



## Analisis Karakteristik Butir Soal Ujian Nasional Matematika SMP/MTS Berdasarkan Pendekatan Teori Tes Klasik

Heni Purwati<sup>1\*</sup>, Heri Retnawati<sup>2</sup>, Jailani<sup>3</sup>, Trie Hartiti Retnowati<sup>4</sup>

<sup>1</sup>Jurusan Penelitian dan Evaluasi Pendidikan, Pascasarjana, Universitas Negeri Yogyakarta, DI Yogyakarta

<sup>2,3</sup>Jurusan Pendidikan Matematika, Fakultas MIPA, Universitas Negeri Yogyakarta, DI Yogyakarta

<sup>4</sup>Jurusan Pendidikan Seni Rupa, Fakultas Bahasa dan Seni, Universitas Negeri Yogyakarta, DI Yogyakarta

\*email korespondensi: [henipurwati\\_01.2017@student.uny.ac.id](mailto:henipurwati_01.2017@student.uny.ac.id)

**Received** : 2 Agustus 2020; **Revised** : 26 Januari 2021; **Accepted** : 11 Juni 2021; **Published** : 19 September 2021

### ABSTRAK

Penelitian ini bertujuan untuk menganalisis karakteristik butir soal ujian nasional matematika SMP/MTs berdasarkan teori klasik. Penelitian ini merupakan penelitian eksploratif dengan pendekatan kuantitatif. Data yang digunakan pada penelitian adalah data sekunder dengan mengambil data ujian nasional mata pelajaran matematika tahun 2015 untuk Propinsi Papua. Pada penelitian ini, responden dikelompokkan menjadi dua kelompok yaitu kelompok tinggi dan kelompok rendah berdasarkan jumlah skor yang diperoleh setiap peserta tes. Untuk menganalisis karakteristik butir soal yang terdiri dari indeks kesukaran, daya pembeda dan reliabilitas, dilakukan dengan bantuan program Excel dan QUEST. Hasil analisis menunjukkan nilai rerata dari indeks kesukaran butir kelompok tinggi berada pada kategori sedang dan kelompok rendah pada kategori sukar, nilai rerata daya pembeda kelompok tinggi dan rendah menunjukkan kategori tidak baik, yang artinya soal tersebut tidak dapat membedakan kemampuan peserta didik berkemampuan tinggi dan yang berkemampuan rendah hal ini dapat disebabkan peserta tes menjawab soal dengan cara menebak sehingga tidak memberikan hasil yang sesuai. Reliabilitas butir kelompok tinggi dan rendah memiliki nilai indeks yang mendekati + 1,00 yang menunjukkan reliabilitas yang tinggi. Berdasarkan hasil analisis butir soal tersebut, dapat disimpulkan bahwa soal tersebut layak digunakan. Dasiil hasil analisis, diberikan saran untuk soal pengujian di tiap-tiap daerah supaya disesuaikan dengan kondisi lingkungan dan peserta didik di masing-masing daerah.

**Kata-kata kunci:** karakteristik butir; teori tes klasik ujian nasional

### PENDAHULUAN

Pendidikan merupakan salah satu penentu kemajuan suatu negara sehingga sudah selayaknya jika pendidikan selalu menjadi perhatian oleh semua pihak mulai dari kebijakan pelaksanaan pendidikan sampai dengan evaluasi. Salah satu evaluasi yang dilaksanakan secara nasional sejak tahun 1950 an adalah ujian nasional yang beberapa kali mengalami perubahan nama dan kebijakan. Kebijakan mengenai pelaksanaan ujian nasional pada hakikatnya merupakan evaluasi terhadap hasil dan proses belajar (Hidayah, 2013, p. 35). Pelaksanaan Ujian Nasional sampai tahun 2020 membuat pro dan kontra karena dianggap tidak sepenuhnya mengukur kompetensi siswa secara menyeluruh karena hanya melihat kemampuan kognitif saja.

Berdasarkan Peraturan Pemerintah Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan pada pasal 68 disebutkan bahwa hasil dari ujian nasional digunakan sebagai salah satu pertimbangan untuk pemetaan mutu program dan/atau satuan pendidikan . Pemetaan tersebut bertujuan untuk melihat kualitas pendidikan tiap-tiap daerah di seluruh Indonesia. Kualitas pendidikan dapat dilihat dari perencanaan, pelaksanaan, dan penilaian serta evaluasi pembelajaran yang dilaksanakan pada pengajaran mata pelajaran di sekolah. Salah satu mata pelajaran yang dipelajari di semua jenjang pendidikan adalah matematika. Hal ini karena matematika dianggap unik dan memiliki peran yang tak tergantikan dalam pembentukan penalaran rasional manusia, dan mendorong kecerdasan manusia (Ministry of Education, 2003). Oleh karena matematika selalu menjadi perhatian karena merupakan dasar dari banyak bidang ilmu yang lain, dimana penguasaan kompetensinya diharapkan dapat membantu mencapai tujuan sesuai yang diharapkan sehingga dapat meningkatkan kualitas pendidikan.

Peningkatan kualitas pendidikan di Indonesia terus ditingkatkan melalui peningkatan kualitas pembelajaran dan kualitas sistem penilaiannya. Proses pembelajaran dan penilaian adalah dua hal yang saling berkaitan dan tidak dapat dipisahkan, sistem pembelajaran yang baik akan menghasilkan kualitas belajar yang baik. Kualitas pembelajaran ini dapat dilihat dari hasil assesmennya. Sistem penilaian atau assesmen yang baik dapat mendorong pendidik untuk mengajar dengan lebih baik dan memotivasi peserta

didik untuk belajar dengan lebih baik. Dalam upaya peningkatan kualitas pendidikan diperlukan sistem assesmen yang baik (Mardapi, 2016). Salah satu sistem penilaian yang pernah dilakukan secara nasional adalah ujian nasional.

Ujian Nasional merupakan salah satu bentuk assesmen untuk mengukur kemampuan kognitif siswa. Alat ukur yang sesuai untuk mengukur aspek kognitif terutama dengan jumlah responden yang sangat banyak karena bersifat nasional adalah soal berbentuk pilihan majemuk. Soal pilihan majemuk yang digunakan harus memiliki kualitas yang bagus dan hasilnya dapat mencerminkan kemampuan siswa yang diukur. Soal yang berkualitas adalah soal yang dapat memberikan informasi setepat-tepatnya sesuai dengan tujuan soal yang dibuat. Untuk memperoleh informasi yang tepat perlu dilakukan analisis butir secara empirik.

Analisis butir soal secara empirik bertujuan untuk memperoleh informasi mengenai karakteristik setiap butir soal. Analisis terhadap butir secara empirik dapat dilakukan dengan dengan dua pendekatan, yaitu dengan pendekatan teori tes klasik dan teori modern atau teori respon butir (Retnawati, 2016). Analisis butir pada penelitian ini menggunakan pendekatan teori tes klasik.

Teori tes klasik atau disebut sebagai teori skor murni klasik (Allen, Mary J & Yen, 1979) didasarkan pada penggunaan model matematika sederhana dengan menghubungkan antara skor amatan (X), skor sebenarnya (T), dan skor kesalahan (*error score*) (E) (Mardapi, 2016). Jika dituliskan dengan pernyataan matematis, maka kalimat tersebut menjadi

$$X = T + E \dots\dots\dots(1)$$

Kesalahan pengukuran yang dilakukan pada teori tes klasik dianggap sebagai kesalahan yang tidak sistematis atau acak. Kesalahan ini merupakan perbedaan dari skor amatan yang diperoleh dengan skor amatan yang diharapkan yang dilihat secara teoritis. Kesalahan pengukuran secara sistematis tidak termasuk kesalahan pengukuran.

Terdapat beberapa dugaan pada teori tes klasik diantaranya; dugaan pertama adalah skor kesalahan pengukuran tidak berkaitan dengan skor sebenarnya, dugaan kedua adalah skor kesalahan tidak berkorelasi dengan skor sebenarnya dan skor-skor kesalahan pada tes-tes yang lain untuk peserta tes yang sama dan dugaan ketiga adalah rata-rata dari skor kesalahan dianggap sama dengan nol (Retnawati, 2016). Pendekatan teori tes klasik dalam penelitian ini digunakan untuk menentukan karakteristik butir soal suatu tes yaitu tingkat kesukaran, daya pembeda, dan reliabilitas tes.

Tingkat kesukaran butir soal disimbolkan dengan  $p_i$ , yaitu termasuk salah satu paramater butir soal yang sangat bermanfaat dalam penganalisan suatu tes. Jika nilai  $p_i$  mendekati 0, maka soal tersebut dikatakan terlalu sulit. Sedangkan jika nilai  $p_i$  mendekati 1, maka soal tersebut dikatakan terlalu mudah, sehingga perlu dieliminasi atau direvisi. Soal yang terlalu sulit atau terlalu mudah tidak dapat membedakan kemampuan seorang siswa satu dengan yang lainnya (Retnawati, 2016).

Tingkat kesukaran butir yang dinyatakan dalam bentuk proporsi memiliki rentang diantara 0 sampai dengan 1. Nilai terletak ditengah-tengah pada  $P_i=0,50$  dikatakan memiliki kriteria sedang yang memisahkan butir yang mudah dari butir sukar. Butir dikatakan mudah jika nilai  $P_i > 0,50$  dan butir dikatakan sukar jika nilai  $P_i < 0,50$  (Naga, 1992). (Allen, Mary J & Yen, 1979) menyatakan bahwa secara umum indeks kesukaran butir yang baik terletak pada rentang 0,3 sampai 0,7. Pada rentang ini informasi tentang kemampuan siswa akan diperoleh secara maksimal. Apabila nilai indeks  $p_i < 0,30$  maka butir soal tersebut dikatakan sukar, apabila terletak pada interval  $0,30 \leq p_i \leq 0,70$  maka butir soal tingkat kesukaran dikatakan sedang, dan jika nilai jika  $p_i > 0,70$  butir soal tersebut dikatakan mudah.

Menentukan besarnya indeks kesukaran dari suatu butir soal dengan penskoran dikotomi, yaitu dengan membandingkan banyaknya peserta yang menjawab benar dengan jumlah peserta tes yang menjawab butir soal tersebut (Retnawati, 2016). Rumus tingkat kesukaran butir soal dapat dituliskan sebagai berikut;

$$p_i = \frac{\sum B}{N} \dots\dots\dots(2)$$

dimana:

- $p_i$  : tingkat kesukatan butir soal ke -i
- $\sum B$  : jumlah peserta tes yang menjawab butir soal dengan benar
- N : jumlah peserta tes yang menjawab butir soal

Daya pembeda atau daya beda butir soal yaitu kemampuan sesuatu butir soal untuk membedakan antara siswa yang berkemampuan tinggi dengan siswa yang berkemampuan rendah. Untuk menentukan

daya pembeda, dapat digunakan indeks korelasi *point biserial*. Terdapat beberapa cara yang dapat digunakan untuk mengetahui besarnya indeks daya pembeda yaitu 1) indeks diskriminasi, 2) indeks korelasi, dan 3) indeks keselarasan (Kartowagiran, 2009). Indeks daya pembeda dikatakan baik jika memiliki nilai lebih dari atau sama dengan 0,3. Indeks daya pembeda suatu butir yang nilainya kecil mengakibatkan butir tersebut tidak dapat membedakan siswa yang berkemampuan tinggi dengan siswa yang berkemampuan rendah (Retnawati, 2016). Koefisien korelasi untuk suatu butir tes dinyatakan dengan rumus :

$$r_{pbis} = \left[ \frac{\bar{X}_i - \bar{X}}{s_x} \right] \sqrt{\frac{p_i}{1-p_i}} \dots\dots\dots(3)$$

dimana:

- $r_{pbis}$  : koefisien korelasi point biserial
- $X_i$  : variabel kontinu
- $\bar{X}_i$  : rerata skor X untuk peserta yang menjawab benar butir tersebut
- $\bar{X}$  : rerata skor X
- $s_x$  : standar deviasi dari skor X
- $p_i$  : proporsi jumlah peserta tes yang menjawab benar butir tersebut

Reliabilitas pada teori tes klasik dapat dipahami dengan beberapa cara. Suatu tes dikatakan memiliki nilai reliabilitas yang tinggi jika skor tampak tes tersebut berkorelasi tinggi dengan skor sebenarnya atau apat juga dikatakan sebagai seberapa tingginya korelasi antara skor tampak pada dua tes yang paralel (Azwar, 2012). Reliabilitas adalah untuk mengetahui sejauh mana hasil suatu proses pengukuran dapat dipercaya. (Nitko, A. J & Brookhart, 2011) "*Reliability is the degree to which students results remain consistent over replications of an assessment procedure.*" Reliabilitas adalah sejauh mana hasil tes siswa tetap konsisten setelah dilakukan berulang kali, sesuai dengan prosedur penilaian.

Reliabilitas ( $\rho$ ) suatu tes pada umumnya diekspresikan secara numerik dalam bentuk koefisien yang besarnya  $-1,00 \leq \rho \leq +,00$ . Koefisien tinggi menunjukkan reliabilitas tinggi. Sebaliknya, jika koefisien suatu skor suatu tes rendah maka reliabilitas tes rendah. Jika suatu reliabilitas sempurna, berarti koefisien tersebut  $+1,00$  (Retnawati, 2016). Estimasi reliabilitasnya dianalisis dengan menggunakan persamaan koefisien alpha sebagai berikut:

$$\alpha = \left( \frac{k}{k-1} \right) \left( 1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right) \dots\dots\dots(4)$$

dengan;

- $\alpha$  : koefisien reliabilitas
- $k$  : banyaknya butir soal
- $\sum \sigma_i^2$  : jumlah varians butir
- $\sigma_t^2$  : varians skor total

Manfaat yang diperoleh dari hasil analisis butir soal adalah untuk mengetahui dan meningkatkan kualitas butir soal serta meningkatkan kompetensi guru dan hasil belajar siswa. Analisis butir soal bermanfaat untuk mengetahui suatu butir soal bermanfaat sesuai dengan yang diharapkan, mengevaluasi kemampuan siswa dan dapat menjadi bahan diskusi di kelas, memberikan informasi kepada guru mengenai materi yang belum dikuasai oleh siswa, memberi masukan kepada pihak-pihak yang berwenang untuk mengevaluasi dan mengembangkan kurikulum, memperbaiki materi yang diujikan, serta meningkatkan kemampuan menyusun soal (Nitko, A. J & Brookhart, 2011)(Fitrianawati, 2017). Analisis butir soal dapat dilakukan untuk semua jenis tes baik yang terstandarisasi seperti Ujian Nasional (UN), maupun tes yang bersifat lokal seperti tes yang dibuat oleh guru di tingkat sekolah masing-masing.

**EKSPERIMEN**

Penelitian ini termasuk penelitian eksploratif dengan menggunakan pendekatan kuantitatif dalam rangka mengetahui karakteristik butir soal. Data yang digunakan pada penelitian ini merupakan data sekunder yaitu data ujian nasional mata pelajaran matematika tahun 2015 tingkat SMP/MTs Propinsi Papua sebanyak 8002 siswa. Kemudian diambil sebagai sampel penelitian sebanyak 4400 siswa. Pada penelitian ini siswa dikelompokkan berdasarkan kelompok tinggi dan kelompok rendah berdasarkan jumlah skor yang diperoleh setiap siswa. Untuk menganalisis karakteristik butir yaitu indeks kesukaran, daya pembeda dan reliabilitas tes dilakukan dengan bantuan program *Excel* dan program *QUEST*.

**HASIL DAN DISKUSI**

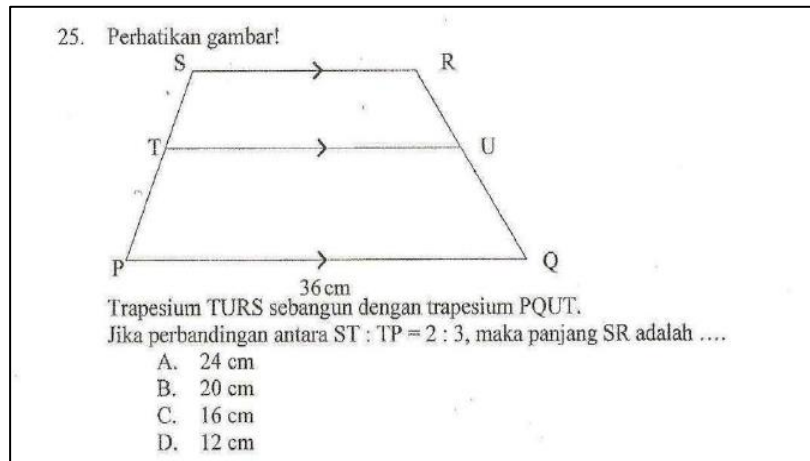
Berdasarkan hasil analisis dari 40 butir dengan bantuan program *Excel* dan pogram *QUEST* diperoleh informasi indeks kesukaran, daya pembeda, serta reliabilitas tes kelompok tinggi dan rendah. Hasil analisis disajikan pada **Tabel 1**.

**Tabel 1.** Indeks Kesukaran, Daya Pembeda, dan Reliabilitas

Butir Soal Nomor	Kelompok Tinggi				Kelompok Rendah			
	P	Kategori	rbis	Kategori	P	Kategori	rbis	Kategori
1	0.92	Mudah	0.14	Tidak baik	0.47	Sedang	0.09	Tidak baik
2	0.63	Sedang	0.22	Tidak baik	0.30	Sedang	0.1	Tidak baik
3	0.55	Sedang	0.18	Tidak baik	0.36	Sedang	0.04	Tidak baik
4	0.85	Mudah	0.19	Tidak baik	0.29	Sukar	0.1	Tidak baik
5	0.82	Mudah	0.21	Tidak baik	0.31	Sedang	0.08	Tidak baik
6	0.87	Mudah	0.16	Tidak baik	0.32	Sedang	0.16	Tidak baik
7	0.60	Sedang	0.11	Tidak baik	0.29	Sukar	0.05	Tidak baik
8	0.72	Mudah	0.22	Tidak baik	0.26	Sukar	0.08	Tidak baik
9	0.82	Mudah	0.18	Tidak baik	0.31	Sedang	0.11	Tidak baik
10	0.82	Mudah	0.11	Tidak baik	0.27	Sukar	0.13	Tidak baik
11	0.76	Mudah	0.21	Tidak baik	0.33	Sedang	0.09	Tidak baik
12	0.75	Mudah	0.23	Tidak baik	0.28	Sukar	0.03	Tidak baik
13	0.83	Mudah	0.17	Tidak baik	0.36	Sedang	0.11	Tidak baik
14	0.65	Sedang	0.21	Tidak baik	0.25	Sukar	0.1	Tidak baik
15	0.77	Mudah	0.14	Tidak baik	0.28	Sukar	0.1	Tidak baik
16	0.65	Sedang	0.22	Tidak baik	0.25	Sukar	0.1	Tidak baik
17	0.77	Mudah	0.24	Tidak baik	0.16	Sukar	0.09	Tidak baik
18	0.73	Mudah	0.15	Tidak baik	0.25	Sukar	0.1	Tidak baik
19	0.81	Mudah	0.16	Tidak baik	0.30	Sedang	0.11	Tidak baik
20	0.75	Mudah	0.17	Tidak baik	0.19	Sukar	0.12	Tidak baik
21	0.87	Mudah	0.18	Tidak baik	0.38	Sedang	0.12	Tidak baik
22	0.60	Sedang	0.18	Tidak baik	0.27	Sukar	0.1	Tidak baik
23	0.74	Mudah	0.24	Tidak baik	0.22	Sukar	0.09	Tidak baik
24	0.80	Mudah	0.26	Tidak baik	0.23	Sukar	0.11	Tidak baik
25	0.36	Sedang	0.15	Tidak baik	0.25	Sukar	0.08	Tidak baik
26	0.84	Mudah	0.26	Tidak baik	0.32	Sedang	0.14	Tidak baik
27	0.65	Sedang	0.32	Baik	0.21	Sukar	0.08	Tidak baik
28	0.67	Sedang	0.27	Tidak baik	0.21	Sukar	0.11	Tidak baik
29	0.65	Sedang	0.28	Tidak baik	0.21	Sukar	0.06	Tidak baik
30	0.54	Sedang	0.23	Tidak baik	0.19	Sukar	0.08	Tidak baik
31	0.62	Sedang	0.19	Tidak baik	0.37	Sedang	0.04	Tidak baik
32	0.66	Sedang	0.18	Tidak baik	0.30	Sukar	0.07	Tidak baik
33	0.54	Sedang	0.29	Tidak baik	0.26	Sukar	0.07	Tidak baik
34	0.45	Sedang	0.21	Tidak baik	0.22	Sukar	0.05	Tidak baik
35	0.69	Sedang	0.19	Tidak baik	0.29	Sukar	0.05	Tidak baik
36	0.52	Sedang	0.23	Tidak baik	0.28	Sukar	0.08	Tidak baik
37	0.66	Sedang	0.24	Tidak baik	0.25	Sukar	0.08	Tidak baik
38	0.88	Mudah	0.16	Tidak baik	0.61	Sedang	0.1	Tidak baik
39	0.67	Sedang	0.23	Tidak baik	0.32	Sedang	0.04	Tidak baik
40	0.68	Sedang	0.28	Tidak baik	0.27	Sukar	0.09	Tidak baik
Rerata	0,703	Sedang	0,204	Tidak baik	0,287	Sukar	0,088	Tidak baik
Reliabilitas			0,99	Tinggi			0,98	Tinggi

Mencermati hasil tersebut, indeks kesukaran butir kelompok tinggi berdasarkan kriteria, maka dari 40 butir soal terdapat 22 butir (55%) yang termasuk dalam kategori mudah, terdapat 18 butir (45%) dengan kategori sedang dan tidak ada butir dengan kategori sukar. Rerata indeks kesukaran butir soal tes sebesar 0,703 dalam kategori sedang, artinya semua butir tes bisa memberikan informasi tentang kemampuan siswa.

Jika ditelusuri lebih jauh lagi pada kelompok tinggi maka butir yang memiliki nilai indeks kesukaran paling kecil adalah butir soal nomor 25 yaitu sebesar 0,36. Hal ini, menunjukkan bahwa dari 2200 peserta tes/siswa hanya 785 atau 35,68% yang menjawab benar. Di bawah ini ditampilkan butir soal nomor 25.



Gambar 1. Butir Soal Nomor 25

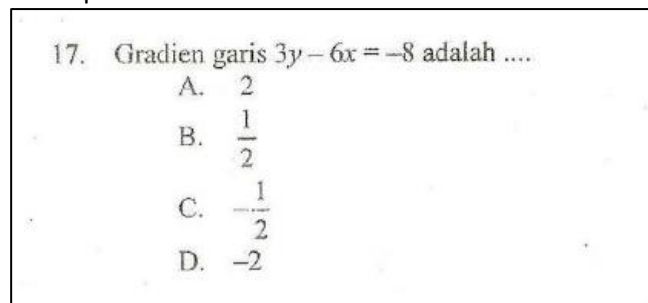
Berdasarkan hasil analisis, faktor penyebab letak kesulitan siswa dalam menyelesaikan soal di atas yaitu; *pertama* belum bisa menghubungkan dan menerapkan rumus geometri pada bangun datar, *kedua* belum paham dengan bentuk-bentuk bangun datar berdasarkan karakteristik sifat-sifatnya, *ketiga* belum bisa mencari, menentukan atau mendapatkan nilai yang tidak diketahui dalam soal, dan *keempat* belum memahami konsep perbandingan.

Indeks daya pembeda pada kelompok tinggi dari 40 butir soal hanya 1 butir (2,5%) dalam kategori baik yaitu butir soal nomor 27, artinya butir ini dapat membedakan siswa yang kemampuannya tinggi dan siswa yang kemampuannya rendah. Namun, secara rerata indeks daya pembeda butir secara keseluruhan sebesar  $0,204 < 0,3$ , artinya semua butir tersebut tidak dapat membedakan siswa yang kemampuannya tinggi dan siswa yang kemampuannya rendah.

Reliabilitas butir tes untuk kelompok tinggi sebesar 0,99 menunjukkan reliabilitas tinggi artinya hasil tes siswa mampu menggambarkan sejauh mana konsistensi hasil ukur yang dilakukan berulang-ulang terhadap responden yang memiliki gejala sama dengan alat ukur yang sama.

Indeks kesukaran butir kelompok rendah berdasarkan kriteria, tidak ada butir dengan kategori mudah, terdapat 14 butir (35%) dengan kategori sedang dan ada 26 butir (65%) dengan kategori sukar. Rerata indeks kesukaran butir soal tes sebesar 0,287 dalam kategori sukar, artinya semua butir tes tidak bisa memberikan informasi tentang kemampuan siswa.

Pada kelompok rendah, butir yang memiliki nilai indeks kesukaran paling kecil adalah butir soal nomor 17 yaitu sebesar 0,16. Hal ini, menunjukkan dari 2200 peserta tes hanya 361 atau 16,41% yang menjawab benar. Di bawah ini ditampilkan butir soal nomor 17.



Gambar 2. Butir Soal Nomor 17

Berdasarkan hasil analisis, faktor penyebab letak kesulitan siswa dalam menyelesaikan soal di atas yaitu; *pertama* belum bisa menerapkan prinsip gradien garis lurus, *kedua* belum bisa mengubah persamaan yang ada kedalam bentuk persamaan  $y = mx \pm c$ , *ketiga* kurang penguasaan dasar-dasar aljabar, dan *keempat* belum memahami unsur-unsur pada persamaan linier satu variabel.

Indeks daya pembeda pada kelompok rendah dari 40 butir soal tidak ada butir dalam kategori baik. Rerata indeks daya pembeda sebesar  $0,088 < 0,3$ , artinya semua butir tersebut tidak dapat membedakan siswa yang kemampuannya tinggi dan siswa yang kemampuannya rendah.

Reliabilitas butir tes untuk kelompok rendah sebesar 0,98 menunjukkan reliabilitas tinggi artinya hasil tes siswa mampu menggambarkan sejauh mana konsistensi hasil ukur yang dilakukan berulang-ulang terhadap responden yang memiliki gejala sama dengan alat ukur yang sama.

## **KESIMPULAN**

Berdasarkan hasil analisis dan pembahasan diperoleh simpulan bahwa karakteristik butir soal matematika ujian nasional tingkat SMP/MTs tahun 2015 Provinsi Papua berdasarkan teori tes klasik ditinjau dari indeks kesukaran, menunjukkan nilai rerata dari indeks kesukaran butir kelompok tinggi berada pada kategori sedang dan kelompok rendah pada kategori sukar. Rerata indeks daya pembeda kelompok tinggi dan rendah pada kategori tidak baik yang artinya soal tersebut tidak dapat membedakan kemampuan peserta didik berkemampuan tinggi dan yang berkemampuan rendah hal ini dapat disebabkan peserta tes menjawab soal dengan cara menebak sehingga tidak memberikan hasil yang sesuai, reliabilitas butir tes kelompok tinggi dan rendah memiliki nilai reliabilitas yang tinggi.

Mencermati hasil penelitian ini, dapat direkomendasikan penelitian ini baru mengetahui karakteristik butir dengan pendekatan teori klasik, sehingga terkait dengan penelitian di masa mendatang, dapat dilakukan dengan pendekatan teori respon butir sehingga dapat diketahui kemampuan masing-masing siswa dan mendapatkan hasil yang lebih detail lagi.

## **DAFTAR PUSTAKA**

Allen, Mary J & Yen, W. M. (1979). *Introduction to Measurement Theory*. Waveland Press.

Azwar, S. (2012). *Reliabilitas dan Validitas*. Pustaka Pelajar.

Fitrianawati, M. (2017). *Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru dan Hasil Belajar Peserta Didik*. 282–295. <https://publikasiilmiah.ums.ac.id/handle/11617/9117>

Hidayah, N. (2013). Ujian nasional dalam Perspektif Kebijakan Publik. *Jurnal Pencerahan*, 1, 35–40. <https://doi.org/https://doi.org/10.13170/jp.7.1.2053>

Kartowagiran, B. (2009). *PENGANTAR TEORI TES KLASIK (TTK)\**.

Mardapi, D. (2016). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Parama Publishing.

Ministry of Education. (2003). *Mathematics Curriculum Standards (Experimental Version)*. <http://ncm.gu.se/media/kursplaner/andralander/kinagym.pdf>

Naga, D. S. (1992). *Pengantar Teori Skor pada Pengukuran Pendidikan*. Gunadarma.

Nitko, A. J & Brookhart, S. M. (2011). *Educational assessment of atudents (6th ed.)*. Pearson Education, Inc.

Retnawati, H. (2016). *Validitas, Reliabilitas, dan Karakteristik Butir*. Parama Publishing Company.