



**Submitted:** 2024-09-02  
**Published:** 2024-11-15

---

## ANALISIS KUALITAS SOAL BERDASARKAN TAKSONOMI BLOOM REVISI

Afifah Fitria<sup>a)</sup>, Zubaidah<sup>b)</sup>, Edy Yusmin<sup>c)</sup>, Halini<sup>d)</sup>, Ade Mirza<sup>e)</sup>

a,b,c,d,e) Program Studi Pendidikan Matematika, Universitas Tanjungpura, Pontianak, Indonesia

**Corresponding Author:** [f1041191016@student.untan.ac.id](mailto:f1041191016@student.untan.ac.id)<sup>a</sup>  
[zubaidah.r@fkip.untan.ac.id](mailto:zubaidah.r@fkip.untan.ac.id), [edy.yusmin@fkip.untan.ac.id](mailto:edy.yusmin@fkip.untan.ac.id), [halini.fkip14@gmail.com](mailto:halini.fkip14@gmail.com),  
[ade.mirza@fkip.untan.ac.id](mailto:ade.mirza@fkip.untan.ac.id)

Article Info	Abstract
<p><b>Keywords:</b> Question Quality; Cognitive Domain Mapping; End of Semester Assessment; Revised Bloom's Taxonomy</p>	<p><i>This study aims to evaluate the quality of test items by considering validity, reliability, difficulty level, discriminating power, and effectiveness of distractors. Furthermore, this study also presents and describes the mapping of cognitive process dimensions and knowledge dimensions based on the revised Bloom's taxonomy. This research employs a mixed methods approach, integrating quantitative and qualitative techniques. The population of this study includes 218 student answer sheets and mathematics test items from the Semester End Assessment (PAS) for grade VIII in the second semester of the 2022-2023 academic year at SMPN 5 Pontianak. To determine a representative sample size from the population, the researcher collected 141 student answer sheets using the Slovin formula. A scoring rubric and a table containing the classification of test items mapped to the cognitive domain mapping were used as instruments. Data analysis techniques were conducted in stages, both quantitatively and qualitatively. The research findings indicate that the test items used have a fairly good level of validity, good reliability, and most of them have a moderate level of difficulty. The items also able to differentiate students with different abilities and have a good level of distractor effectiveness. However, in terms of cognitive processes, it has not been mapped</i></p>

*proportionally. The items only map levels C1, C2, and C3. Thus, higher-order cognitive process abilities cannot be measured by these items. The mapping of knowledge dimension levels has been included proportionally.*

**Kata Kunci:**

Kualitas Soal;  
Pemetaan Ranah  
Kognitif; Penilaian Akhir  
Semester; Taksonomi  
Bloom Revisi

---

Penelitian ini dimaksudkan mengevaluasi kualitas butir-butir soal dengan mempertimbangkan validitas, reliabilitas, tingkat kesukaran, tingkat daya pembeda, serta efektivitas pengecohnya. Selain itu, penelitian ini juga memaparkan dan menguraikan pemetaan dimensi proses kognitif dan dimensi pengetahuan yang didasarkan pada taksonomi Bloom revisi. Penelitian ini menggunakan metodologi campuran (*mixed methods*) yang mengintegrasikan teknik kuantitatif dan kualitatif. Populasi dalam penelitian ini meliputi 218 lembar jawaban siswa serta butir-butir soal matematika dari Penilaian Akhir Semester (PAS) kelas VIII pada semester genap tahun ajaran 2022-2023 di SMPN 5 Pontianak. Untuk menentukan jumlah sampel yang representatif dari populasi tersebut, peneliti mengumpulkan 141 lembar jawaban siswa dengan menerapkan rumus Slovin. Pedoman penilaian dan tabel yang berisi klasifikasi butir-butir pertanyaan yang dipetakan ke dalam pemetaan ranah kognitif merupakan instrumen yang digunakan. Teknik analisis data dilakukan secara bertahap, baik secara kuantitatif maupun kualitatif. Temuan penelitian menunjukkan bahwa soal-soal yang digunakan memiliki tingkat validitas yang cukup baik, reliabilitas yang baik, dan sebagian besar memiliki tingkat kesulitan sedang. Soal-soal juga mampu membedakan siswa dengan kemampuan yang berbeda, serta memiliki tingkat efektivitas pengecoh yang baik. Namun, pada aspek proses kognitif belum dipetakan secara proporsional. Soal hanya memetakan tingkatan C1, C2, dan C3. Dengan demikian, kemampuan proses kognitif tingkat tinggi belum dapat diukur dengan soal-soal tersebut. Pemetaan tingkat dimensi pengetahuan telah memuat secara proporsional.

## PENDAHULUAN

Peningkatan kualitas pendidikan sangat bergantung pada pengembangan sistem penilaian yang efektif dan terpercaya (Sudjana, 2014). Berdasarkan Peraturan Menteri Pendidikan Nasional Nomor 21 Tahun 2022 tentang Standar Penilaian Pendidikan, penilaian tidak hanya sekadar proses pengumpulan data, tetapi juga mencakup analisis yang mendalam untuk mengidentifikasi hasil belajar, capaian pembelajaran, dan kebutuhan siswa. Hal ini menempatkan penilaian sebagai salah satu komponen fundamental dalam penyelenggaraan pendidikan (Mansyur, 2015).

Penilaian Akhir Semester (PAS) merupakan salah satu bentuk evaluasi kunci untuk mengukur pemahaman siswa terhadap materi serta kemampuan mereka dalam menerapkan konsep dan keterampilan yang telah dipelajari (Arbiatin & Mulabbiyah, 2020; Azizah dkk, 2021). PAS memungkinkan siswa untuk mengevaluasi tingkat pencapaian mereka, yang dapat memberikan dorongan bagi mereka untuk terus meningkatkan pengetahuan dan pemahaman mereka (Himmah, 2019). Namun demikian, untuk mencapai tujuan ini, instrumen penilaian yang digunakan harus memenuhi standar kualitas tertentu agar mampu memberikan hasil yang valid dan reliabel.

Dalam praktiknya, pengembangan soal PAS sering kali masih menghadapi berbagai kendala terkait dengan kualitas soal yang dibuat. Menurut Nurrada (2020), meskipun soal disusun oleh tim guru mata pelajaran, seringkali standar baku dan uji coba yang diperlukan untuk menjamin kualitas soal belum dilakukan dengan memadai. Waty (2017) menekankan bahwa tanpa adanya standar yang kuat, soal yang digunakan untuk PAS belum tentu mampu mengukur kompetensi siswa secara akurat. Herawati (2021) menambahkan bahwa soal yang dirancang dengan kurang baik tidak hanya mengurangi validitas dan reliabilitas hasil evaluasi, tetapi juga dapat menyebabkan bias dalam pengukuran capaian pembelajaran.

Selain itu, meskipun berbagai penelitian telah membahas kualitas soal matematika dari sudut pandang teknis, seperti validitas, reliabilitas, tingkat kesukaran, dan daya pembeda, masih terdapat gap penelitian dalam memahami bagaimana aspek-aspek ini berhubungan dengan kompleksitas kognitif berdasarkan taksonomi Bloom revisi. Dalam penelitian ini belum banyak penelitian yang mengeksplorasi hubungan antara kualitas soal matematika dengan tingkatan berpikir siswa sesuai kerangka kerja taksonomi Bloom. Padahal, pemetaan soal ke dalam dimensi kognitif ini sangat penting untuk

memastikan bahwa siswa tidak hanya mampu mengingat informasi, tetapi juga dapat menganalisis dan mengevaluasi konsep-konsep yang dipelajari.

Taksonomi Bloom yang telah direvisi terdapat klasifikasi tujuan pembelajaran dan kemampuan kognitif yang berbeda. Taksonomi ini mengelompokkan proses kognitif ke dalam enam kategori, yaitu mengingat (C1), memahami (C2), menerapkan (C3), menganalisis (C4), mengevaluasi (C5), dan mencipta (C6). Meskipun model ini telah banyak digunakan dalam pengembangan instrumen penilaian, beberapa penelitian menunjukkan bahwa belum semua soal yang dikembangkan mampu mengukur seluruh tingkat kognitif siswa secara komprehensif. Ropii dan Fahrurrozi (2017) serta Wulandari dan Pramusinto (2020) menggarisbawahi pentingnya mengevaluasi soal tidak hanya berdasarkan kualitas teknisnya tetapi juga berdasarkan kapasitasnya untuk mengukur kemampuan berpikir siswa pada berbagai level kognitif.

Penelitian yang telah dilakukan cenderung lebih fokus pada aspek teknis, seperti analisis statistik untuk menilai validitas dan reliabilitas soal. Namun, aspek yang berkaitan dengan bagaimana soal-soal ini dapat mengukur kemampuan berpikir tingkat tinggi siswa belum banyak diteliti secara empiris. Hal ini mengindikasikan adanya kebutuhan untuk penelitian yang lebih mendalam yang menggabungkan

analisis kualitas soal dengan pemetaan dimensi kognitif sesuai taksonomi Bloom revisi.

Malawi & Maruti (2016) dan Arikunto (2016) menunjukkan bahwa kualitas soal dapat dilihat dari beberapa indikator utama seperti validitas, reliabilitas, tingkat kesulitan, dan daya pembeda. Namun, keterkaitan antara indikator-indikator ini dengan taksonomi Bloom dalam konteks soal matematika untuk siswa SMP masih jarang dieksplorasi. Bahkan, kajian literatur oleh Rurisfiani dkk (2019) menunjukkan bahwa dimensi proses kognitif yang tinggi, seperti analisis dan mencipta, masih jarang ditemui dalam soal PAS matematika, yang lebih banyak berfokus pada kemampuan kognitif dasar.

Oleh karena itu, penelitian ini bertujuan untuk melakukan analisis menyeluruh terhadap kualitas soal matematika dalam PAS Genap kelas VIII dengan mempertimbangkan dimensi kognitif pada taksonomi Bloom revisi. Dengan pendekatan ini, diharapkan penelitian dapat memberikan kontribusi yang signifikan dalam pengembangan instrumen penilaian yang lebih efektif, yang pada akhirnya dapat mendukung peningkatan kualitas pendidikan secara keseluruhan.

Penelitian ini diharapkan dapat memberikan wawasan baru bagi guru dan pengembang kurikulum dalam memahami pentingnya analisis kualitas soal berdasarkan standar teknis sekaligus

pemetaan kognitif. Hasil penelitian ini juga diharapkan dapat menjadi dasar bagi peningkatan mutu penilaian di berbagai tingkatan pendidikan. Dengan menciptakan instrumen penilaian yang lebih valid dan reliabel, siswa akan memiliki kesempatan lebih besar untuk mengembangkan kemampuan berpikir kritis mereka sesuai dengan tuntutan pembelajaran abad ke-21.

## **METODE**

Untuk mengumpulkan, menganalisis, dan menginterpretasikan data serta untuk desain penelitian dan pengumpulan data, penelitian ini memakai metodologi campuran (*mixed methods*) yang mengintegrasikan teknik kuantitatif dan kualitatif (Iskandar dkk., 2021). Dengan demikian, penelitian ini mengombinasikan data dan analisis yang bersifat kuantitatif dan kualitatif. Bentuk penelitian yang digunakan ialah desain sekuensial eksplanatori. Menurut Iskandar, dkk. (2021) desain eksplanatori merupakan satu diantara desain metode campuran yang bersifat sekuensial, di mana proses penelitian dilakukan dalam dua tahap, yakni tahap kuantitatif dan kualitatif. Melalui bentuk penelitian ini, peneliti ingin mengumpulkan dan menganalisis data kuantitatif dan kualitatif guna untuk memberikan gambaran yang lebih lengkap dan menghasilkan pemahaman yang lebih

komprehensif serta mendalam tentang kualitas soal dan pemetaan ranah kognitif soal yang ingin diteliti.

Validitas butir soal, reliabilitas butir soal, tingkat kesukaran, tingkat daya pembeda, tingkat efektivitas pengecoh, dan persentase pemetaan pada tingkat dimensi proses kognitif dan dimensi pengetahuan semuanya dianalisis dalam analisis data kuantitatif butir soal. Para peneliti menggunakan metode Korelasi Point Biserial untuk pilihan ganda dalam analisis validitas butir soal. Untuk pertanyaan pilihan ganda dalam analisis reliabilitas butir soal, para peneliti menggunakan rumus Kuder-Richardson Formula 20 (KR-20). Dengan menggunakan masing-masing formulasi tes ini, tingkat kesulitan, tingkat daya pembeda, dan tingkat efektivitas pengecoh dianalisis. Kajian terhadap data kualitatif dilakukan oleh peneliti dengan menggunakan taksonomi Bloom yang telah direvisi untuk mengukur aspek pengetahuan dan proses kognitif siswa. Berdasarkan aspek pengetahuan (pengetahuan faktual, konseptual, dan prosedural) dan proses kognitif (C1, C2, C3, C4, C5, dan C6), peneliti mengkaji dan mengkategorikan setiap butir soal ke dalam tingkatan kemampuan berpikir mahasiswa. Selain itu, peneliti mendeskripsikan secara rasional sesuai dengan keadaan yang ada, dan dua orang dosen telah mengevaluasi deskripsi

mereka. Selanjutnya, para peneliti menentukan proporsi tingkat pemetaan yang berkaitan dengan proses kognitif dan dimensi pengetahuan, serta menarik kesimpulan tentang hasil pemetaan di setiap tingkat.

Para peneliti memakai acuan penilaian, yang mencakup indikator evaluasi untuk setiap tingkat proses kognitif dan dimensi pengetahuan, sebagai alat pengumpul data selama tahap analisis kualitatif. Tujuan penggunaan pedoman ini adalah untuk membantu peneliti dalam memberikan penilaian yang objektif terhadap kriteria-kriteria yang telah ditetapkan. Sebelum menggunakan pedoman penelitian tersebut, terlebih dahulu divalidasi oleh dua dosen. Setelah pedoman penelitian telah divalidasi oleh dosen tersebut, peneliti menyajikan hasil analisis menggunakan tabel hasil analisis klasifikasi pemetaan ranah kognitif dan pemetaan ranah kognitif butir soal dan divalidasi oleh dua dosen.

Populasi penelitian ini ialah butir soal matematika pada PAS kelas VIII tahun pelajaran 2022/2023 semester genap di SMPN 5 Pontianak dan 218 lembar jawaban siswa. Selanjutnya, peneliti mengambil sampel dengan cara probability sampling, khususnya dengan menggunakan metode pengambilan sampel acak dasar. Sebanyak 141 lembar jawaban siswa dikumpulkan oleh para peneliti dengan menggunakan rumus Slovin untuk menentukan jumlah

sampel yang akan diambil dari populasi yang ada.

## HASIL DAN PEMBAHASAN

### Analisis Soal Berdasarkan Kualitas Soal PAS

#### Uji Validitas Butir

Ukuran standar yang dapat mengukur tingkat ketepatan, kebenaran, keabsahan pada setiap butir soal dalam mengukur kemampuan siswa ialah adanya uji validitas butir. Kemampuan siswa yang akan diukur akan dapat ditentukan oleh tingkat validitas butir soal yang cukup baik (Arikunto, 2016). Pertanyaan yang valid memperlihatkan seberapa tepat dan akurat sebuah pertanyaan dalam mengukur kemampuan atau pengetahuan siswa, yang merupakan persyaratan untuk butir-butir pertanyaan yang baik.

Analisis validitas butir soal PAS Genap Matematika kelas VIII tahun pelajaran 2022/2023 di SMPN 5 Pontianak menggunakan rumus Point Biserial Correlation ( $r_{bis}$ ), dengan memperoleh  $r_{bis}$  di setiap butir soal selanjutnya mengambil keputusan tiap butir soal yang valid maupun invalid dengan membandingkan  $r_{bis}$  dengan  $r_{tabel}$ . Soal dapat dinyatakan valid jika  $r_{bis} > r_{tabel}$  dan soal dapat dinyatakan invalid jika sebaliknya. Nilai  $r_{tabel}$  dapat dilihat pada tabel  $r$ ,  $df = n - 2$  dengan  $n$  ialah jumlah siswa dan  $\alpha = 0,05$ . Sehingga  $r_{tabel}$  dengan  $df = 141 - 2 = 139$ , dan  $\alpha = 0,05$  adalah 0,1654.

Berdasarkan temuan analisis validitas butir soal PAS dari 40 butir soal secara keseluruhan ada 29 butir soal (72,5%) yang valid dengan nomor soal 16, 17, 18, 19, 20, 22, 23, 27, 28, 29, 32, 34, 35, 36, 37, 38, dan 39. Disisi lain, butir soal yang invalid ada 11 butir soal (27,5%) berkategori invalid dengan nomor soal 9, 10, 11, 21, 24, 25, 26, 30, 31, 33, dan 40. Dengan ini, butir soal yang valid lebih dominan dibandingkan butir soal yang invalid.

Dalam hal validitas butir soal, temuan penelitian ini hampir sama dengan penelitian Azizah dkk. (2021) di mana 70% soal pilihan ganda termasuk dalam kelompok valid dan 30% sisanya tidak valid. Ini memperlihatkan kalau mayoritas pertanyaan bisa diandalkan dan mampu mengukur data yang relevan. Hasil ini sesuai dengan gagasan yang dikemukakan oleh Sudijono (2013) yang menyatakan bahwa ketepatan suatu butir soal dalam mengukur apa yang seharusnya diukur merupakan hal yang menentukan validitasnya.

Setelah peneliti meninjau data, 72,5% dari pertanyaan dianggap valid, yang menggambarkan bahwa butir-butir pertanyaan tersebut dapat secara akurat mengukur seberapa baik siswa telah memahami materi pelajaran yang diajarkan atau diujikan. Sebaliknya, 27,5% dari pertanyaan lainnya dianggap tidak valid,

yang menunjukkan bahwa pertanyaan-pertanyaan tersebut memiliki hubungan yang sangat lemah atau tidak berhubungan dengan materi pelajaran yang sedang diuji. Validitas butir soal dapat dipengaruhi oleh beberapa faktor, antara lain faktor respon siswa, faktor instrumen evaluasi, dan faktor administrasi evaluasi (Siyoto & Sodik, 2015).

Beberapa variabel jawaban siswa, seperti kecenderungan untuk menjawab dengan cepat, ketidakmampuan untuk membaca pertanyaan secara menyeluruh, kesalahan aritmatika, kurangnya ketelitian dalam bekerja, dan menebak-nebak saat menjawab, kemungkinan besar berdampak pada item pertanyaan yang tidak valid (Azizah dkk., 2021). Faktor pelaksanaan penilaian juga berpengaruh, misalnya kemungkinan siswa mencontek yang dapat memengaruhi validitas butir soal. Hal-hal seperti ini harus dihindari, contohnya dengan pengawasan ketat agar tidak ada siswa yang curang.

Faktor lain yang dapat memengaruhi validitas soal adalah tingkat kesukaran dan daya pembeda soal. Soal yang terlalu mudah atau terlalu sulit bagi sebagian besar siswa tidak dianggap valid (Riinawati, 2021). Selain itu, soal dengan daya pembeda rendah, yang tidak mampu membedakan antara siswa dengan tingkat pengetahuan atau keterampilan yang berbeda, juga

mengakibatkan soal tersebut tidak valid. Ini terjadi ketika banyak siswa menjawab dengan benar atau semua siswa menjawab salah.

Berdasarkan pemaparan sebelumnya, dapat disimpulkan bahwa soal PAS Genap Matematika kelas VIII tahun pelajaran 2022/2023 di SMPN 5 Pontianak tergolong soal yang valid dengan persentase 72,5%. Hal ini berarti soal PAS sudah cukup menjadi instrumen tes yang baik. Soal-soal yang tidak valid dapat digunakan kembali setelah dimodifikasi, namun tidak dapat dihapus dari evaluasi berikutnya. Soal-soal yang baik dan sah dapat disimpan di bank soal karena dapat menilai kemampuan siswa, memberikan informasi yang relevan, dan menilai tingkat kebenaran dan keabsahan yang dimiliki oleh setiap soal, sehingga soal-soal yang baik dapat ditambahkan ke bank soal dan digunakan pada evaluasi berikutnya. Sebaliknya, soal yang tidak valid tidak dapat ditambahkan ke bank soal.

### **Uji Reliabilitas Butir**

Standar yang digunakan untuk mengukur tingkat keajegan atau ketepatan serta keabsahan tinggi suatu butir soal berkaitan dengan kemampuannya memberikan hasil yang dapat diandalkan dalam mengukur kemampuan siswa. Reliabilitas butir soal berhubungan dengan kepercayaan terhadap soal tersebut dan kesesuaiannya dengan materi yang telah ditetapkan (Zainuddin, 2021). Soal yang

baik ialah soal yang reliabel, yakni soal yang selalu memberikan hasil yang konsisten ketika diujikan pada siswa yang sama di waktu atau kesempatan yang berbeda.

Analisis reliabilitas butir soal PAS Genap Matematika kelas VIII tahun pelajaran 2022/2023 di SMPN 5 Pontianak dilaksanakan dengan memakai rumus Kuder-Richardson Formula 20 (KR-20). Temuan analisis tersebut diperolehnya koefisien reliabilitas ( $r_{11}$ ) dan diinterpretasikan kedalam lima kategori tingkat reliabilitas yakni kategori sangat tidak reliabel/sangat buruk, tidak reliabel/buruk, cukup reliabel/cukup baik, berkategori reliabel/baik, sangat reliabel/sangat baik (Kusaeri & Suprananto, 2012).

Berdasarkan hasil analisis tersebut diperoleh nilai  $r_{11} = 0,736$  dengan rentang  $0,70 \leq r_{11} < 0,90$  yang memiliki kriteria butir soal yakni reliabel/baik. Temuan ini selaras dengan penelitian oleh Arbiatin dan Mulabbiyah (2020) yang memperlihatkan kalau soal pilihan ganda termasuk dalam kategori soal yang reliabel. Dengan demikian, butir soal ini cenderung konsisten dan dapat diandalkan dalam penilaian kemampuan siswa. Temuan ini sesuai dengan teori Arifin (2012) yang menyatakan bahwa reliabilitas butir merupakan salah satu karakteristik penting yang harus dimiliki oleh suatu tes agar dapat dianggap baik.

Ketika seperangkat pertanyaan telah dievaluasi dan terbukti reliabel, hal ini



menunjukkan bahwa pertanyaan tersebut dapat menjalankan fungsi yang dimaksudkan, memiliki tingkat validitas dan kebenaran yang tinggi, dan dapat digunakan berulang kali untuk menilai tujuan pembelajaran siswa yang sama karena reliabilitasnya yang kuat. Hal ini sejalan dengan pendapat Riinawati (2021) yang menyatakan bahwa suatu instrumen atau tes dikatakan reliabel jika diberikan kepada siswa yang sama beberapa kali dan menghasilkan temuan yang konsisten.

Kuantitas pertanyaan yang sah dapat berdampak pada tingkat reliabilitas butir soal. Ini selaras dengan gagasan yang diajukan oleh Arikunto (2016) yang menyatakan bahwa sebuah tes dapat saja valid tetapi tidak dapat dipercaya, atau sebaliknya. Ini memperlihatkan bahwa, sama halnya dengan nilai validitas yang tinggi akan menghasilkan nilai reliabilitas yang tinggi, nilai validitas butir soal yang rendah juga akan menghasilkan nilai reliabilitas butir soal yang rendah. Reliabilitas dan validitas butir soal berkorelasi, artinya koefisien reliabilitas suatu tes meningkat seiring dengan jumlah butir soal yang valid (Zein dkk., 2013).

### Uji Tingkat Kesukaran

Tingkat kesukaran soal merujuk pada probabilitas siswa dengan kemampuan tertentu untuk menjawab soal dengan

benar, yang biasanya dinyatakan dalam bentuk indeks. Indeks ini diperoleh dengan membandingkan jumlah siswa yang menjawab benar dengan total jumlah siswa yang mengikuti tes. Semakin banyak siswa yang menjawab benar, semakin rendah tingkat kesukaran soal tersebut. Dalam pengembangan soal, disarankan untuk menggunakan butir soal dengan tingkat kesukaran yang seimbang, dengan proporsi soal mudah, sedang, dan sukar ialah 3:4:3 (Sudjana, 2014).

Analisis tingkat kesukaran soal PAS Genap Matematika kelas VIII tahun ajaran 2022/2023 di SMPN 5 Pontianak dilakukan dengan menggunakan rumus dari Astuti (2022). Hasilnya diperoleh indeks tingkat kesukaran (ITK) yang diinterpretasikan ke dalam tiga kategori, yaitu soal sukar, sedang, dan mudah.

29 butir soal (72,5%) diklasifikasikan sebagai soal sedang dan 11 butir soal (27,5%) diklasifikasikan sebagai soal yang menantang berdasarkan tingkat kesulitan belajar. Hasilnya, butir soal dengan tingkat kesulitan sedang lebih banyak dibandingkan dengan butir soal yang mudah atau menantang. Kesimpulan ini hampir sama dengan penelitian Oktafiana dkk. (2020) yang menunjukkan bahwa tidak ada tingkat kesulitan yang mudah dan mayoritas soal memiliki tingkat kesulitan sedang. Dengan demikian, soal tersebut tidak memenuhi

syarat perbandingan soal yang baik dalam hal tingkat kesukaran.

Arifin (2012) menyatakan bahwa sebuah soal dianggap baik dan dapat secara akurat menilai kemampuan siswa jika tingkat kesulitannya seimbang, atau proporsional. Tingkat kesukaran yang berbeda memungkinkan adanya penyesuaian dengan tingkat kemampuan antar siswa (Sari, 2013) sehingga siswa dengan kemampuan beragam dapat diuji secara tepat, tanpa merasa terlalu mudah atau terlalu sulit.

Febriana (2019) menyatakan bahwa dalam mempertimbangkan tingkat kesukaran yang baik, bukan hanya disesuaikan dengan perbandingan distribusi tingkat kesukaran, tetapi juga disesuaikan dengan tujuan tes. Sebagai contoh, tes harus memiliki tingkat kesulitan yang lebih tinggi jika digunakan untuk seleksi. Oleh karena itu, pertanyaan yang terlalu mudah atau terlalu rumit tidak dapat menghasilkan banyak informasi.

### **Uji Daya Pembeda**

Uji daya pembeda dipakai guna menentukan sejauh mana setiap butir soal dapat membedakan kemampuan siswa, yaitu antara siswa yang telah memahami dan yang belum memahami materi yang diajarkan oleh guru. Kemampuan sebuah butir soal untuk membedakan antara siswa yang menguasai materi dengan yang tidak

berkorelasi dengan koefisien daya pembedanya.

Analisis tingkat daya pembeda butir soal PAS Genap Matematika kelas VIII tahun ajaran 2022/2023 di SMPN 5 Pontianak dilakukan dengan menggunakan rumus dari Kusaeri & Suprananto (2012). Hasilnya diperoleh indeks tingkat daya pembeda (ITK) yang diinterpretasikan ke dalam lima kategori, yakni semuanya tidak baik, tidak baik, cukup, baik, dan baik sekali.

Berdasarkan analisis, terdapat 1 butir soal (2,5%) yang dikategorikan sebagai sangat baik, 14 butir soal (35,0%) dikategorikan baik, dan 13 butir soal (32,5%) dikategorikan cukup. Butir soal ini mampu membedakan siswa dengan kemampuan rendah dan tinggi. Namun, terdapat 7 butir soal (17,5%) yang dikategorikan kurang baik dan 5 butir soal (12,5%) yang dikategorikan tidak baik, yang gagal membedakan tingkat kemampuan siswa. Secara keseluruhan, butir soal dengan daya pembeda tinggi (cukup, baik, dan sangat baik) lebih dominan dibandingkan dengan butir soal yang memiliki daya pembeda rendah.

Temuan ini hampir sama dengan penelitian Halik dkk. (2019) yang menunjukkan bahwa daya pembeda soal lebih banyak ditemukan pada level tinggi daripada level rendah. Sesuai dengan temuan Arifin (2012) kemampuan butir soal untuk membedakan kemampuan siswa meningkat seiring dengan daya pembeda

yang lebih besar. Sebuah soal memiliki kemampuan yang kuat untuk membedakan instruksi jika sebagian besar atau semua siswa dapat menjawabnya dengan benar.

Pentingnya memperhatikan tingkat daya pembeda dalam merancang soal evaluasi atau tes, karena soal dengan daya pembeda yang baik bisa memberikan informasi lebih akurat tentang kemampuan siswa dan dasar yang kuat untuk mengevaluasi kemampuan siswa (Ahmad, 2015). Selain itu, dengan memantau daya pembeda, guru dapat menilai keefektifan soal dalam mengukur pemahaman siswa.

Temuan analisis memperlihatkan kalau tingkat kesukaran mempengaruhi nilai daya pembeda, sesuai dengan teori Arikunto (2016) yang menyatakan bahwa hubungan antara tingkat kesukaran (P) dan daya pembeda (D) dapat diilustrasikan melalui rumus  $D=2P$ .

### Uji Efektivitas Pengecoh

Untuk menilai seberapa efektif pilihan jawaban pengecoh (*distractor*) dalam mengelabui siswa, terutama mereka yang memiliki kemampuan di bawah rata-rata, perlu dilakukan analisis efektivitas pengecoh tersebut. Jika pengecoh tidak ada atau tidak berfungsi dengan baik, siswa akan dapat menjawab pertanyaan dengan sangat mudah. Oleh karena itu, pengecoh memainkan peran yang sangat penting

dalam penilaian akhir semester berbentuk pilihan ganda.

Analisis tingkat efektivitas pengecoh butir soal PAS Genap Matematika kelas VIII tahun ajaran 2022/2023 di SMPN 5 Pontianak dilakukan dengan menggunakan rumus dari Ahmad (2015). Dari hasil tersebut, diperoleh indeks pengecoh (IP) pada masing-masing opsi jawaban yang kemudian diinterpretasikan ke dalam lima kategori, yakni sangat baik, baik, kurang baik, jelek, dan sangat jelek. Untuk mengevaluasi efektivitas pengecoh pada setiap butir soal, diterapkan kriteria yang disesuaikan dari skala Likert, terdiri dari lima kriteria: sangat baik, baik, cukup baik, kurang baik, dan tidak baik.

Soal PAS terdiri dari 40 soal pilihan ganda dengan empat pilihan jawaban (A, B, C, dan D). Tiga dari pilihan tersebut merupakan jawaban yang salah yang berfungsi sebagai pengecualian, sedangkan satu adalah jawaban yang benar atau kunci jawaban. Siswa yang tidak dapat memberikan jawaban yang tepat akan memilih pengecoh dalam jumlah yang setara dengan jawaban yang benar. Sebaliknya, pengecoh untuk soal yang buruk akan dipilih secara tidak seimbang. Jika proporsi siswa yang memilih pengecoh sama atau hampir sama dengan jumlah ideal, maka pengecoh tersebut dianggap baik.

Berdasarkan hasil analisis, terdapat 21 butir soal (52,5%) berkategori sangat baik, 7 butir soal (17,5%) berkategori baik, 11 butir soal (27,5%) berkategori cukup baik, dan 1 butir soal (2,5%) berkategori kurang baik. Tidak ada butir soal yang memiliki tingkat efektivitas pengecoh dalam kategori tidak baik.

Temuan penelitian ini hampir sama dengan temuan penelitian Halik dkk. (2019) yang meneliti tentang seberapa efektif pemicu pada soal pilihan ganda. Menurut Halik dkk. (2019) tidak ada soal yang memiliki pengecoh yang tidak memadai atau buruk, dan 70% soal memiliki pengecoh yang sangat baik, 22,5% memiliki pengecoh yang baik, dan 7,5% memiliki pengecoh yang buruk. Hal ini menunjukkan bahwa pemicu pertanyaan memiliki tingkat efektivitas yang lebih tinggi ketika digunakan dengan baik daripada ketika tidak. Jika dibandingkan dengan penelitian ini, penelitian Halik dkk. (2019) memiliki persentase kualitas yang lebih tinggi.

Ini merupakan butir soal yang bagus, dan siswa yang memberikan jawaban yang salah akan sama-sama memilih pengecoh. Di sisi lain, pengecoh dipilih secara tidak konsisten untuk soal yang kurang baik. Pengecoh yang berhasil harus menyerupai jawaban yang benar dan mendorong siswa yang kurang berhati-hati untuk memilihnya. Pilihan jawaban yang terlalu jelas berbeda dari kunci dapat mengurangi efektivitas pengecoh (Sari, 2013). Pengecoh yang baik

mencerminkan kesalahan umum yang sering dibuat siswa dalam memahami materi atau konsep tertentu (Zainuddin, 2021). Ini membantu mengidentifikasi kesalahpahaman siswa. Pengecoh efektif juga harus dapat membedakan siswa dengan tingkat kemampuan berbeda.

Penting mempertimbangkan efektivitas pengecoh agar soal dapat memberikan informasi relevan dan akurat tentang kemampuan siswa. Pengecoh yang baik membantu mengukur pemahaman dan pengetahuan siswa secara lebih mendalam, sehingga meningkatkan kualitas evaluasi atau tes.

### **Pemetaan Dimensi Proses Kognitif Soal Berdasarkan Taksonomi Bloom Revisi Dimensi Proses Kognitif**

Hasil analisis tingkat dimensi proses kognitif soal PAS Genap Matematika kelas VIII tahun ajaran 2022/2023 di SMPN 5 Pontianak yang berjumlah 40 soal pilihan ganda diperoleh 4 kategori dimensi proses kognitif, yaitu C1 "Mengingat", C2 "Memahami", C3 "Mengaplikasikan", dan C4 "Menganalisis". Kategori C1 terdapat 4 butir soal (10%), kategori C2 terdapat 15 butir soal (37,5%), kategori C3 terdapat 21 butir soal (52,5%), dan tidak ditemukan soal pada tingkatan C4, C5, dan C6.

Hanya ditemukan 4 butir soal kategori C1 karena soal pada tingkat ini umumnya untuk tingkat sekolah dasar dan jarang diberikan di SMP, serta dianggap terlalu

mudah untuk PAS. Tingkatan C1 sudah termasuk dalam tingkat kognitif yang lebih tinggi, karena untuk menyelesaikan masalah pada tingkat kognitif yang lebih tinggi, keterampilan yang terkait dengan C1 harus sudah dikuasai terlebih dahulu. Soal PAS didominasi oleh kategori C2 dan C3.

Kategori C2 mendorong siswa tidak hanya mengingat informasi, tetapi juga merangkum, menyimpulkan, menjelaskan, atau mengklarifikasi informasi yang dipelajari. Soal pada tingkat ini mendorong siswa berpikir lebih mendalam tentang konsep yang dipelajari.

Kategori C3 mendorong siswa mengaplikasikan pengetahuan dan konsep untuk memecahkan masalah, menjawab pertanyaan, atau menyelesaikan soal kontekstual. Tidak ditemukan butir soal pada tingkatan C4, C5, dan C6, padahal tingkatan ini akan mendorong siswa mengembangkan pemahaman yang kuat dan berpikir tingkat tinggi.

Temuan penelitian ini memperlihatkan kesamaan dengan studi yang dilakukan oleh Azizah dkk. (2021) ditinjau dari perspektif dimensi proses kognitif dalam soal pilihan ganda, yaitu 10% soal berada pada tingkat C1, 15% pada tingkat C2, 70% pada tingkat C3, dan 5% pada tingkat C4. Soal PAS tersebut memiliki tingkat ranah kognitif C4 yang rendah dan tidak memuat tingkat C5 dan C6. Hal ini menunjukkan pemetaan

dimensi proses kognitif dominan pada tingkatan yang rendah. Jika dibandingkan dengan penelitian ini, penelitian Azizah dkk. (2021) lebih dominan pada persentase tingkatan rendah (C1, C2, dan C3) yaitu 95%, sedangkan pada penelitian ini mencapai 100%.

Helmawati (2019) menyatakan bahwa soal yang disajikan harus mencakup pertanyaan yang menuntut siswa untuk berpikir pada tingkat tinggi, seperti menganalisis, mengevaluasi, dan mencipta. Proporsi soal tersebut harus seimbang. Menurut Giani dan Hiltrimartin (2015) proporsi soal yang baik adalah 30% untuk C1 dan C2, 40% untuk C3 dan C4, serta 30% untuk C5 dan C6. Namun, hasil analisis menunjukkan bahwa distribusi dimensi proses kognitif pada soal yang dianalisis belum memenuhi proporsi ideal ini, dengan beberapa persentase yang sangat berbeda dari yang direkomendasikan.

Pentingnya mengembangkan butir soal yang proporsional pada tingkat dimensi proses kognitif adalah agar guru dapat mengukur pemahaman peserta didik secara komprehensif. Soal dengan tingkat kognitif beragam memungkinkan pengukuran yang lebih holistik tentang kemampuan peserta dalam memahami dan menerapkan pengetahuan.

Dapat disimpulkan bahwa soal PAS Genap Matematika kelas VIII tahun ajaran 2022/2023 di SMPN 5 Pontianak tergolong soal yang memiliki pemetaan dimensi proses kognitif tidak proporsional, dengan 47,5% untuk C1 dan C2, 52,5% untuk C3 dan C4, serta 0% untuk C5 dan C6. Artinya, soal PAS ini belum cukup menjadi instrumen tes yang baik karena berada pada tingkatan proses kognitif rendah.

### **Dimensi Pengetahuan**

Hasil analisis tingkat dimensi pengetahuan soal PAS Genap Matematika kelas VIII tahun ajaran 2022/2023 di SMPN 5 Pontianak yang berjumlah 40 soal pilihan ganda diperoleh tiga kategori dimensi pengetahuan, yakni pengetahuan faktual, konseptual, dan prosedural. Kategori pengetahuan faktual terdapat 9 butir soal (22,5%), kategori pengetahuan konseptual terdapat 10 butir soal (25%), dan kategori pengetahuan prosedural terdapat 21 butir soal (52,5%).

Hanya ditemukan 22,5% soal pada kategori pengetahuan faktual karena soal pada tingkat ini digunakan untuk mengukur pemahaman dasar siswa tentang istilah, pengetahuan khusus, dan elemen dasar terkait suatu informasi. Pada tingkatan ini sebagai tingkat pengetahuan paling rendah, biasanya hanya sedikit soal diperlukan untuk menguji pemahaman siswa, karena dianggap sebagai fondasi bagi tingkat pemikiran yang lebih tinggi. Karena

kemampuan pada tingkat faktual harus dikomunikasikan secara tidak langsung untuk menjawab soal-soal pada tingkat yang lebih tinggi, maka tingkat pengetahuan faktual telah dibahas pada tingkat di atasnya.

Kategori pengetahuan konseptual pada penelitian ini sebesar 25% dengan 10 butir soal. Soal pada tingkat ini merupakan pengetahuan tentang ide-ide atau konsep-konsep yang lebih luas dengan melibatkan pengorganisasian informasi dan pemahaman tentang hubungan antar konsep. Pemahaman konsep yang mendalam memungkinkan siswa untuk lebih mudah mengaitkan dan menghubungkan berbagai informasi, serta menerapkan pengetahuan mereka dalam beragam konteks dan situasi. Oleh sebab itu, soal-soal yang dirancang untuk menguji pengetahuan konseptual secara efektif berfungsi untuk mengukur sejauh mana siswa memahami materi yang telah dipelajari.

Berdasarkan data yang ada, pengetahuan prosedural menempati posisi dominan, dengan persentase lebih dari 50%, yaitu 52,5%. Lebih dari separuh soal mengacu pada tingkat pengetahuan ini. Pengetahuan prosedural merujuk pada pemahaman tentang cara melakukan sesuatu, meliputi keterampilan, algoritma, teknik, dan metode, atau yang biasa disebut dengan prosedur (Alexander dalam Gunawan, 2012). Fokus dari soal PAS ini

sebagian besar pada metode, prosedur, atau pendekatan untuk menyelesaikan masalah matematika. Ini sejalan dengan jumlah tingkat penerapan kognitif yang disertakan dalam soal, yang sering kali membutuhkan tingkat keahlian prosedural tertentu untuk menyelesaikannya.

Temuan penelitian ini sangat mirip dengan penelitian Oktavia (2021) yang menunjukkan bahwa 35% soal pilihan ganda membutuhkan pengetahuan faktual, 40% membutuhkan pengetahuan konseptual, dan 25% membutuhkan pengetahuan prosedural. Tingkat pemetaan dimensi pengetahuan dominan pada tingkatan konseptual. Jika dibandingkan dengan penelitian ini, penelitian Oktavia (2021) lebih dominan pada persentase tingkatan konseptual 40%, sedangkan penelitian ini lebih dominan pada tingkatan prosedural 52,5%.

Menurut Helmawati (2019) dengan mencakup berbagai tingkatan dalam dimensi pengetahuan, soal-soal dapat memberikan gambaran yang lebih lengkap tentang pemahaman dan kemampuan siswa, serta membantu mengukur pemahaman siswa secara menyeluruh, mulai dari pemahaman dasar hingga kemampuan berpikir tingkat tinggi. Soal yang baik juga harus dirancang dengan jelas, relevan dengan tujuan pembelajaran, dan menghindari bias atau ambigu.

Dengan demikian, soal PAS Genap Matematika kelas VIII tahun ajaran 2022/2023 di SMPN 5 Pontianak tergolong soal yang memiliki pemetaan dimensi pengetahuan yang sudah proporsional, dengan 22,5% untuk tingkatan pengetahuan faktual, 25% untuk konseptual, dan 52,5% untuk prosedural. Artinya, soal PAS ini sudah cukup menjadi instrumen tes yang baik dan berada pada tingkatan pengetahuan tingkat tinggi.

## **PENUTUP**

### **Simpulan**

Analisis kualitas soal matematika pada penilaian akhir semester genap kelas VIII tahun pelajaran 2022/2023 menunjukkan hasil yang beragam. Validitas butir soal dominan sudah valid, dengan reliabilitas yang baik, namun masih terdapat beberapa soal yang invalid. Daya pembeda soal sebagian besar cukup hingga sangat baik, meskipun ada yang kurang baik. Tingkat kesukaran soal belum proporsional, didominasi soal sukar dan sedang tanpa adanya soal mudah. Pengecoh soal umumnya efektif, namun beberapa masih perlu perbaikan. Pemetaan dimensi proses kognitif berdasarkan taksonomi Bloom revisi belum proporsional, hanya mencakup tingkatan C1 hingga C3, tanpa soal pada tingkatan C4 sampai C6. Sementara itu, pemetaan dimensi pengetahuan sudah

proporsional, mencakup pengetahuan faktual, konseptual, dan prosedural.

### Saran

Soal yang sudah baik dapat dimasukkan ke bank soal dan digunakan kembali, soal kurang baik dapat direvisi, sedangkan soal tidak baik sebaiknya dibuang atau direvisi jika memungkinkan. Soal yang telah direvisi harus diujikan kembali dan jika sudah baik, dapat dimasukkan ke bank soal. Untuk pemetaan taksonomi Bloom revisi, sebaiknya soal disajikan secara proporsional pada setiap tingkatan kognitif agar kemampuan siswa diuji secara menyeluruh.

### DAFTAR PUSTAKA

- Ahmad, N. (2015). *Buku Evaluasi Pembelajaran*. Yogyakarta: Interpena.
- Arbiatin, E., & Mulabbiyah. (2020). Analisis Kelayakan Butir Soal Tes Penilaian Akhir Semester Mata Pelajaran Matematika Kelas VI Di SDN 19 Ampenan Tahun Pelajaran 2019/2020. *Jurnal PGMI*, 12(2), 146-171.  
<https://doi.org/10.20414/elmidad.v12i2.2627>
- Arifin, Z. (2012). *Evaluasi Pembelajaran*. Jakarta: Direktorat Jenderal Pendidikan Islam Kementerian Agama.
- Arikunto, S. (2016). *Dasar-dasar evaluasi pendidikan*. Jakarta: PT Rineka Cipta.
- Astuti, M. (2022). *Evaluasi Pendidikan*. Yogyakarta: Deepublish.
- Azizah, N., Hanifah, & Sumardi, H. (2021). Analisis Kualitas Dan Tingkat Kognitif Soal Matematika Penilaian Akhir Semester (PAS) Ganjil Kelas IX di SMPN 10 Kota Bengkulu Tahun 2020/2021. *Journal Mathematics Education Sigma [JMES]*, 2(2).  
<https://doi.org/10.30596/jmes.v2i2.7936>
- Febriana, R. (2019). *Evaluasi Pembelajaran*. Jakarta: Bumi Aksara.
- Giani, Z., & Hiltrimartin, C. (2015). Analisis Tingkat Kognitif Soal-Soal Buku Teks Matematika Kelas VII Berdasarkan Taksonomi Bloom. *Jurnal Pendidikan Matematika*, 66, 37-39.
- Halik, A. S., Mania, S., & Nur, F. (2019). Analisis Butir Soal Ujian Akhir Sekolah (UAS) Mata Pelajaran Matematika Pada Tahun Ajaran 2015/2016 SMP Negeri 36 Makassar. *Journal of Islamic Education*, 1(1), 11-17.
- Helmawati. (2019). *Pembelajaran dan Penilaian berbasis HOTS*. Bandung: PT Remaja Rosdakarya.
- Herawati, N. (2021). Kemampuan Guru Dalam Membuat Soal HOTS Dalam Ujian Tengah Semester. *Jurnal Pendidikan Guru Sekolah Dasar*, 10(6), 1689-1694.
- Himmah, W. I. (2019). Analisis Soal Penilaian Akhir Semester Mata Pelajaran Matematika Berdasarkan Level Berpikir. *Journal of Medives: Journal of Mathematics Education IKIP Veteran Semarang*, 3(1), 55.  
<https://doi.org/10.31331/medivesveteran.v3i1.698>



- Iskandar, Nehru, & Riantoni, C. (2021). *Metode Penelitian Campuran*. Pekalongan: PT. Nasya Expanding Management.
- Kusaeri, & Suprananto. (2012). *Pengukuran dan penilaian pendidikan*. Yogyakarta: Graha Ilmu.
- Malawi, I., & Maruti, E. S. (2016). *Evaluasi Pendidikan*. Magetan: CV Ae Media Grafika.
- Mansyur. (2015). *Asesmen Pembelajaran Sekolah*. Yogyakarta: Pustaka Pelajar.
- Nenty, H. J., Adedoyin, O. O., Odili, J. N., & Major, T. E. (2007). Primary Teachers' Perceptions of Classroom Assessment Practices as Means of Providing Quality Primary/Basic Education in Botswana and Nigeria. *Educational Research and Review*, 2.
- Nurrada, A. (2021). Kualitas Soal Penilaian Akhir Semester (PAS) Buatan Guru Mata Pelajaran Kimia Kelas X MIA SMA Di Kabupaten Kotawaringin Barat Pada Semester Ganjil Tahun Ajaran 2018/2019. *Jurnal Ilmiah Kanderang Tingang*, 12(1), 47–55. <https://doi.org/10.37304/jikt.v12i1.121>
- Oktafiana, L., Holsin, I., & Mursyidah, H. (2020). Analisis Soal Matematika Tipe Higher Order Thinking Skills (HOTS) Tingkat SMP. *Jurnal Pemikiran dan Penelitian Pendidikan Matematika (JP3M)*, 2(2), 112–129. <https://doi.org/10.36765/jp3m.v2i2.24>
- Oktavia, A. P. (2021). Analisis Kualitas Butir Soal Matematika pada Ujian Akhir Semester Gasal Berdasarkan Taksonomi Bloom Kelas VIII SMP Negeri 1 Patikraja Kabupaten Banyumas. *Skripsi*. Institut Agama Islam Negeri Purwokerto.
- Ropii, M., & Fahrurrozi, M. (2017). *Evaluasi Hasil Belajar*. Pancor: Universitas Hamzanwadi Press.
- Rurisfiani, Ramly, & Sultan. (2019). Level Berpikir Pertanyaan Guru pada Pembelajaran Bahasa Indonesia. *Jurnal Keilmuan Pendidikan Bahasa dan Sastra Indonesia*, 1(2), 111–119. <https://doi.org/10.26499/bahasa.v1i2.37>
- Siyoto, S., & Sodik, M. A. (2015). *Dasar Metodologi Penelitian*. Yogyakarta: Literasi Media Publishing.
- Sopiah, A., Sidauruk, S., & Asi, N. (2020). Kualitas Soal Penilaian Akhir Semester (PAS) Buatan Guru Mata Pelajaran Kimia Kelas X MIA SMA di Kabupaten Kota Waringin Barat Pada Semester Ganjil Tahun Ajaran 2018/2019. *Jurnal Ilmiah Kanderang Tingang*, 10(2), 110-126.
- Sriyanti, I. (2019). *Evaluasi Pembelajaran Matematika*. Subang: Uwais Inspirasi Indonesia.
- Sudijono, A. (2013). *Pengantar Evaluasi Pendidikan*. Yogyakarta: Rajagrafindo Persada.
- Sudjana, N. (2014). *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya.
- Sukiman. (2017). *Sistem Penilaian Pembelajaran*. Yogyakarta: Media Akademi.

- Sunaga, E. (2020). Meningkatkan Kompetensi Guru Dalam Menyusun Soal Penilaian Akhir Semester (PAS) Melalui Workshop Di Madrasah Binaan Min 2 Simalungun. *ANSIRU PAI: Pengembangan Profesi Guru*, 169–181.
- Suryaningtyas, S., & Setyaningrum, W. (2020). Analisis kemampuan metakognitif siswa SMA kelas XI program IPA dalam pemecahan masalah matematika. *Jurnal Riset Pendidikan Matematika*, 7(1), 74–87. <https://doi.org/10.21831/jrpm.v7i1.16049>
- Suryanto, A. (2018). *Evaluasi Pembelajaran*. Tangerang: CV. Gerina Prima.
- Susanto, H., Rinaldi, A., & Novalia. (2015). Analisis Validitas Reabilitas Tingkat Kesukaran dan Daya Beda pada Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika. *Jurnal Pendidikan Matematika*, 6, 203–217.
- Wahyuni, S., & Dinisa Putri, N. (2020). Karakteristik Instrumen Penilaian Akhir Semester Mata Pelajaran Matematika Di Smkn 1 Braja Sebah. *Journal of Mathematics Education*, 1(2), 126. <https://doi.org/10.32332/linear.v1i2.2958>
- Waty, E. R. K. (2017). Menelaah Kualitas Soal Ujian Sekolah Buatan Guru Pada Pencapaian Kelulusan Siswa. *Jurnal Pendidikan Dan Pemberdayaan Masyarakat*, 4(2), 11–17.
- Wulandari, A. R., & Pramusinto, H. (2020). Analisis Kualitas Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Otomatisasi Tata Kelola Sarana dan Prasarana Kelas XI Otomatisasi dan Tata Kelola Perkantoran. *Economic Education Analysis Journal*, 9(2), 366–378. <https://doi.org/10.15294/eeaj.v9i2.39000>
- Zainuddin. (2021). *Pengembangan dan Evaluasi Pendidikan*. Lombok: CV. Alliv Renteng Mandiri.
- Zein, A., Fadillah, M., & Novianti, R. (2013). Hubungan Antara Validitas Butir, Reliabilitas, Tingkat Kesukaran dan Daya Pembeda Soal Ujian Semester Genap Bidang Studi Biologi Kelas Xi SMA / MA Negeri Di Kota Padang Tahun Pelajaran 2010 / 2011. *Semirata 2013 FMIPA Unila*, 2009, 39–47.