

Pengembangan dan Validasi Instrumen Motivasi Belajar Matematika Siswa SD Berbasis Item Response Theory

Yuliana Yuliana^{1*}, Syukrul Hamdi², Risky Setiawan³, Siswantoyo Siswantoyo⁴,
Edy Istiyono⁵, Farida Agus Setiawati⁶

^{1,2,3,4,5,6}Universitas Negeri Yogyakarta

*Corresponding author, e-mail: yulianathen07@gmail.com

Abstrak

Motivasi belajar merupakan salah satu faktor penting yang memengaruhi keberhasilan siswa dalam memahami konsep dan memecahkan masalah. Namun, ketersediaan instrumen yang memiliki karakteristik psikometrik yang kuat untuk mengukur motivasi belajar matematika siswa sekolah dasar masih terbatas. Penelitian ini bertujuan untuk mengembangkan dan memvalidasi instrumen motivasi belajar Matematika siswa sekolah dasar menggunakan pendekatan *Item Response Theory* (IRT). Jenis penelitian yang digunakan adalah penelitian dan pengembangan (R & D) dengan model pengembangan 4-D yaitu pendefinisian (*define*), perancangan (*design*), pengembangan (*develop*), dan penyebaran (*disseminate*). Sampel penelitian sebanyak 734 siswa kelas IV, V, dan VI berasal dari dua puluh satu sekolah dasar yang tersebar di tujuh kecamatan yang ada di Kabupaten Belitung Timur. Sampel uji coba sebanyak 352 siswa dan sampel pengukuran sebanyak 382 siswa. Data dikumpulkan dengan menggunakan kuesioner. Analisis data meliputi uji validitas konten menggunakan indeks Aiken's V, validitas konstruk melalui *Confirmatory Factor Analysis* (CFA), dan analisis butir menggunakan *Graded Response Model* (GRM) dalam kerangka IRT. Hasil penelitian menunjukkan bahwa instrumen memiliki validitas konten yang sangat baik dengan indeks Aiken's V antara 0,92 hingga 1,00 dengan sembilan orang ahli validator berasal dari dua orang dosen dan tujuh orang pengawas SD Kabupaten Belitung Timur. Validitas konstruk pada CFA menunjukkan 20 dari 25 butir pernyataan yang diuji memenuhi kriteria kelayakan model dengan indeks kecocokan yang baik (*Chi-Square* 0,000, CFI 0,991, TLI 0,989, RMSEA 0,044, dan SRMR 0,057). Hasil *Graded Response Model* menunjukkan bahwa parameter ambang kategori respons berfungsi secara berurutan dan sebagian besar butir memiliki daya diskriminasi yang memadai. Hasil pengukuran menunjukkan bahwa 72,7% siswa berada pada tingkat motivasi sedang, 14,2% siswa memiliki motivasi belajar yang tinggi, dan hanya 13,1% siswa yang memiliki motivasi rendah. Temuan ini menunjukkan bahwa instrumen yang dikembangkan memiliki karakteristik psikometrik yang baik sehingga valid, reliabel, dan layak digunakan sebagai alat ukur motivasi belajar matematika siswa sekolah dasar. Instrumen ini juga dapat dimanfaatkan oleh pendidik sebagai dasar dalam merancang strategi pembelajaran yang adaptif sesuai dengan karakteristik motivasi belajar siswa.

Kata kunci: Instrumen, Matematika; Motivasi Belajar; Reliabilitas; Validitas.

Abstract

Learning motivation is an important factor influencing students' success in understanding concepts and solving problems in mathematics. However, the availability of instruments with strong psychometric properties to measure elementary school students' mathematics learning motivation is still limited. This study aimed to develop and validate an instrument for measuring elementary school students' mathematics learning motivation using the *Item Response Theory* (IRT) approach. This research employed a Research and Development (R&D) design using the 4-D development model, which consists of the stages of defining, designing, developing, and disseminating. The research sample consisted of 734 students in grades IV, V, and VI from twenty-one elementary schools across seven districts in East Belitung Regency. The pilot testing involved 352 students, while the measurement stage involved 382 students. Data were collected using a questionnaire. Data analysis included content validity testing using Aiken's V index, construct validity testing through *Confirmatory Factor Analysis* (CFA), and item analysis using the *Graded Response Model* (GRM) within the framework of *Item Response Theory*. The results showed that the instrument demonstrated excellent content validity, with Aiken's V values ranging from 0.92 to 1.00, evaluated by nine expert validators consisting of two university lecturers and seven elementary school

supervisors from East Belitung Regency. The CFA results indicated that 20 out of 25 items met the model fit criteria with satisfactory fit indices (Chi-Square = 0.000, CFI = 0.991, TLI = 0.989, RMSEA = 0.044, and SRMR = 0.057). The GRM analysis indicated that the response category threshold parameters functioned sequentially and that most items showed adequate discrimination power. The measurement results revealed that 72.7% of students had a moderate level of motivation, 14.2% had high motivation, and 13.1% had low motivation. These findings indicate that the developed instrumen has strong psychometric properties and is therefore valid, reliable, and appropriate for measuring elementary school students' mathematics learning motivation. The instrumen can also be utilized by educators as a basis for designing adaptive learning strategies that align with students' motivational characteristics.

Keywords: Instrumen; Learning Motivation; Mathematics; Reliability; Validity.

How to Cite: Yuliana, Y. et al. (2026). Pengembangan dan Validasi Instrumen Motivasi Belajar Matematika Siswa SD Berbasis Item Response Theory. *Naradidik: Journal of Education & Pedagogy*, 5(2), 278-293.



This is an open access article distributed under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2026 by author.

Pendahuluan

Era globalisasi dan perkembangan teknologi saat ini yang sudah memasuki revolusi industri 5.0, mengharuskan generasi muda memiliki keterampilan abad 21 yang disingkat dengan 6C, yaitu *character* (karakter), *citizenship* (kewarganegaraan), *critical thinking* (berpikir kritis), *creativity* (kreativitas), *collaboration* (kolaborasi), dan *communication* (komunikasi). Di samping itu, generasi muda juga harus memiliki keterampilan literasi abad 21, yaitu literasi informasi, literasi media, dan literasi teknologi. Pendidikan berperan penting dalam membentuk generasi muda untuk memiliki keterampilan 6C dan literasi abad 21 dan mempersiapkan peserta didik agar mampu menghadapi tantangan global melalui proses pembelajaran yang mendorong pengembangan kemampuan berpikir tingkat tinggi dan pemecahan masalah.

Matematika merupakan salah satu mata pelajaran fundamental yang berperan penting dalam mengembangkan kemampuan berpikir logis, sistematis, analitis, dan kreatif yang diperlukan dalam berbagai bidang kehidupan (Badjeber & Purwaningrum, 2018; Izzah & Azizah, 2019). Hal ini dikarenakan matematika menjadi salah satu mata pelajaran yang menuntut siswa untuk memahami konsep, menggunakan penalaran, memecahkan masalah, mengkomunikasikan dan saling menghargai (Dirgantoro, 2018). Tujuan dari adanya pembelajaran matematika di sekolah dasar siswa di sekolah dasar yakni agar siswa mampu berpikir secara berpikir kritis, sistematis, logis, analitis, dan kreatif serta memiliki kemauan kerja yang efektif (Wiryanto, 2020) dimana kemampuan ini merupakan suatu keterampilan berpikir tingkat tinggi atau *Higher Order Thinking Skills* (Dinni, 2021). Oleh karena itu, pembelajaran matematika di sekolah dasar memiliki peran strategis dalam membangun fondasi kemampuan berpikir siswa sejak dini.

Namun demikian, berbagai laporan pendidikan menunjukkan bahwa kemampuan matematika siswa Indonesia masih berada pada tingkat yang relatif rendah. Hasil Programme for International Student Assessment (PISA) tahun 2023 menunjukkan bahwa skor matematika siswa Indonesia sebesar 366, masih berada di bawah rata-rata internasional dengan selisih sekitar 106 poin (OECD, 2023). Selain itu, sekitar 82% siswa Indonesia masih berada pada tingkat kemampuan di bawah level dua, yang menunjukkan bahwa sebagian besar siswa mengalami kesulitan dalam memahami konsep matematika dasar serta menyelesaikan masalah yang memerlukan penalaran matematis. Kondisi tersebut juga tercermin dalam hasil Asesmen Nasional (AN) pada jenjang sekolah dasar. Laporan Rapor Pendidikan Kabupaten Belitung Timur Tahun 2025 menunjukkan skor rerata kemampuan numerasi siswa SD Se-Kabupaten Belitung Timur sebesar 52,32 dengan skor kemampuan pada domain bilangan, aljabar, geometri, data dan ketidakpastian berkisar 50,56-52,24 berada pada peringkat bawah di Provinsi Kepulauan Bangka Belitung (Dinas Pendidikan Kabupaten Belitung Timur, 2025). Temuan tersebut menunjukkan bahwa peningkatan kualitas pembelajaran matematika menjadi salah satu prioritas penting dalam upaya meningkatkan kemampuan numerasi siswa.

Rendahnya numerasi ini sering kali berakar pada stigma bahwa matematika adalah pelajaran yang sulit dan tidak diminati, yang sangat dipengaruhi oleh faktor motivasi belajar karena dasar dari kemampuan numerasi dipelajari dalam mata pelajaran matematika, yaitu bagaimana langkah operasi hitung dari domain bilangan, aljabar, geometri dan pengukuran, serta data dan ketidakpastian (Fachrudin, 2022). Sebagian besar siswa menganggap matematika menjadi salah satu mata pelajaran yang dianggap sulit dan kurang diminati. Salah satu penyebabnya adalah motivasi belajar (Novita et al., 2025). Siswa yang memiliki motivasi belajar

yang tinggi akan memiliki kompetensi yang baik saat ujian maupun tugas lainnya (Ryan & Deci, 2020). Motivasi belajar merupakan daya penggerak internal yang mendorong, mengarahkan, dan mempertahankan perilaku belajar siswa dalam mencapai tujuan akademik tertentu (Schunk et al., 2014). Dalam pembelajaran matematika, motivasi memiliki peran penting dalam menentukan tingkat keterlibatan siswa, ketekunan dalam menghadapi kesulitan, serta keberhasilan dalam mencapai hasil belajar yang optimal. Penelitian menunjukkan bahwa siswa yang memiliki motivasi belajar tinggi cenderung menunjukkan minat yang lebih besar terhadap pembelajaran matematika, lebih gigih dalam menyelesaikan tugas, serta memiliki kepercayaan diri yang lebih baik dalam menghadapi tantangan akademik (Ryan & Deci, 2020; Sivrikaya, 2019). Pada jenjang Sekolah Dasar (SD) kelas tinggi, siswa mulai dihadapkan pada konsep matematika yang lebih kompleks dan menuntut pemikiran logis serta pemecahan masalah yang lebih mendalam. Oleh karena itu, penting untuk secara akurat mengukur dan memahami dinamika motivasi belajar matematika pada kelompok usia ini agar intervensi pedagogis dapat dirancang secara efektif (Mukhoiyaroh & Aziz, 2024).

Secara teoritis, motivasi belajar dapat dijelaskan melalui berbagai pendekatan psikologis, salah satunya adalah *Self-Determination Theory* (SDT). Teori ini menjelaskan bahwa motivasi intrinsik siswa dipengaruhi oleh terpenuhinya tiga kebutuhan psikologis dasar, yaitu kompetensi, otonomi, dan keterhubungan akademik (Ryan & Deci, 2020). Dalam konteks pembelajaran matematika, motivasi belajar dapat tercermin melalui beberapa dimensi penting, seperti nilai instrinsik, pengaturan diri, kemampuan diri, nilai guna, dan kecemasan saat ujian (Fiorella et al., 2021). *Self-efficacy* menggambarkan keyakinan individu terhadap kemampuannya dalam menyelesaikan tugas belajar dan merupakan salah satu faktor psikologis penting yang memengaruhi keterlibatan serta prestasi siswa dalam pembelajaran matematika (Bjerke, 2026). Siswa yang memiliki *self-efficacy* tinggi cenderung lebih percaya diri dalam menghadapi tantangan akademik dan menunjukkan ketekunan yang lebih tinggi dalam proses belajar.

Sejumlah penelitian empiris menunjukkan bahwa motivasi belajar dan *self-efficacy* memiliki hubungan yang signifikan dengan pencapaian akademik siswa dalam matematika. Penelitian longitudinal menunjukkan bahwa *self-efficacy* matematika menjadi prediktor penting terhadap performa akademik dan kecemasan matematika siswa dalam proses pembelajaran (Lehikoinen et al., 2025). Studi lain juga menemukan bahwa motivasi belajar memiliki pengaruh yang signifikan terhadap kemampuan pemecahan masalah matematika dan hasil belajar siswa (Fitriyah, 2024). Selain itu, kajian sistematis mengenai motivasi belajar menunjukkan bahwa penguatan keyakinan diri siswa terhadap kemampuan matematika dapat meningkatkan keterlibatan belajar dan pencapaian akademik secara signifikan (Granello, 2025). Penelitian lain juga menegaskan bahwa motivasi belajar matematika merupakan faktor yang memediasi hubungan antara lingkungan belajar dan performa matematika siswa (Street et al., 2022). Temuan ini menunjukkan bahwa motivasi belajar merupakan faktor penting yang perlu diperhatikan dalam upaya meningkatkan kualitas pembelajaran matematika.

Meskipun penelitian mengenai motivasi belajar telah banyak dilakukan, sebagian besar instrumen yang digunakan dalam penelitian tersebut masih dikembangkan menggunakan pendekatan *Classical Test Theory* (CTT), seperti pengembangan instrumen motivasi belajar matematika (Andayana et al., 2021), dan pengembangan motivasi belajar IPA (Listiwani et al., 2021; Suhudi et al., 2024) dan hanya berfokus pada siswa kelas V. Pendekatan CTT memiliki keterbatasan dalam mengevaluasi karakteristik setiap butir instrumen secara lebih mendalam serta menghasilkan estimasi parameter yang bergantung pada karakteristik sampel penelitian (Thompson, 2023). Sebagai alternatif, pendekatan *Item Response Theory* (IRT) dalam psikometri modern menawarkan keunggulan dalam pengembangan instrumen pengukuran karena mampu memberikan estimasi parameter item yang lebih akurat dibandingkan pendekatan CTT (Ouyang et al., 2025) dan dikarenakan CTT tergantung pada sampel dan berpotensi menyebabkan estimasi yang bias. Keuntungan utama IRT adalah mampu menghasilkan fungsi informasi item dan tes, sehingga memudahkan dalam penelitian untuk menentukan item mana yang berkontribusi paling baik pada skala tersebut. Hal ini memungkinkan penghapusan item yang kurang efektif, sehingga menghasilkan instrumen pengukuran yang lebih pendek, efisien, dan tepat (Wang et al., 2024). Pendekatan ini juga memungkinkan analisis yang lebih komprehensif terhadap karakteristik butir instrumen, seperti tingkat kesulitan item, daya diskriminasi, serta fungsi kategori respons (Santoso et al., 2024). Melalui pendekatan ini, kualitas instrumen dapat dievaluasi secara lebih akurat sehingga menghasilkan alat ukur yang lebih valid dan reliabel.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengembangkan dan memvalidasi instrumen motivasi belajar matematika siswa sekolah dasar khususnya kelas tinggi menggunakan pendekatan *Item Response Theory* (IRT). Instrumen yang dikembangkan dalam penelitian ini mengintegrasikan dimensi nilai intrinsik, pengaturan diri, kemampuan diri, nilai guna, dan kecemasan. Sebanyak 25 indikator yang berasal dari pengembangan kuesioner motivasi matematika (Fiorella et al., 2021) dan indikator yang dikembangkan sendiri sesuai dengan langkah-langkah pengembangan instrumen afektif (Istiyono, 2020). Instrumen dianalisis menggunakan model *Graded Response Model* (GRM) dalam kerangka IRT. Hasil penelitian ini diharapkan dapat menghasilkan instrumen yang memiliki karakteristik psikometrik

yang kuat dan dapat digunakan sebagai alat ukur yang valid dan reliabel untuk mengidentifikasi tingkat motivasi belajar matematika siswa sekolah dasar serta menjadi dasar bagi guru dan pengambil kebijakan pendidikan dalam merancang strategi pembelajaran yang lebih efektif dan adaptif.

Metode Penelitian

Jenis penelitian yang digunakan adalah *Research And Development* (R & D) berdasarkan prosedur pengembangan instrumen non tes yang dikembangkan oleh Istiyono (2020). *Research and Development* adalah suatu proses atau langkah-langkah untuk mengembangkan suatu produk baru atau menyempurnakan produk yang telah ada, yang dapat dipertanggungjawabkan. Penelitian ini mengacu pada model penelitian dan pengembangan 4D (Four-D) yang dikembangkan Thiagarajan et al., (1974) yang terdiri atas 4 tahap utama yakni pendefinisian (*define*), perancangan (*design*), pengembangan (*develop*), dan penyebaran (*disseminate*).

Tahap *define* (pendefinisian) bertujuan untuk mengidentifikasi kebutuhan pengembangan instrumen dan menentukan konstruk yang akan diukur. Pada tahap ini dilakukan analisis literatur mengenai teori motivasi belajar serta analisis karakteristik siswa sekolah dasar. Sampel penelitian berasal dari siswa kelas IV, V, dan VI dari 3 sekolah dasar yang jaraknya jauh di satu kecamatan. Sebanyak 21 sekolah terpilih di Kabupaten Belitung Timur. Demografi sampel penelitian disajikan pada Tabel 1. Berdasarkan kajian teori motivasi belajar, instrumen dalam penelitian ini mengintegrasikan beberapa dimensi utama motivasi belajar, yaitu *self-efficacy*, atribusi terhadap keberhasilan dan kegagalan, serta tujuan belajar yang diyakini memiliki hubungan dengan keterlibatan belajar dan prestasi akademik siswa (Ryan & Deci, 2020; Street et al., 2024). Dari 3 dimensi utama dibagi menjadi 5 dimensi yang lebih spesifik, yaitu *self-efficacy* terdiri dari kemampuan diri dan kecemasan, atribusi terhadap keberhasilan dan kegagalan terdiri dari pengaturan diri, dan tujuan belajar terdiri dari nilai intrinsik.

Tabel 1. Demografi Siswa SD Kelas Tinggi Kabupaten Belitung Timur

| Kategori | Sub Kategori | Total Siswa | | Uji Coba | | Pengukuran | |
|---------------|-------------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | | n | % | n | % | n | % |
| Jenis Kelamin | Perempuan | 418 | 56,95 | 216 | 56,54 | 202 | 57,39 |
| | Laki-laki | 316 | 43,05 | 166 | 43,46 | 150 | 42,61 |
| | Jumlah | 734 | 100,00 | 382 | 100,00 | 352 | 100,00 |
| Kelas | IV | 260 | 35,42 | 134 | 35,08 | 126 | 35,80 |
| | V | 236 | 32,15 | 118 | 30,89 | 118 | 33,52 |
| | VI | 238 | 32,43 | 130 | 34,03 | 108 | 30,68 |
| | Jumlah | 734 | 100,00 | 382 | 100,00 | 352 | 100,00 |
| Usia | 9 tahun | 24 | 3,27 | 14 | 3,66 | 10 | 2,84 |
| | 10 tahun | 239 | 32,56 | 116 | 30,37 | 123 | 34,94 |
| | 11 tahun | 240 | 32,70 | 129 | 33,77 | 111 | 31,53 |
| | 12 tahun | 209 | 28,47 | 114 | 29,84 | 95 | 26,99 |
| | 13 tahun | 22 | 3,00 | 9 | 2,36 | 13 | 3,69 |
| | Jumlah | 734 | 100,00 | 382 | 100,00 | 352 | 100,00 |
| Kecamatan | Damar | 99 | 13,49 | 89 | 23,30 | 10 | 2,84 |
| | Dendang | 83 | 11,31 | 51 | 13,35 | 32 | 9,09 |
| | Gantung | 129 | 17,57 | 46 | 12,04 | 83 | 23,58 |
| | Kelapa Kampit | 108 | 14,71 | 61 | 15,97 | 47 | 13,35 |
| | Manggar | 171 | 23,30 | 71 | 18,59 | 100 | 28,41 |
| | Simpang Pesak | 54 | 7,36 | 20 | 5,24 | 34 | 9,66 |
| | Simpang Renggiang | 90 | 12,26 | 44 | 11,52 | 46 | 13,07 |
| Jumlah | 734 | 100,00 | 382 | 100,00 | 352 | 100,00 | |

Sumber: Data Penelitian (2025)

Partisipan didominasi oleh siswa berjenis kelamin perempuan (56,95%) dengan persebaran tingkat kelas yang relatif merata antara kelas IV (35,42%), kelas V (32,15%), dan kelas VI (32,43%). Dari sisi usia, mayoritas siswa berada pada rentang 10 hingga 11 tahun yang secara kumulatif mencakup lebih dari 65% total sampel. Sesuai dengan jumlah siswa di setiap kecamatan, Kecamatan Manggar menjadi kontributor

sampel terbesar dengan jumlah 171 siswa (23,30%), sedangkan Kecamatan Simpang Pesak memberikan kontribusi terkecil yakni sebanyak 54 siswa (7,36%).

Tahap *design* (perancangan) dilakukan dengan menyusun kisi-kisi instrumen berdasarkan indikator dari setiap dimensi motivasi belajar yang telah ditentukan. Selanjutnya dilakukan penulisan butir pernyataan yang merepresentasikan setiap indikator konstruk yang diukur. Setiap dimensi terdapat 5 indikator, sehingga terdapat 25 indikator yang terdiri dari 19 indikator berasal dari kuesioner motivasi matematika (Fiorella et al., 2021) dan 6 indikator yang dikembangkan sendiri sesuai dengan langkah-langkah pengembangan instrumen afektif (Istiyono, 2020) seperti yang disajikan pada Tabel 2. Instrumen yang dikembangkan menggunakan skala Likert untuk mengukur tingkat persetujuan responden terhadap setiap pernyataan, yang umum digunakan dalam pengukuran konstruk psikologis dan pendidikan (Joshi et al., 2015). Skala instrumen yang digunakan adalah skala Likert dengan skor 1 – 4 dengan ketentuan untuk pernyataan positif : [1] Tidak setuju, [2] Kurang Setuju, [3] Setuju, [4] Sangat Setuju, sedangkan pernyataan negatif: [1] Sangat setuju, [2] Setuju, [3] Kurang Setuju, [4] Tidak Setuju.

Tabel 2. Kisi-kisi dan Butir Instrumen

| Dimensi Motivasi | Nomor Butir/Indikator | Pernyataan Butir |
|------------------|--|------------------|
| Nilai intrinsik | 1. Saya merasa senang belajar Matematika. | Positif |
| | 7. Saya senang mencoba soal-soal pelajaran yang sedikit sulit. | Positif |
| | 18. Saya suka Matematika yang menantang saya. | Positif |
| | 22. Saya tidak tertarik dan mudah bosan saat pelajaran Matematika. | Negatif |
| | 23. Saya bangga jika saya bisa memahami pelajaran yang sulit. | Positif |
| Pengaturan Diri | 2. Saya cenderung mudah menyerah atau mencontek saat mendapatkan soal Matematika yang sulit | Negatif |
| | 5. Jika saya mengalami kesulitan dalam mempelajari Matematika, saya akan mencari contoh dan penyelesaian soal sejenis. | Positif |
| | 8. Saya berusaha cukup keras untuk mempelajari Matematika | Positif |
| | 20. Saya mempersiapkan diri dengan baik untuk ujian dan ulangan Matematika. | Positif |
| | 25. Saya menggunakan strategi/cara yang memastikan saya belajar Matematika dengan baik. | Positif |
| Kemampuan Diri | 3. Meskipun sudah selesai mengerjakan soal Matematika, saya sering tidak yakin jawaban saya benar. | Negatif |
| | 12. Saya yakin saya bisa mendapatkan nilai di atas kriteria ketuntasan pada mata pelajaran Matematika. | Positif |
| | 16. Saya yakin saya akan berhasil dalam ujian Matematika. | Positif |
| | 21. Saya yakin bisa mengerjakan soal-soal PR Matematika dengan baik. | Positif |
| | 24. Saya yakin saya dapat menguasai pengetahuan dan keterampilan dalam mata pelajaran Matematika. | Positif |
| Nilai Guna | 10. Saya belajar Matematika agar membantu saya mendapatkan pekerjaan yang baik. | Positif |
| | 11. Saya percaya Matematika yang saya pelajari akan bermanfaat bagi saya. | Positif |
| | 15. Saya merasa belajar Matematika hanya untuk lulus ujian. | Negatif |
| | 17. Belajar Matematika itu penting karena berguna untuk cita-citaku nanti. | Positif |
| | 19. Pengetahuan Matematika membantu saya saat melakukan kegiatan seperti berbelanja. | Positif |
| Kecemasan | 4. Saya gugup saat mengerjakan ujian Matematika. | Negatif |
| | 6. Saya menjadi cemas ketika tiba saatnya mengikuti ujian Matematika. | Negatif |
| | 9. Saya berani menjawab pertanyaan dari guru meskipun tidak yakin jawabannya benar. | Positif |
| | 13. Saya khawatir gagal dalam ujian Matematika. | Negatif |
| | 14. Saya khawatir siswa lain lebih baik dalam nilai Matematika. | Negatif |

Sumber: Data Penelitian (2025)

Tahap *develop* (pengembangan) meliputi proses validasi ahli, revisi instrumen, serta uji coba empiris. Validasi ahli dilakukan untuk menilai kesesuaian isi, kejelasan bahasa, serta relevansi butir instrumen terhadap konstruk yang diukur. Validator dalam penelitian ini terdiri dari sembilan ahli, yaitu dua orang ahli atau pakar akademisi yang terdiri dari 1 dosen doktor dari UPN “Veteran” Jakarta dan 1 dosen magister dari Universitas Bina Nusantara yang memiliki keahlian di bidang evaluasi pendidikan dan psikometri dan tujuh orang praktisi pendidikan, yaitu pengawas sekolah dasar Dinas Pendidikan Kabupaten Belitung Timur yang memiliki pengalaman dalam bidang supervisi akademik dan penilaian pendidikan. Penilaian validitas konten dilakukan menggunakan indeks Aiken’s V, yang digunakan untuk menilai tingkat kesepakatan para ahli terhadap relevansi setiap butir instrumen (Aiken, 1985).

Tahap *disseminate* (penyebaran) dilakukan melalui pengujian instrumen pada sampel penelitian untuk memperoleh bukti empiris mengenai validitas dan reliabilitas instrumen. Ukuran sampel yang direkomendasikan untuk menggunakan IRT minimal 250 orang (Istiyono, 2020). Sampel penelitian terdiri atas 734 siswa kelas IV, V, dan VI yang berasal dari 21 sekolah dasar di tujuh kecamatan di Kabupaten Belitung Timur. Sampel dibagi menjadi dua kelompok, yaitu 352 siswa sebagai sampel uji coba dan 382 siswa sebagai sampel pengukuran seperti yang disajikan pada Tabel 1. Data penelitian dikumpulkan menggunakan kuesioner motivasi belajar matematika yang telah dikembangkan. Analisis data dilakukan melalui beberapa tahapan. Pertama, validitas konten dianalisis menggunakan indeks Aiken’s V untuk menilai tingkat kesepakatan para ahli terhadap relevansi setiap butir instrumen (Aiken, 1985). Perhitungan indeks Aiken’s V menggunakan perangkat lunak *Microsoft Excel*. Kedua, validitas konstruk dianalisis menggunakan analisis faktor eksploratori (*Exploratory Factor Analysis/EFA*) dan *Confirmatory Factor Analysis* (CFA) untuk menguji kesesuaian model pengukuran dengan data empiris (Brown, 2015). Analisis EFA menggunakan perangkat lunak SPSS versi 31 dan analisis CFA dilakukan menggunakan perangkat lunak RStudio, dengan kriteria kelayakan model meliputi nilai Comparative Fit Index (CFI) $\geq 0,90$, Tucker-Lewis Index (TLI) $\geq 0,90$, Root Mean Square Error of Approximation (RMSEA) $\leq 0,08$, dan Standardized Root Mean Square Residual (SRMR) $\leq 0,08$ (Hamdi et al., 2026). Selanjutnya, analisis karakteristik butir dilakukan menggunakan pendekatan Item Response Theory (IRT) dengan model *Graded Response Model* (GRM) untuk mengevaluasi parameter diskriminasi dan ambang kategori respons setiap butir data politomus (Istiyono et al., 2024). Model GRM merupakan salah satu model IRT yang banyak digunakan untuk menganalisis data skala Likert dalam penelitian pendidikan (Issayeva, 2023). Analisis IRT dilakukan menggunakan perangkat lunak Rstudio, dan instrumen dinyatakan memiliki kualitas yang baik apabila parameter diskriminasi berada pada kategori sedang hingga tinggi serta parameter ambang kategori respons menunjukkan urutan yang logis.

Hasil dan Pembahasan

Validitas Konten

Tahap awal dalam pengembangan instrumen motivasi belajar matematika adalah melakukan validasi konten oleh para ahli. Validasi konten bertujuan untuk memastikan bahwa setiap butir instrumen telah merepresentasikan konstruk teoritis yang diukur secara tepat. Dalam penelitian ini, proses validasi dilakukan oleh sembilan validator yang terdiri dari dua dosen ahli/pakar praktisi dan tujuh pengawas sekolah dasar Dinas Pendidikan Kabupaten Belitung Timur. Penilaian dilakukan menggunakan indeks Aiken’s V untuk mengukur tingkat kesepakatan para ahli terhadap relevansi butir instrumen.

Hasil analisis menunjukkan bahwa nilai Aiken’s V berkisar antara 0,92 hingga 1,00, yang menunjukkan bahwa seluruh butir instrumen memiliki tingkat validitas konten yang sangat tinggi. Nilai tersebut berada di atas batas minimal yang direkomendasikan yaitu 0,80 pendidikan (Istiyono, 2020), sehingga dapat disimpulkan bahwa setiap butir instrumen telah memenuhi kriteria validitas isi. Validitas konten yang tinggi menunjukkan bahwa indikator yang digunakan telah sesuai dengan konstruk motivasi belajar yang ingin diukur.

Hasil ini sejalan dengan penelitian pengembangan instrumen motivasi belajar yang dilakukan oleh Mudanta et al. (2020) yang juga memperoleh nilai validitas konten yang sangat tinggi melalui proses validasi ahli. Validasi konten merupakan tahap penting dalam pengembangan instrumen karena memastikan bahwa butir yang disusun benar-benar mencerminkan konsep teoritis yang mendasari instrumen tersebut pendidikan (DeVellis & Thorpe, 2021). Dengan demikian, instrumen motivasi belajar matematika yang dikembangkan dalam penelitian ini telah memenuhi persyaratan awal sebagai alat ukur yang valid sebelum dilakukan pengujian empiris pada responden.

Validitas Konstruk

Validitas konstruk menggunakan analisis faktor eksploratori (*Exploratory Factor Analysis/EFA*) dengan *Statistical Package for the Social Sciences* (SPSS) versi 31 pada 382 data responden. Analisis ini bertujuan untuk

mengidentifikasi struktur faktor yang terbentuk secara empiris dari data responden (Hamdi et al., 2026). Hasil uji Kaiser Meyer Olkin (KMO) sebesar 0,888 ($> 0,50$) dan nilai Signifikansi Bartlett's Test sebesar 0,000 ($< 0,05$) menunjukkan bahwa korelasi antar variabel cukup kuat sehingga analisis faktor dapat dilakukan secara tepat dan data memiliki kecukupan sampel yang sangat baik. Menurut Hair et al. (2019), nilai KMO di atas 0,80 menunjukkan tingkat kecukupan sampel yang sangat baik untuk analisis faktor. Penggunaan analisis faktor ini membuktikan bahwa dimensi motivasi belajar yang diukur memang terbentuk secara empiris. Suhudi et al. (2024) dalam penelitiannya tentang kuesioner motivasi belajar IPA di SDIT BPMAA Pekanbaru juga menekankan pentingnya uji validitas dan reliabilitas ulang untuk memastikan akurasi hasil pengukuran pada objek penelitian yang spesifik. Data hasil uji coba terdapat pada Tabel 3.

Tabel 3. Hasil Uji KMO dan Bartlett's Test

| KMO and Bartlett's Test | | |
|--|--------------------|----------|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0,888 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 3310,027 |
| | df | 300 |
| | Sig. | 0,000 |

Sumber: Data Penelitian (2025)

Selain KMO, terdapat MSA yang menunjukkan seberapa jauh data yang terkumpul untuk setiap variabel mencukupi. MSA digunakan untuk mengetahui dan menentukan variabel mana saja yang layak dalam analisis faktor. Nilai MSA antar variabel harus $> 0,50$. Jika nilai MSA mendekati 1 menunjukkan hubungan yang sangat kuat antar variabel dan dapat diprediksi oleh variabel lain dengan sedikit kesalahan. Nilai MSA $< 0,50$ menunjukkan bahwa hubungan antar variabel lemah, maka variabel yang dimaksud tidak dapat dianalisis lebih mendalam dengan analisis faktor atau dihilangkan dalam analisis. Dalam praktik analisis data, nilai MSA terdapat pada *Anti-Image Matrices* nilai yang diberi tanda ^a (Asra, 2017) yang disajikan sebagai nilai pada diagonal utama matrik anti image (Cheung et al., 2024). Nilai MSA pada uji coba instrumen disajikan pada Tabel 4. Nilai MSA $> 0,5$ dan mendekati 1 artinya semua variabel lolos untuk dilakukan analisis faktor.

Tabel 4. Nilai MSA Uji Coba Instrumen

| Indikator | Nilai MSA | Indikator | Nilai MSA | Indikator | Nilai MSA | Indikator | Nilai MSA | Indikator | Nilai MSA |
|----------------|-----------|-----------------|-----------|-----------------|-----------|-----------------|-----------|-----------------|-----------|
| x ₁ | 0,925 | x ₆ | 0,851 | x ₁₁ | 0,857 | x ₁₆ | 0,908 | x ₂₁ | 0,917 |
| x ₂ | 0,788 | x ₇ | 0,924 | x ₁₂ | 0,889 | x ₁₇ | 0,856 | x ₂₂ | 0,880 |
| x ₃ | 0,797 | x ₈ | 0,916 | x ₁₃ | 0,857 | x ₁₈ | 0,900 | x ₂₃ | 0,891 |
| x ₄ | 0,796 | x ₉ | 0,913 | x ₁₄ | 0,768 | x ₁₉ | 0,912 | x ₂₄ | 0,921 |
| x ₅ | 0,928 | x ₁₀ | 0,871 | x ₁₅ | 0,785 | x ₂₀ | 0,919 | x ₂₅ | 0,928 |

Sumber: Data Penelitian (2025)

Untuk melihat berapa banyak komponen yang terbentuk dari variabel yang ada, maka dapat dilihat pada Total Varian yang Dijelaskan (*Total Variance Explained*) pada Tabel 5 dan pada Tabel 6 Matrik Komponen menunjukkan nilai korelasi antara masing-masing variabel dengan faktor yang terbentuk. yang terbentuk ada 5 komponen dilihat dari nilai Eigen (Total) > 1 dan Plot Layar pada Gambar 1.

Tabel 5. Total Varian yang Dijelaskan

| Compon ent | Total Variance Explained | | | | | | | | |
|---------------|--------------------------|---------------|--------------|-------------------------------------|---------------|--------------|-----------------------------------|---------------|--------------|
| | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 7,177 | 28,707 | 28,707 | 7,177 | 28,707 | 28,707 | 4,377 | 17,508 | 17,508 |
| 2 | 2,781 | 11,123 | 39,830 | 2,781 | 11,123 | 39,830 | 3,329 | 13,317 | 30,826 |
| 3 | 1,404 | 5,616 | 45,446 | 1,404 | 5,616 | 45,446 | 2,375 | 9,501 | 40,327 |

| | | | | | | | | | |
|----|-------|-------|---------|-------|-------|--------|-------|-------|--------|
| 4 | 1,049 | 4,195 | 49,641 | 1,049 | 4,195 | 49,641 | 2,166 | 8,665 | 48,992 |
| 5 | 1,038 | 4,153 | 53,795 | 1,038 | 4,153 | 53,795 | 1,201 | 4,803 | 53,795 |
| 6 | 0,970 | 3,882 | 57,676 | | | | | | |
| 7 | 0,909 | 3,635 | 61,312 | | | | | | |
| 8 | 0,828 | 3,312 | 64,623 | | | | | | |
| 9 | 0,804 | 3,216 | 67,840 | | | | | | |
| 10 | 0,768 | 3,070 | 70,910 | | | | | | |
| 11 | 0,727 | 2,908 | 73,818 | | | | | | |
| 12 | 0,692 | 2,767 | 76,585 | | | | | | |
| 13 | 0,671 | 2,684 | 79,269 | | | | | | |
| 14 | 0,613 | 2,451 | 81,720 | | | | | | |
| 15 | 0,558 | 2,233 | 83,954 | | | | | | |
| 16 | 0,522 | 2,086 | 86,040 | | | | | | |
| 17 | 0,504 | 2,017 | 88,057 | | | | | | |
| 18 | 0,477 | 1,908 | 89,965 | | | | | | |
| 19 | 0,466 | 1,862 | 91,827 | | | | | | |
| 20 | 0,404 | 1,616 | 93,443 | | | | | | |
| 21 | 0,380 | 1,521 | 94,965 | | | | | | |
| 22 | 0,354 | 1,417 | 96,382 | | | | | | |
| 23 | 0,346 | 1,386 | 97,767 | | | | | | |
| 24 | 0,286 | 1,144 | 98,912 | | | | | | |
| 25 | 0,272 | 1,088 | 100,000 | | | | | | |

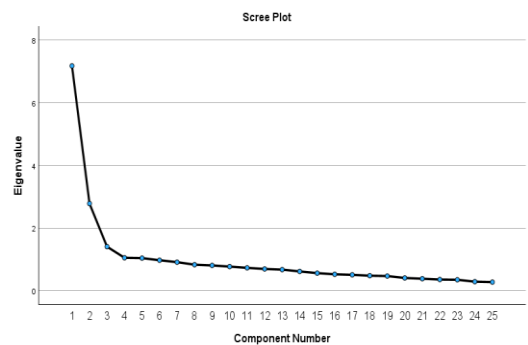
Sumber: Data Penelitian (2025)

Tabel 6. Matrik Komponen

| Indikator | Component Matrix ^a | | | | |
|-----------------|-------------------------------|--------|--------|--------|--------|
| | 1 | 2 | 3 | 4 | 5 |
| X ₁ | 0,665 | 0,017 | -0,182 | -0,286 | 0,030 |
| X ₂ | 0,313 | 0,388 | 0,396 | -0,234 | -0,179 |
| X ₃ | 0,329 | 0,586 | 0,018 | -0,280 | -0,084 |
| X ₄ | 0,334 | 0,697 | -0,065 | -0,103 | 0,089 |
| X ₅ | 0,540 | -0,103 | 0,040 | 0,018 | 0,439 |
| X ₆ | 0,415 | 0,649 | 0,087 | 0,057 | 0,013 |
| X ₇ | 0,567 | 0,042 | -0,265 | -0,130 | 0,184 |
| X ₈ | 0,579 | -0,296 | 0,058 | -0,233 | 0,230 |
| X ₉ | 0,456 | -0,064 | -0,339 | -0,056 | 0,335 |
| X ₁₀ | 0,533 | -0,249 | 0,145 | -0,104 | -0,239 |
| X ₁₁ | 0,564 | -0,307 | 0,388 | 0,019 | 0,092 |
| X ₁₂ | 0,583 | 0,036 | -0,273 | -0,083 | -0,408 |
| X ₁₃ | 0,411 | 0,570 | -0,109 | 0,297 | 0,057 |
| X ₁₄ | 0,207 | 0,558 | -0,083 | 0,473 | -0,103 |
| X ₁₅ | 0,221 | 0,310 | 0,497 | 0,157 | 0,393 |
| X ₁₆ | 0,700 | -0,076 | -0,114 | 0,126 | -0,233 |
| X ₁₇ | 0,575 | -0,307 | 0,370 | 0,018 | -0,151 |
| X ₁₈ | 0,630 | 0,044 | -0,314 | -0,188 | 0,110 |
| X ₁₉ | 0,608 | -0,199 | -0,141 | 0,088 | 0,155 |
| X ₂₀ | 0,669 | -0,203 | 0,218 | 0,115 | 0,080 |
| X ₂₁ | 0,703 | -0,132 | -0,175 | 0,327 | -0,030 |
| X ₂₂ | 0,511 | 0,265 | 0,288 | -0,328 | -0,129 |
| X ₂₃ | 0,488 | -0,180 | 0,235 | 0,292 | -0,080 |
| X ₂₄ | 0,705 | -0,082 | -0,025 | 0,126 | -0,185 |
| X ₂₅ | 0,601 | -0,278 | -0,123 | 0,045 | -0,140 |

Sumber: Data Penelitian (2025)

Gambar 1. Plot Layar



Factor Loading pada sampel uji coba ($n = 382$), maka harus $> 0,30$ (Hair et al., 2019). Kriteria *Factor Loading* dapat dilihat pada Tabel 7 dan *Factor Loading* pada Matrik Komponen yang Diputar $> 0,30$. Pada Tabel 8 nilai *Factor Loading* $> 0,30$ maka semuanya masuk ke dalam 5 faktor, namun tidak mengelompok ke dalam satu faktor yang sama sesuai dengan awal pengembangan instrumen artinya indikator yang ada belum konsisten, sehingga harus dikelompokkan ke dalam 5 faktor dengan indikator yang baru. Jika ada indikator termasuk dalam 2 faktor maka dimasukkan ke dalam faktor dengan nilai *Factor Loading* tertinggi.

Tabel 7. Nilai *Factor Loading* dan Jumlah Sampel

| Faktor loading | Jumlah Sampel |
|----------------|---------------|
| 0,30 | 350 |
| 0,35 | 250 |
| 0,40 | 200 |
| 0,45 | 150 |
| 0,50 | 120 |
| 0,55 | 100 |
| 0,60 | 85 |
| 0,65 | 70 |
| 0,70 | 60 |
| 0,75 | 50 |

(Hair et al., 2019)

Tabel 8. Matrik Komponen yang Diputar

| Indikator | Rotated Component Matrix ^a | | | | |
|-----------|---------------------------------------|-------|-------|-------|--------|
| | 1 | 2 | 3 | 4 | 5 |
| X1 | | 0,582 | 0,331 | | |
| X2 | | | 0,655 | | |
| X3 | | | 0,651 | | |
| X4 | | | 0,545 | 0,481 | |
| X5 | 0,309 | 0,537 | | | 0,334 |
| X6 | | | 0,517 | 0,547 | |
| X7 | | 0,608 | | | |
| X8 | 0,417 | 0,519 | | | |
| X9 | | 0,652 | | | |
| X10 | 0,576 | | | | |
| X11 | 0,656 | | | | |
| X12 | 0,379 | | | | -0,511 |
| X13 | | | | 0,696 | |
| X14 | | | | 0,762 | |
| X15 | | | | | 0,655 |
| X16 | 0,579 | 0,328 | | | |
| X17 | 0,729 | | | | |
| X18 | | 0,647 | | | |
| X19 | 0,431 | 0,514 | | | |
| X20 | 0,648 | 0,309 | | | |
| X21 | 0,579 | 0,423 | | 0,337 | |
| X22 | | | 0,654 | | |
| X23 | 0,618 | | | | |
| X24 | 0,606 | 0,309 | | | |
| X25 | 0,543 | 0,360 | | | |

Sumber: Data Penelitian (2025)

Pada analisis faktor konfirmatori (*Confirmatory Factor Analysis/CFA*) di RStudio kelima komponen di atas tidak bisa dilanjutkan dikarenakan ada data yang bernilai negatif, jika item dibuang maka data tidak bisa dianalisis lebih lanjut karena dalam satu faktor hanya ada 2 butir (minimal 3 butir). Data dikembalikan seperti semula dan dilakukan analisis reliabilitas di R Studio dengan melihat nilai r drop pada item statistics

pada Tabel 9. Jika nilai r drop negatif dan $< 0,25$, maka butir akan dihilangkan bertujuan untuk meningkatkan nilai Cronbach Alpha. Hasil analisis reliabilitas menunjukkan bahwa butir $x_2 = 0,21$, $x_3 = 0,18$, $x_9 = 0,22$, $x_{15} = 0,0091$, dan $x_{23} = 0,24$, sehingga kelima butir ini harus dihilangkan. Estimasi reliabilitas mengacu pada kemampuan alat ukur untuk tetap stabil dan konsisten dalam memberikan hasil yang sama setiap kali pengukuran dilakukan (Sürücü & Maslakçi, 2020). Hasil analisis reliabilitas semua butir menunjukkan nilai Cronbach Alpha 0,84 untuk $x_1, x_{12}, x_{16}, x_{18}, x_{20}, x_{21}, x_{24}$, selain dari itu bernilai 0,85. Hal ini menunjukkan konsistensi internal yang sangat baik. Hasil ini setara dengan instrumen motivasi yang dikembangkan oleh Mudanta et al. (2020) yang memiliki koefisien reliabilitas 0,86. Hal ini membuktikan bahwa instrumen ini memiliki stabilitas yang baik untuk mengukur motivasi belajar siswa sekolah dasar secara konsisten.

Tabel 9. Kesesuaian Model pada Uji Coba

| Indeks Kesesuaian | Hasil | Kriteria | Keterangan |
|---|-------|-------------|--------------|
| Chi-Square / P-Values | 0,00 | $\geq 0,05$ | Tidak Sesuai |
| CFI (Comparative Fit Index) | 0,91 | $\geq 0,90$ | Sesuai |
| TLI (Tucker Lewis Index) | 0,90 | $\geq 0,90$ | Sesuai |
| RMSEA (Root Mean Square Error of Approximation) | 0,06 | $< 0,80$ | Sesuai |
| SRMR (Standardized Root Means Square) | 0,05 | $< 0,80$ | Sesuai |

Sumber: Data Penelitian (2025)

Hasil analisis faktor konfirmatori (*Confirmatory Factor Analysis/CFA*) disajikan pada Tabel 10. Nilai p -values diperoleh sebesar 0,00 ($\geq 0,05$) menunjukkan bahwa ditemukan perbedaan signifikan antara data empiris dan model estimasi, sehingga model dianggap tidak sesuai. Hal ini bisa disebabkan karena sampel data yang besar (Rachmawati & Siswanto, 2025) sehingga harus memperhatikan kesesuaian dengan indeks lainnya. Hasil CFI, TLI, RMSEA, dan SRMR menunjukkan bahwa data sudah sesuai dan instrumen dapat digunakan.

Berdasarkan hasil analisis faktor, terdapat lima faktor utama dengan nilai eigenvalue lebih besar dari satu, yang secara kumulatif menjelaskan sekitar 53,795% varians total. Hal ini menunjukkan bahwa sebagian besar variasi data dapat dijelaskan oleh lima dimensi motivasi belajar yang terbentuk dari hasil analisis faktor. Struktur faktor ini memperkuat pandangan bahwa motivasi belajar merupakan konstruk multidimensional yang terdiri dari berbagai aspek psikologis yang saling berkaitan. Temuan ini konsisten dengan penelitian Fiorella et al. (2021) yang menunjukkan bahwa motivasi belajar siswa dapat dijelaskan melalui beberapa dimensi utama seperti nilai intrinsik, pengaturan diri, kemampuan diri, nilai guna, dan kecemasan. Dengan demikian, struktur faktor yang terbentuk dalam penelitian ini mendukung konsep bahwa motivasi belajar tidak hanya dipengaruhi oleh satu faktor tunggal, tetapi oleh berbagai komponen psikologis yang bekerja secara simultan.

Setelah struktur faktor terbentuk, langkah selanjutnya adalah melakukan analisis reliabilitas untuk menilai konsistensi internal instrumen. Reliabilitas menunjukkan sejauh mana instrumen mampu menghasilkan hasil pengukuran yang stabil dan konsisten. Hasil analisis reliabilitas menunjukkan bahwa beberapa butir memiliki nilai *corrected item-total correlation (r-drop)* kurang dari 0,25, sehingga lima butir (x_2, x_3, x_9, x_{15} , dan x_{23}) dieliminasi dari instrumen. Eliminasi butir dengan korelasi rendah bertujuan untuk meningkatkan konsistensi internal instrumen serta memastikan bahwa setiap butir memiliki kontribusi yang signifikan terhadap konstruk yang diukur. Setelah proses seleksi butir dilakukan, nilai Cronbach's Alpha sebesar 0,84–0,85, yang menunjukkan bahwa instrumen memiliki tingkat reliabilitas yang sangat baik. Menurut Sürücü & Maslakçi (2020), nilai Cronbach's Alpha di atas 0,80 menunjukkan bahwa instrumen memiliki konsistensi internal yang tinggi. Hasil ini juga sejalan dengan penelitian pengembangan instrumen motivasi belajar yang dilakukan oleh Mudanta et al. (2020) yang memperoleh koefisien reliabilitas sebesar 0,86. Dengan demikian, instrumen motivasi belajar matematika yang dikembangkan dalam penelitian ini memiliki tingkat reliabilitas yang baik dan dapat digunakan untuk mengukur motivasi belajar siswa secara konsisten.

Interpretasi Parameter Butir GRM

Setelah analisis validitas dan reliabilitas dilakukan, karakteristik butir dianalisis menggunakan pendekatan Item Response Theory (IRT) dengan model Graded Response Model (GRM). Model GRM digunakan karena instrumen menggunakan skala Likert yang bersifat politomus (Istiyono et al., 2024). Penggunaan model *Graded Response Model (GRM)* dalam penelitian ini memberikan informasi yang lebih detail dibandingkan teori klasik. Analisis menunjukkan bahwa butir-butir tes fit dengan model, selaras dengan temuan Istiyono et al. (2020) yang menyatakan bahwa analisis IRT memberikan parameter yang

tidak tergantung pada sampel (*item invariant*). Hasil GRM pada uji coba tersaji pada Tabel 11. Parameter ambang batas (Extrmt1, Extrmt2, Extrmt3) atau sering disebut "extremity parameters" di RStudio. Dalam model GRM yang ideal, nilai ambang batas harus meningkat secara monoton, yaitu $\text{Extrmt1} < \text{Extrmt2} < \text{Extrmt3}$. Jika tidak, ini menunjukkan masalah pada butir tersebut. Pada $x_1, x_4, x_5, x_6, x_7, x_8, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}, x_{21}, x_{24}, x_{25}$ semua ambang batasnya terurut dengan baik ($\text{Extrmt1} < \text{Extrmt2} < \text{Extrmt3}$). Kondisi ini menunjukkan bahwa kategori respons pada skala Likert berfungsi secara logis dalam membedakan tingkat motivasi siswa. Menurut De Ayala (2022), urutan ambang kategori yang monoton merupakan indikator bahwa responden dapat membedakan tingkat respons secara konsisten. Namun pada x_{22} , Extrmt3 masih bernilai negatif artinya butir ini masih "relatif mudah" dibandingkan dengan butir lain. Selain parameter ambang, analisis juga menghasilkan parameter diskriminasi (*discrimination parameter*) yang menunjukkan kemampuan suatu butir dalam membedakan responden dengan tingkat motivasi yang berbeda. $\text{Dscrmn} < 1$ menunjukkan diskriminasi rendah, butir ini kurang efektif dalam membedakan individu. Ini terlihat pada $x_4, x_6, x_7, x_{13}, x_{14}, x_{22}$. Nilai Dscrmn x_{14} paling rendah, artinya butir ini paling tidak efektif dalam membedakan responden. $\text{Dscrmn} \approx 1-2$ artinya diskriminasi sedang hingga baik, yaitu $x_1, x_5, x_8, x_{10}, x_{11}, x_{12}, x_{17}, x_{18}, x_{19}, x_{20}$, dan x_{25} . $\text{Dscrmn} > 2$ artinya diskriminasi sangat tinggi, butir ini sangat baik dalam membedakan kemampuan responden terdapat pada x_{16}, x_{21} , dan x_{24} .

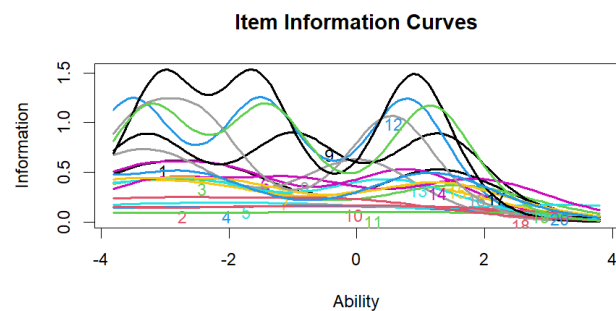
Diskriminasi rendah pada beberapa butir dapat disebabkan oleh beberapa faktor, seperti redaksi pernyataan yang kurang jelas, interpretasi responden yang berbeda, atau indikator konstruk yang kurang spesifik. Oleh karena itu, butir dengan diskriminasi rendah perlu dipertimbangkan untuk direvisi pada pengembangan instrumen selanjutnya. Sebaliknya, beberapa butir menunjukkan parameter diskriminasi yang sangat tinggi sangat efektif dalam membedakan responden dengan tingkat motivasi yang berbeda. Menurut Baker & Kim (2019), butir dengan diskriminasi tinggi memiliki kontribusi yang besar dalam meningkatkan presisi pengukuran instrumen. Secara keseluruhan, hasil analisis IRT menunjukkan bahwa sebagian besar butir dalam instrumen memiliki kualitas psikometrik yang baik sehingga mampu mengukur motivasi belajar siswa secara akurat.

Tabel 10. Hasil GRM Uji Coba

| Indikator | Extrmt 1 | Extrmt 2 | Extrmt 3 | Dscrmn |
|-----------|-------------|-------------|-------------|--------|
| x_1 | -3.309 | -2.162 | 1.291 | 1.452 |
| x_4 | -4.356 | -1.000 | 1.034 | 0.746 |
| x_5 | -3.575 | -2.182 | 1.577 | 1.211 |
| x_6 | -4.311 | -1.224 | 0.599 | 0.726 |
| x_7 | -3.451 | -0.986 | 3.323 | 0.821 |
| x_8 | -3.424 | -2.145 | 0.797 | 1.449 |
| x_{10} | -3.813 | -2.459 | 0.915 | 1.192 |
| x_{11} | -3.658 | -3.021 | 0.021 | 1.584 |
| x_{12} | -3.310 | -1.000 | 1.280 | 1.875 |
| x_{13} | -4.612 | -1.044 | 1.318 | 0.749 |
| x_{14} | -5.051 | -0.874 | 2.029 | 0.582 |
| x_{16} | -3.517 | -1.498 | 0.795 | 2.223 |
| x_{17} | -5.075 | -2.660 | 0.508 | 1.302 |
| x_{18} | -3.021 | -1.001 | 1.910 | 1.304 |
| x_{19} | -3.797 | -2.072 | 1.259 | 1.258 |
| x_{20} | -3.359 | -2.490 | 0.547 | 2.064 |
| x_{21} | -3.036 | -1.600 | 0.905 | 2.438 |
| x_{22} | -5.528 | -2.782 | -0.211 | 0.945 |
| x_{24} | -3.229 | -1.398 | 1.158 | 2.162 |
| x_{25} | -4.810 | -2.655 | 1.163 | 1.405 |

Sumber: Data Penelitian (2025)

Gambar 2. Kurva Informasi Butir



Pengukuran

Hasil pengukuran dianalisis menggunakan CFA untuk mengetahui kesesuaian butir dan GRM untuk mengetahui seberapa baik suatu item membedakan antara individu pada berbagai tingkatan kemampuan (parameter diskriminasi) dan memperkirakan ambang batas (parameter lokasi). Hasil analisis CFA terlihat pada Tabel 12 menunjukkan kesesuaian model dan Tabel 13 parameter diskriminasi $\text{Dscrmn} < 1$ menunjukkan diskriminasi rendah pada butir $x_2, x_4, x_5, x_{10}, x_{11}$, dan x_{18} , diskriminasi sedang antara 1 – 2

terdapat pada butir $X_1, X_3, X_6, X_7, X_8, X_9, X_{13}, X_{14}, X_{15}$, dan X_{20} , serta diskriminan tinggi >2 terdapat pada butir X_{12}, X_{16}, X_{17} , dan X_{19} .

Tabel 11. Kesesuaian Model pada Pengukuran

| Indeks Kesesuaian | Hasil | Kriteria | Keterangan |
|---|-------|-------------|--------------|
| Chi-Square / P-Values | 0,000 | $\geq 0,05$ | Tidak Sesuai |
| CFI (Comparative Fit Index) | 0,991 | $\geq 0,90$ | Sesuai |
| TLI (Tucker Lewis Index) | 0,989 | $\geq 0,90$ | Sesuai |
| RMSEA (Root Mean Square Error of Approximation) | 0,044 | $< 0,80$ | Sesuai |
| SRMR (Standardized Root Means Square) | 0,057 | $< 0,80$ | Sesuai |

Sumber: Data Penelitian (2025)

Tabel 12. Hasil GRM Pengukuran

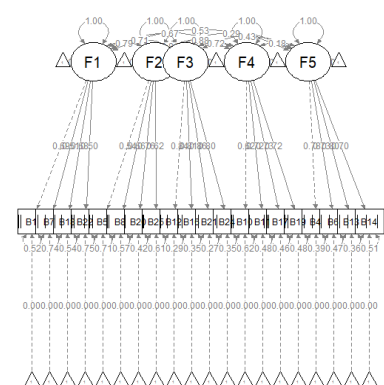
Item

| Indikator | Extrmt 1 | Extrmt 2 | Extrmt 3 | Dscrmn |
|-----------------|-------------|-------------|-------------|--------|
| X ₁ | -3.309 | -2.162 | 1.291 | 1.452 |
| X ₂ | -4.356 | -1.000 | 1.034 | 0.746 |
| X ₃ | -3.575 | -2.182 | 1.577 | 1.211 |
| X ₄ | -4.311 | -1.224 | 0.599 | 0.726 |
| X ₅ | -3.451 | -0.986 | 3.323 | 0.821 |
| X ₆ | -3.424 | -2.145 | 0.797 | 1.449 |
| X ₇ | -3.813 | -2.459 | 0.915 | 1.192 |
| X ₈ | -3.658 | -3.021 | 0.021 | 1.584 |
| X ₉ | -3.310 | -1.000 | 1.280 | 1.875 |
| X ₁₀ | -4.612 | -1.044 | 1.318 | 0.749 |
| X ₁₁ | -5.051 | -0.874 | 2.029 | 0.582 |
| X ₁₂ | -3.517 | -1.498 | 0.795 | 2.223 |
| X ₁₃ | -5.075 | -2.660 | 0.508 | 1.302 |
| X ₁₄ | -3.021 | -1.001 | 1.910 | 1.304 |
| X ₁₅ | -3.797 | -2.072 | 1.259 | 1.258 |
| X ₁₆ | -3.359 | -2.490 | 0.547 | 2.064 |
| X ₁₇ | -3.036 | -1.600 | 0.905 | 2.438 |
| X ₁₈ | -5.528 | -2.782 | -0.211 | 0.945 |
| X ₁₉ | -3.229 | -1.398 | 1.158 | 2.162 |
| X ₂₀ | -4.810 | -2.655 | 1.163 | 1.405 |

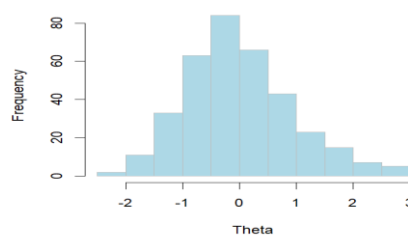
Sumber: Data Penelitian (2025)

Berdasarkan hasil dari CFA pada Gambar 3 mayoritas indikator memiliki *factor loading* yang tinggi (di atas 0.70) menunjukkan bahwa indikator-indikator tersebut memiliki validitas konvergen yang kuat. Tingkat korelasi yang sangat tinggi antara beberapa faktor (misalnya 0.88 antara F3 dan F4) menimbulkan kekhawatiran tentang validitas diskriminan (apakah faktor-faktor tersebut benar-benar berbeda). Indikator seperti B10 (residual variance 0.290, $R^2 = 0.71$) adalah pengukur yang sangat baik, sementara indikator seperti B1 (residual variance 0.520, $R^2 = 0.48$) adalah pengukur yang lebih lemah.

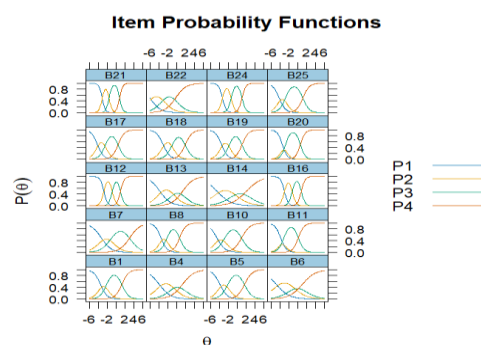
Gambar 3. Hasil CFA 20 Butir



Distribusi Kemampuan (Theta EAP)



Gambar 4. Distribusi Kemampuan Theta (EAP)



Gambar 5. Item Probability Functions (IPF)

Berdasarkan Gambar 4, hasil estimasi kemampuan partisipan menggunakan metode *Expected A Posteriori* (EAP) menunjukkan distribusi skor theta (θ) yang mendekati distribusi normal. Mayoritas responden memiliki tingkat kemampuan yang terpusat pada rentang -1 hingga +1 menunjukkan bahwa sebagian besar siswa memiliki tingkat motivasi yang berada di sekitar rata-rata. Rata-rata kemampuan sampel berada di sekitar nilai 0, yang mengindikasikan bahwa tingkat kesulitan instrumen secara umum setara dengan rata-rata kemampuan kelompok responden yang diuji. Menurut [Embretson & Reise \(2019\)](#), instrumen yang baik seharusnya memiliki tingkat kesulitan yang sesuai dengan karakteristik populasi yang diukur sehingga dapat memberikan informasi yang optimal. Meskipun demikian, terdapat sebaran frekuensi yang cukup kecil pada skor ekstrem kanan (kemampuan tinggi di atas +2), menunjukkan bahwa instrumen ini mampu membedakan individu dengan tingkat kemampuan tinggi, namun jumlah responden di kategori tersebut relatif terbatas dalam sampel penelitian ini.

Analisis karakteristik butir dilakukan dengan meninjau *Item Probability Functions* (IPF) pada Gambar 5. Kurva ini merepresentasikan probabilitas responden memilih kategori jawaban tertentu (P1 hingga P4) berdasarkan tingkat kemampuan (θ) siswa. Hasil observasi IPF menunjukkan (1) sebagian besar item (seperti B12, B14, dan B21) menunjukkan kurva kategori yang tumpang tindih secara proporsional. Hal ini menandakan bahwa pilihan jawaban (skala Likert/peringkat) berfungsi dengan baik dan setiap kategori memiliki peluang untuk dipilih pada level kemampuan tertentu. (2) Item seperti B1, B4, B7, dan B12 menunjukkan kurva yang lebih curam dan berdekatan di area theta negatif hingga nol, menandakan item ini cenderung lebih mudah disetujui atau dijawab oleh responden dengan kemampuan rendah hingga sedang. Sebaliknya, item seperti B22 dan B6 memiliki kurva yang lebih landai dan bergeser ke arah kanan (theta positif), yang mengindikasikan bahwa item tersebut memiliki tingkat kesulitan yang lebih tinggi atau membutuhkan tingkat kemampuan yang lebih besar untuk mencapai probabilitas jawaban pada kategori tertinggi (P4). (3) Bentuk kurva yang tertata rapi pada sebagian besar item mengonfirmasi bahwa instrumen memenuhi asumsi monotonisitas, di mana semakin tinggi kemampuan seseorang (θ), semakin besar probabilitas mereka untuk memilih kategori jawaban yang lebih tinggi.

Berdasarkan hasil analisis IRT dan asumsi distribusi kemampuan siswa mendekati normal, pembagian persentase kemampuan siswa adalah 72,7% siswa berada pada tingkat motivasi belajar matematika sedang, 14,2% siswa memiliki motivasi belajar matematika yang tinggi, dan hanya 13,1% siswa yang memiliki motivasi belajar matematika rendah. Temuan ini menunjukkan bahwa sebagian besar siswa memiliki tingkat motivasi yang cukup dalam pembelajaran matematika, namun masih diperlukan upaya untuk meningkatkan motivasi belajar siswa agar dapat mencapai tingkat yang lebih tinggi. Penelitian sebelumnya menunjukkan bahwa motivasi belajar memiliki hubungan yang signifikan dengan prestasi akademik siswa ([Ryan & Deci, 2020](#)). Butir-butir instrumen memiliki sebaran yang baik karena mampu mengukur siswa dari tingkat motivasi sangat rendah hingga sangat tinggi. Hal ini diperkuat dengan merujuk pada penelitian [Mediartika & Aznam \(2018\)](#) juga menekankan bahwa instrumen yang berkualitas harus memiliki rentang parameter kesukaran yang luas. Secara praktis, temuan bahwa mayoritas siswa memiliki motivasi tingkat sedang (72,7%) menjadi catatan penting bagi pendidik. [Yuniarto \(2017\)](#) dalam studinya menunjukkan bahwa tingkat motivasi memiliki korelasi langsung dengan pencapaian nilai akademik. Oleh karena itu, instrumen yang dikembangkan dalam penelitian ini dapat digunakan sebagai alat diagnostik awal bagi guru untuk meningkatkan motivasi siswa, sebagaimana disarankan oleh [Listiawani et al. \(2024\)](#) dalam pengembangan instrumen motivasi belajar IPA.

Penggunaan skala politomus dalam penelitian ini didukung oleh [Ahmad et al. \(2025\)](#) yang menggunakan Model Rasch untuk validasi psikometrik dengan menekankan pentingnya analisis *outfit* untuk mendeteksi butir yang tidak berfungsi dengan baik dalam mengukur kemampuan siswa. IRT mampu mengatasi kelemahan CTT dalam hal ketergantungan parameter pada sampel ([Selçuk & Demir, 2024](#)).

Selain itu, IRT sangat efektif dalam pengembangan tes berbasis *High Order Thinking Skills* (HOTS) maupun variabel afektif karena mampu memetakan posisi siswa (theta) dan posisi butir pada skala yang sama (Istiyono et al., 2018). Selain itu, penggunaan pendekatan *Item Response Theory* (IRT) dalam penelitian ini memberikan keuntungan dalam hal estimasi parameter yang tidak bergantung pada karakteristik sampel. Hal ini membuat instrumen yang dikembangkan memiliki tingkat generalisasi yang lebih baik dibandingkan dengan instrumen yang hanya menggunakan pendekatan teori tes klasik (Selçuk & Demir, 2024). Dengan demikian, instrumen motivasi belajar matematika yang dikembangkan dalam penelitian ini dapat digunakan sebagai alat diagnostik untuk mengidentifikasi tingkat motivasi belajar siswa. Informasi yang diperoleh dari instrumen ini dapat membantu guru dalam merancang strategi pembelajaran yang lebih efektif dan adaptif sesuai dengan karakteristik motivasi siswa.

Kesimpulan

Penelitian ini menghasilkan instrumen penilaian motivasi belajar matematika untuk siswa SD kelas tinggi di Kabupaten Belitung Timur terbukti valid dan reliabel. Dari 25 butir pernyataan awal sebanyak 20 butir dinyatakan valid dan reliabel dan tersebar dalam lima dimensi motivasi belajar, yaitu nilai intrinsik, pengaturan diri, kemampuan diri, nilai guna, dan kecemasan. Validitas konten yang diperoleh melalui indeks Aiken's V berada pada rentang 0,92 hingga 1,00, menunjukkan kesesuaian isi instrumen dengan konstruk yang diukur. Analisis faktor konfirmatori (CFA) pada 382 sampel juga mengonfirmasi kelayakan konstruk dengan nilai CFI, TLI, RMSEA, dan SRMR dalam kategori baik. Sementara itu, analisis IRT menggunakan model GRM menunjukkan bahwa seluruh butir berfungsi sesuai harapan, dengan tingkat diskriminasi bervariasi dari rendah hingga tinggi. Temuan penelitian juga menunjukkan bahwa mayoritas siswa berada pada tingkat motivasi sedang (72,7%), diikuti motivasi tinggi (14,2%) dan motivasi rendah (13,1%). Kondisi ini mengindikasikan bahwa sebagian besar siswa kelas IV, V, dan VI telah memiliki dorongan internal yang cukup untuk belajar matematika, namun tetap diperlukan strategi pembelajaran yang mampu mendorong peningkatan motivasi ke tingkat yang lebih tinggi.

Pengembangan instrumen ini memberikan kontribusi penting bagi guru, sekolah, dan peneliti dalam mendiagnosis tingkat motivasi belajar matematika siswa secara lebih akurat dan berbasis bukti, membantu merancang intervensi pedagogis yang lebih tepat sasaran sesuai kebutuhan siswa pada tiap dimensi motivasi, mendukung persiapan siswa menghadapi asesmen seperti Tes Kemampuan Akademik (TKA) dan menyediakan instrumen baku yang dapat digunakan sebagai dasar pemetaan motivasi belajar di tingkat sekolah maupun daerah. Penelitian ini memiliki beberapa keterbatasan, yaitu sampel penelitian hanya mencakup siswa SD di Kabupaten Belitung Timur, sehingga generalisasi hasil ke daerah lain perlu dilakukan dengan hati-hati, instrumen hanya dikembangkan untuk kelas tinggi (IV–VI), sehingga belum mencerminkan kebutuhan motivasi belajar siswa pada kelas rendah, dan pengukuran dilakukan dalam konteks mata pelajaran matematika saja. Penelitian selanjutnya dapat melakukan uji coba lanjutan pada wilayah atau populasi berbeda untuk mendapatkan generalisasi instrumen yang lebih kuat, mengembangkan instrumen motivasi belajar matematika untuk siswa kelas rendah, dan menerapkan instrumen ini pada mata pelajaran lain untuk menguji konsistensi struktur motivasi pada domain pembelajaran berbeda.

Daftar Pustaka

- Ahmad, N. Q., Arthur, R., Rahayu, W., & Rahmadhani, E. (2025). Pengembangan instrumen kemampuan pemecahan masalah matematika siswa menggunakan model rasch. *Juring (Journal for Research in Mathematics Learning)*, 8(1), 61–74. <http://dx.doi.org/10.24014/juring.v8i1.34022>
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Andayana, I. K. A., Margunayasa, I. G., & Yudianta, K. (2021). Pengembangan Instrumen Pengukuran Motivasi Belajar Matematika Siswa Kelas V Sekolah Dasar. *Mimbar Pendidikan Indonesia (MPI)*, 2(1), 173–179. <https://doi.org/10.23887/mpi.v2i1.30058>
- Asra, A., Utomo, A. P., Asikin, M., dan Puspongoro, N. H. (2017). Analisis Mutivariabel: Suatu Pengantar. Edisi 1. Bogor: IN MEDIA.
- Baker, F. B., & Kim, S. H. (2017). *The basics of item response theory using R*. Springer.
- Badjeber, R., & Purwaningrum, J. P. (2018). Pengembangan Higher Order Thinking Skills dalam Pembelajaran Matematika di SMP. *Guru Tua: Jurnal Pendidikan dan Pembelajaran*, 1(1), 36–43. <https://doi.org/10.31970/gurutua.v1i1>
- Bjerke, A. H. (2026). Sources of mathematics self-efficacy in primary and secondary students: A systematic review. *Education Sciences*, 16(2), 182. <https://doi.org/10.3390/educsci16020182>

-
- Cheung, G.W., Cooper-Thomas, H.D., Lau, R.s., & Wang, L.C. (2024). Reporting reliability, convergent and discriminant validity with structure equation modelling: A review and best-practice recommendations. *Asia Pasific Journal of Management*, 41(2), 745-783.
- De Ayala, R. J. (2022). *The theory and practice of item response theory* (2nd ed.). Guilford Press.
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). Sage.
- Dinas Pendidikan Kabupaten Belitung Timur, 2025. Laporan Rapor Pendidikan Kabupaten Belitung Timur Tahun 2025.
- Dinni, H. N. (2021). HOTS (High Order Thinking Skills) dan Kaitannya dengan Kemampuan Literasi Matematika. *PRISMA: Prosiding Seminar Nasional Matematika*, 1, Universitas Negeri Semarang.
- Dirgantoro, K. P. S. (2018). Kompetensi Guru Matematika Dalam Mengembangkan Kompetensi Matematis Siswa. *Scholaria: Jurnal Pendidikan dan Kebudayaan*, 8(2), 157-166. <https://doi.org/10.24246/j.js.2018.v8.i2.p157-166>
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Psychology Press
- Fachrudin, A.D. (2022). Modul Pelatihan Peningkatan Kompetensi Numerasi untuk Guru Pengetahuan Numerasi: Proses, Konten, dan Konteks. Jakarta: Pusat Asesmen dan Pembelajaran, Badan Penelitian dan Pengembangan dan Perbukuan Kementerian Pendidikan dan Kebudayaan
- Fiorella, L., Yoon, S.Y., Atit, K., Power, J.R., Panther, G., Sorby, S., Uttal, D.H., & Veurink, N. (2021). Validation of the Mathematics Motivation Questionnaire (MMQ) for secondary school students. *International Journal of STEM Education*, 8 (52). doi: 10.1186/s40594-021-00307-x
- Fitriyah, L. (2024). The relationship between self-efficacy and students' mathematical problem-solving ability. *Golden Ratio of Social Science and Education*, 4(1), 45-56. <https://doi.org/10.52970/grsse.v4i1.421>
- Granello, D. H. (2025). Self-efficacy and student engagement in mathematics learning: A systematic review. *Educational Psychology Review*, 37(1), 55-73. <https://doi.org/10.1007/s10648-024-09745-1>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th Ed.). Cengage Learning.
- Hamdi, S., Rahman, E. F. S. A., Laksita, G. D., Qamariah, N. (2026). *Analisis faktor menggunakan JASP, Jamnovi, dan R implementasinya dalam pendidikan matematika*. Filosofis Indonesia Press.
- Hasana, N., & Sylvia, I. (2025). Pengembangan soal penalaran peserta didik pada pembelajaran sosiologi fase F SMA N 1 Lengayang. *Naradidik: Journal of Education & Pedagogy*, 4(2), 229-236. <https://doi.org/10.24014/nara.v4i2.282>
- Issayeve, L. (2023). *The graded response model – Samejima (1969)*. Assessment Systems. <https://assess.com/graded-response-model/>
- Istiyono, E. (2020). Pengembangan Instrumen Penilaian dan Analisis Hasil Belajar Fisika dengan Teori Tes Klasik dan Modern. UNY Press.
- Istiyono, E., Dwandaru, W. B., & Rahayu, F. (2018). *Pengembangan tes creative thinking skills fisika SMA (PhysCreTHOTS) berdasarkan teori tes modern*. Program Pascasarjana, Universitas Negeri Yogyakarta.
- Istiyono, E., Setiawan, R., & Harun. (2020). Pelatihan penyusunan instrumen tes dan analisisnya secara modern bagi guru-guru IPA SMP. *J. Pengabdian Masyarakat MIPA dan Pendidikan MIPA*, 4(1), 102-108.
- Istiyono, E., Supahar, Pebriana, I.N. (2024). Teori Respon Butir Contoh dan Aplikasinya pada Pembelajaran dan Penelitian. UNY Press.
- Izzah, K. H., & Azizah, M. (2019). Analisis Kemampuan Penalaran Siswa dalam Pemecahan Masalah Matematika Siswa Kelas IV. *Indonesian Journal Of Educational Research and Review*, 2(2), 210. <https://doi.org/10.23887/ijerr.v2i2.17629>
- Joshi, A., Kale, S., Chandel, S., & Pal, D. (2015). Likert scale: Explored and explained. *British Journal of Applied Science & Technology*, 7(4), 396-403
- Lehikoinen, A., Räsänen, P., & Aunola, K. (2025). Mathematics self-efficacy and performance among students: A longitudinal study. *Scandinavian Journal of Educational Research*, 69(2), 215-230. <https://doi.org/10.1080/00313831.2025.2558722>
- Listiawani, Z., Berlian, M., Anwar, A., & Vebrianto, R. (2024). Pengembangan instrumen untuk mengukur motivasi belajar siswa sekolah dasar. *JIIP (Jurnal Ilmiah Ilmu Pendidikan)*, 7(8), 8836-8844. <https://doi.org/10.52208/klasikal.v5i1.629>
- Mediartika, N., & Aznam, N. (2018). Pengembangan instrumen penilaian portofolio berbasis multiple intelligence untuk mengukur kemampuan berpikir kritis dan sikap ilmiah. *Jurnal Inovasi Pendidikan IPA*, 4(1), 52-63.
- Mudanta, K. A., Astawan, I. G., & Jayanta, I. N. L. (2020). Instrumen penilaian motivasi belajar dan hasil belajar IPA siswa kelas V sekolah dasar. *Jurnal Mimbar Ilmu*, 25(2), 262-270.
-

-
- Mukhoiyaroh, M., & Aziz, Y. (2024). Effects of Student Self-Determination Motivation on Student Academic Performance. *Scaffolding: Jurnal Pendidikan Islam dan Multikulturalisme*, 6(3), 104-122. DOI: 10.37680/scaffolding.v6i3.5150
- Novita S, Setyowibowo H, Wijayanti PAK, Erwina W, Yudiana W, Purba FD, et al. (2025). Exploring the impact of digital reality based mathematics learning on students' motivation: Protocol for a systematic review and meta-analysis. *PLoS One*, 20(4): e0319664. doi:10.1371/jurnal.pone.0319664
- OECD. (2023). PISA 2022 Results (Volume I and II) - Country Notes: Indonesia https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/indonesia_c2e1ae0e-en.html
- Ouyang, J., Chen, Y., Li, C., & Xu, G. (2025). *Statistical analysis of large-scale item response data under measurement non-invariance: A statistically consistent method and its application to PISA 2022* (arXiv:2505.16608v2). arXiv. <https://doi.org/10.48550/arXiv.2505.16608>
- Rachmawati, E., & Siswanto, S. (2025). Developing an instrument to measure students' anti-corruption character using modern test theory. *Jurnal Penelitian dan Evaluasi Pendidikan*, 29(1), 26-42 doi: <https://doi.org/10.21831/pep.v29i1.84560>
- Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Annual Review of Organizational Psychology and Organizational Behavior*, 7, 223-248. <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Santoso, H., Marlina, E., & Putri, R. (2024). Application of item response theory in educational assessment: A psychometric perspective. *Journal of Educational Measurement*, 61(1), 45-60. <https://doi.org/10.1111/jedm.1234>
- Schunk, D. H., Meece, J. L., & Pintrich, P. R. (2014). *Motivation in education: Theory, research, and applications* (4th ed.). Pearson.
- Selçuk, E., & Demir, E. (2024). Comparison of item response theory ability and item parameters according to classical and Bayesian estimation methods. *International Journal of Assessment Tools in Education (IJATE)*, 11(2), 213-248. <https://doi.org/10.21449/ijate.1290831>
- Sivrikaya, A. H. (2019). The Relationship between Academic Motivation and Academic Achievement of the Students. *Asian Journal of Education and Training*, 5(2), 309-315. DOI: 10.20448/journal.522.2019.52.309.315
- Street, K. E. S., Malmberg, L. E., & Schukajlow, S. (2024). Students' mathematics self-efficacy: A scoping review. *ZDM—Mathematics Education*, 56(3), 455-472. <https://doi.org/10.1007/s11858-024-01548-0>
- Suhudi, Radeswandri, Herlinda, & Vebrianto, R. (2024). Pengembangan instrumen motivasi belajar siswa: Kuesioner. *Jurnal Gentala Pendidikan Dasar*, 9(1), 83-95.
- Sürücü, L., & Maslakçı, A. (2020). Validity and Reliability in Quantitative Research. *Business & Management Studies: An International* <https://doi.org/10.15295/bmij.v8i3.1540>
- Thiagarajan, S., Semmel, D., & Semmel, M. (1974). *Instructional development for training teachers of exceptional children*. Indiana University.
- Thompson, N. (2023). Classical Test Theory vs. Item Response Theory. *Assessment Systems*. <https://assess.com/what-is-item-response-theory/>
- Wang, X., Zhang, S., & Xin, T. (2022). Item Response Theory Analysis of the Dark Factor of Personality Scale for College Students in China. *International Journal of Environmental Research and Public Health*, 19(19), 12787.
-