text{ cm}$ Luas juring kecil AB = $\frac{44}{132} \times 1.386 = 462 \text{ cm}^2$
	1	Ada indikasi perhitungan $\frac{22}{7} \times 21^2$, $2 \times \frac{22}{7} \times 21$, luas juring kecil AB $\frac{\text{panjang busur kecil}}{\text{panjang busur besar}} \times \text{luas lingkaran}$, tetapi ada kesalahan dalam perhitungan
	0	Jawaban benar 462 tanpa langkah penyelesaian, jawaban dan langkah penyelesaian salah/kosong/ada coretan.

Soal yang diskor menggunakan rubrik penskoran model 2 mempunyai tiga kategori yaitu 2, 1, dan 0 tergantung pedoman penskoran dari tiap butir soal. Skor 2 menunjukkan jika jawaban peserta didik benar disertai langkah penyelesaian, skor 1 bila jawaban siswa benar sebagian, dan skor 0 bila sebagian kecil jawaban peserta didik benar

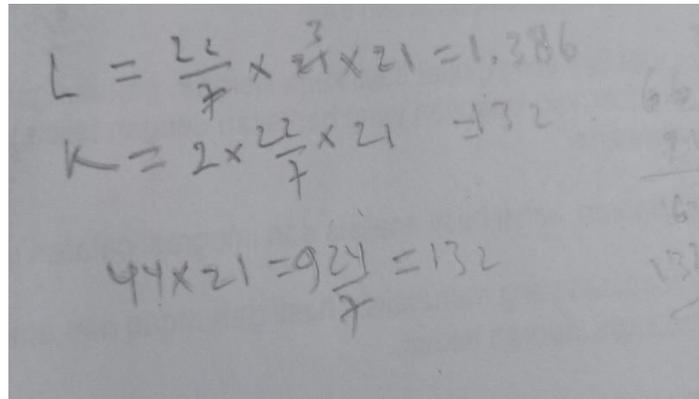
atau jawaban peserta didik salah. Berdasarkan penskoran yang dilakukan menggunakan rubrik penskoran model 2, rata-rata peserta didik menjawab benar soal tersebut sebesar 33,85%. Berikut rubrik penskoran model 3 yang digunakan untuk menskor soal yang sama.

Tabel 7. Rubrik penskoran model 3

No. Soal	Kriteria/Kunci Jawaban	Skor
	Luas lingkaran = $\frac{22}{7} \times 21^2 = 1.386 \text{ cm}^2$	1
	Keliling lingkaran = $2 \times \frac{22}{7} \times 21 = 132 \text{ cm}$	1
	Luas juring kecil AB = $\frac{44}{132} \times 1.386 = 462 \text{ cm}^2$	1
Skor Maksimum		3

Soal yang diskor menggunakan rubrik penskoran model 3 mempunyai skor maksimum berbeda dari setiap rubrik butir soal tergantung pedoman penskorannya. Berdasarkan hasil penskoran yang dilakukan menggunakan rubrik penskoran model 3, rata-rata peserta didik menjawab benar soal tersebut sebesar 36,43%.

Walaupun soal ini termasuk pemahaman, sebanyak 63,57% masih merasa kesulitan atau tidak dapat menyelesaikan soal tersebut. Sebagian kecil peserta didik hanya dapat menghitung luas lingkaran dan keliling tanpa dapat menghitung luas juring.



Sejalan dengan hasil penelitian yang dilakukan oleh Dessy P. dan Wahyu S.(2020), peserta didik hanya menghafal rumus dan belum memahami konsep dengan baik. Siswa kurang

cermat membaca dan memahami maksud soal dengan baik. Berikut contoh soal yang mengukur kemampuan level kognitif aplikasi.

Soal 2:

<p>Dani dan Abi masing-masing mempunyai kawat sama panjang. Dari kawat tersebut, Abi membuat sebuah lingkaran, sedangkan Dani membuat sebuah persegi. Bila luas lingkaran yang dibuat Abi adalah 154 cm^2, berapakah luas persegi yang dibuat Dani?</p>	<p>Domain konten : Geometri Materi : Luas Level Kognitif : Aplikasi Daya Serap : 33,41%</p>
--	---

Soal tersebut mengukur kemampuan peserta didik dalam menyelesaikan soal cerita yang mengukur dua kemampuan, yaitu menghitung keliling lingkaran untuk mendapatkan panjang kawat dan selanjutnya menghitung luas persegi dengan terlebih dahulu menghitung panjang sisi persegi. Sebanyak 33,41% peserta didik dapat menjawab benar soal tersebut. Tingkat kesukaran

soal tersebut termasuk kategori sedang. Peserta didik kurang memahami soal yang disajikan dalam bentuk cerita serta konsep atau rumus yang digunakan untuk menyelesaikannya. Menurut Kusnita, Rismayanti, Indah Puspitasari (2018), persentase kemampuan peserta didik dalam mengerjakan soal tes kemampuan pemecahan masalah masih tergolong rendah.

Soal 3:

<p>Perhatikan gambar!</p> <p>Diketahui O pusat lingkaran, $AB \parallel CD$, panjang garis $AB = 20 \text{ cm}$, $CD = 12 \text{ cm}$, dan panjang busur kecil $CD = 22 \text{ cm}$. Berapakah luas daerah yang diarsir?</p>	<p>Domain konten : Geometri Materi : Luas Level Kognitif : Penalaran Daya Serap : 9,32%</p>
--	---

Soal tersebut mengukur kemampuan peserta didik dalam menghitung luas tembereng lingkaran yang diketahui panjang diameter dan panjang garis

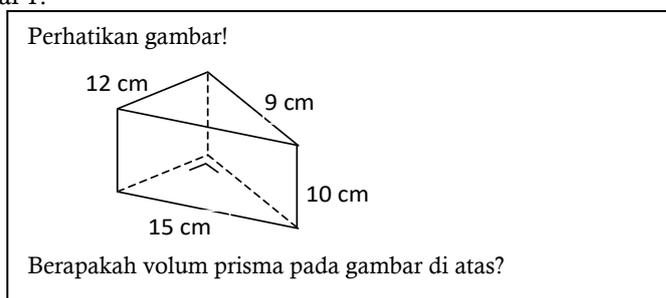
yang sejajar dengan diameternya. Untuk menghitung luas tembereng, peserta didik harus mencari luas juring dan luas segitiga OCD.

Sebanyak 9,32% peserta didik dapat menjawab benar soal tersebut. Soal tersebut termasuk ke dalam soal dengan tingkat kesukaran sulit.

Bangun Ruang

Berikut disajikan contoh soal, pedoman penskoran, dan persentase daya serap materi berdasarkan level kognitif.

Soal 1:



Domain konten	: Geometri
Materi	: Volum
Level Kognitif	: Pengetahuan dan Pemahaman
Rerata Daya Serap	: 78,36%

Soal tersebut mengukur kemampuan pemahaman peserta didik menghitung volum prisma yang alasnya berbentuk segitiga. Untuk menghitung volum prisma, peserta didik pertama-tama menghitung luas segitiga, kemudian hasilnya

dikalikan dengan tinggi prisma. Sebanyak 78,36% peserta didik dapat menjawab benar soal tersebut. Soal tersebut termasuk ke dalam soal dengan tingkat kesukaran mudah.

Tabel 8. Rubriks penskoran model 1

No. Soal	Kriteria/Kunci Jawaban	Skor
	Luas alas segitiga = $\frac{1}{2} \times 12 \times 9 = 54 \text{ cm}^2$	1
	Volum prisma = luas alas x tinggi = $54 \times 10 = 540 \text{ cm}^3$	1
	Skor Maksimum	1

Berdasarkan penskoran yang dilakukan menggunakan rubriks penskoran model 1, rata-rata peserta didik menjawab benar soal tersebut sebesar

77,88%. Rubrik penskoran model 2 disajikan pada tabel 9 berikut.

Tabel 9. Rubriks penskoran model 2

No. Soal	Skor	Kriteria/Kunci Jawaban
2		Jawaban benar 540 cm^3 beserta langkah penyelesaian Luas alas segitiga = $\frac{1}{2} \times 12 \times 9 = 54 \text{ cm}^2$ Volum prisma = luas alas x tinggi = $54 \times 10 = 540 \text{ cm}^3$
1		Perhitungan luas segitiga $\frac{1}{2} \times 12 \times 9 = \dots$ dan volum prisma benar = $\dots \times 10$, tetapi salah dalam penyelesaian akhir
0		Jawaban benar 540 tanpa langkah penyelesaian, jawaban dan langkah penyelesaian salah/kosong/ada coretan.

Berdasarkan hasil penskoran yang dilakukan menggunakan rubriks penskoran model 2, rata-rata peserta didik menjawab benar soal tersebut sebesar

77,99%. Rubrik penskoran model 3 disajikan pada tabel 10 berikut.

Tabel 10. Rubriks penskoran model 3

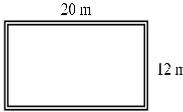
No. Soal	Kriteria/Kunci Jawaban	Skor
	Luas alas segitiga = $\frac{1}{2} \times 12 \times 9 = 54 \text{ cm}^2$	1
	Volum prisma = luas alas x tinggi = $54 \times 10 = 540 \text{ cm}^3$	1
	Skor Maksimum	2

Berdasarkan hasil penskoran yang dilakukan menggunakan rubrik penskoran model 2, rata-rata peserta didik menjawab benar soal tersebut sebesar

79,20%. Berikut contoh soal yang mengukur kemampuan level kognitif aplikasi.

Contoh 2:

Di bawah ini adalah gambar sebuah kolam renang dilihat dari atas dan samping.



Dilihat dari atas



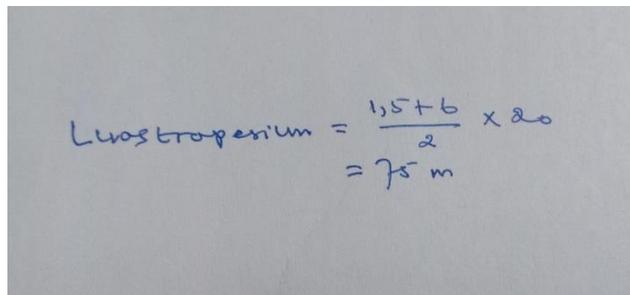
Terdalam
Dilihat dari samping

Berapakah volum kolam renang tersebut?

Domain konten : Geometri
Materi : Volum
Level Kognitif : Aplikasi
Daya Serap : 18,17%

Soal tersebut mengukur kemampuan peserta didik dalam menghitung volum suatu kolam renang. Kolam renang tersebut disajikan dalam bentuk sketsa dilihat dari sisi atas dan sisi samping.

Rata-rata peserta didik menjawab benar soal tersebut sebesar 18,17%. Soal tersebut termasuk kategori sulit.



Peserta didik sulit memahami pernyataan atau memahami gambar yang terdapat dalam soal tersebut. Sebagian kecil peserta didik menggunakan

konsep luas trapesium tetapi salah dalam perhitungan. Sebagian besar peserta didik tidak dapat menjawab soal tersebut.

Contoh 3:

Luas tiga sisi yang berbeda dari sebuah balok adalah 3 cm^2 , 6 cm^2 , dan 8 cm^2 . Berapakah volum balok tersebut?

Jawab:

Domain konten : Geometri
Materi : Volum
Level Kognitif : Penalaran
Daya Serap : 20,34%

Soal tersebut mengukur kemampuan peserta didik dalam menghitung volum balok yang diketahui ukuran luas ketiga sisi-sisinya. Rata-rata peserta didik menjawab benar soal tersebut sebesar

20,34% dan sebanyak 79,66% tidak menjawab soal tersebut (kosong). Soal tersebut termasuk ke dalam soal dengan tingkat kesukaran sulit. Perhitungan reliabilitas tes disajikan pada tabel 11 berikut.

Tabel 11. Reliabilitas tes matematika bentuk soal uraian berdasarkan rubrik penskoran model 1, 2, dan 3

Rubrik	$\sum i$	$\sum i^2$	$(\sum i)^2$	$\sum X^2$	$\sum Y^2$	S_e^2	S_s^2	Reliabilitas
Model 1	3.507	3.507	12.299.049	32.407	774.329	0,1645	0,5238	0,6859
Model 2	7.290	13.946	53.144.100	137.394	3.352.988	0,5942	1,9974	0,7025
Model 3	12.351	44.239	152.547.201	410.539	8.997.605	2,2186	7,3619	0,6986

N = 452, k = 22

Dari hasil perhitungan reliabilitas tes menggunakan ketiga rubrik dapat disimpulkan bahwa ketiga model rubrik menghasilkan nilai reliabilitas tes relatif sama dengan reliabilitas tertinggi adalah 0,7026 berasal dari rubrik penskoran model 2.

SIMPULAN

Dari hasil pembahasan tersebut dapat disimpulkan bahwa kemampuan peserta didik dalam menyelesaikan soal uraian materi bangun datar dan bangun ruang masih rendah, yaitu di bawah 50%. Hanya soal uraian level kognitif pengetahuan dan pemahaman yang mengukur

materi volum mendapatkan kemampuan di atas 70%. Kemampuan siswa menyelesaikan soal yang mengukur level penalaran sebesar 12,20%. Peserta didik kurang memahami maksud dari soal, kurang menguasai konsep, salah melakukan perhitungan, serta mengubah soal cerita ke dalam model matematika. Reliabilitas tes uraian matematika berdasarkan rubrik penskoran model 1 adalah 0,6859, model 2 adalah 0,7025, dan reliabilitas tes uraian matematika berdasarkan rubriks penskoran model 3 adalah 0,6986. Dari ketiga reliabilitas tes tersebut dapat disimpulkan bahwa reliabilitas tes dari ketiga rubrik penskoran relatif sama dan termasuk kategori tinggi atau reliabel.

REFERENSI

- Anastasi, Anne, and Urbina, Susana. 1988. *Psychological Testing 6th*. New York: Macmillan Publishing Company.
- Arikunto, Suharsimi. 2003. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Azwar, Saifuddin. 2000. *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Azwar, Saifuddin. 1987. *Tes Prestasi*. Yogyakarta: Liberty.
- Bond, T. G., & Fox, C. M. 2007. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. (2nd Ed.). Mahwah: Lawrence Erlbaum Associates, Publishers.
- Brenan, Robert L. 2006. *Educational Measurement*. Washington: American Council on Education Praeger.
- Charles D. Hopkins dan Richard L. Antes. 1979. *Classroom Testing Construction*. Illionos: F. E. Peacock.
- Crosker, Linda, James Algina. 1986. *Introduction To Classical & Modern Test Theory*. Holt, Rinehart dan Winston. Inc. New York.
- Departemen Pendidikan dan Kebudayaan. 1993. *Penyusunan, Penskoran dan Penggunaan Tes Prestasi Belajar Bentuk Uraian*. Jakarta: Pusat Pengujian.
- Dessy P. dan Wahyu S..2020. Analisis Kesalahan Siswa SMP Kelas IX dalam Menyelesaikan Soal pada materi Pokok Bangun Ruang Sisi Datar. *Jurnal Ilmiah Pendidikan Matematika*, Volume 7 No. 1, Maret 2020. <http://ejournal.stkipbbm.ac.id/index.php/mtk/article/view/427/375>
- DeMars, C. 2010. *Item Response Theory*. New York: Oxford University Press, Inc.
- Fahmi and Idris HM Noor. 2019. The analysis of students' error and difficulty level of mathematics essay test. *International Journal of Educational Policy Research and Review* Vol. 6 (4), July, 2019 <https://www.journalissues.org/IJEPRR/>
- Gronlund, N.E., Linn, D. R.L. 1990. *Measurement and assessment in teaching*. New York: Macmillan Publishing Company.
- Kusnita D, Rismayanti, dan Indah P. 2018. Analisis Kemampuan dan pemecahan Masalah Matematik Siswa MTs pada Materi Bangun Ruang Sisi Datar. *Jurnal Pembelajaran Matematika Inovatif*. Volume 1, No. 5, September 2018. <https://journal.ikipsiliwangi.ac.id/index.php/jpmi/article/view/1695>
- Kementerian Pendidikan dan Kebudayaan. 2019. *Panduan Penilaian Tes Tertulis*. Jakarta: Puspendik.
- Kementerian Pendidikan dan Kebudayaan, 2012. *Buku Panduan Pemanfaatan Pamer UN-2012*, Jakarta: Pusat Penilaian Pendidikan.
- Lee J. Cronbach. 1960. *Essential of Psychological Testing*, 3th ed, New York: Harper & Row.
- Litwin, Marks S. 1995. *How to Measure Survey Reliability and Validity* London: Sage Publications.
- Mary J. Allen, Wendy M. Yen. 1979. *Introduction Measurment Theory*, California: Montrey.
- Mehrens, WA and I. J. Lehman. 1991. *Measurment and Evaluation in Educational and Psychology*. New York: Hold, Rincchart and Winston, Inc.,
- M. Chabib Thoha. 1996. *Teknik Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada.
- Muhammad Khumaedi. 2012. *Reliabilitas Instrumen Penelitian Pendidikan*. Semarang: UNNES.
- JPTM. *Jurna Pendidikan Teknis Mesin* Vol. 12 No. 1 Juni 2012. <https://journal.unnes.ac.id/nju/index.php/JPTM/article/view/5273>
- Nitko, Anthony J. 1996. *Educational Assessment of Students*. Engliwood Cliffs, Prentise-Hall, Inc.,
- OECD. 2010. *TIMSS Field-test Scoring Guides Grade 8*. Boston Collge.
- Suryabrata, Sumadi. 1987. *Pengembangan Tes Hasil Belajar*. Jakarta: Rajawali.
- Triton,P.B. (2006). *SPSS 13.0 Terapan: Riset Statistik Parametrik*. Yogyakarta: Andi.
- Umar dan Hayat. 2000. *Penelitian Penggunaan Soal Bentuk Obyektif dan Uraian di Sekolah*. Jakarta: Pusat Penelitian dan Pengembangan Sistem Pengujian.