

Pengembangan Dan Validasi Instrumen Asesmen Berpikir Kreatif Matematis Berbasis Soal *Open-Ended* Menggunakan Pendekatan Classical Test Theory

Dewi Wahyuni^{1)*}

1) Akuntansi, Ekonomi dan Bisnis, Universitas Dharmawangsa, Indonesia

*Corresponding Email: dewi.wahyuni@dharmawangsa.ac.id

ABSTRAK - Fokus penelitian ini adalah menghasilkan perangkat asesmen untuk mengukur kemampuan berpikir kreatif matematis siswa kelas XI SMK melalui penyusunan soal-soal *open-ended*, kemudian menelaah kelayakan dan validitas instrumen tersebut dengan menggunakan pendekatan *Classical Test Theory* (CTT). Tahap awal penelitian difokuskan pada identifikasi kebutuhan sebagai landasan dalam merancang produk yang akan dikembangkan. Untuk mencapai tujuan tersebut, penelitian menerapkan pendekatan Research and Development (R&D), penyusunan konstruk dan indikator berpikir kreatif matematis, pengembangan butir soal, validasi ahli, uji coba instrumen kepada 32 siswa, serta analisis karakteristik butir berdasarkan CTT yang mencakup validitas, reliabilitas, daya pembeda, dan tingkat kesukaran. Evaluasi terhadap kualitas instrumen memperlihatkan bahwa seluruh butir soal telah memenuhi kriteria yang ditetapkan. Koefisien validitas setiap butir berada pada rentang 0,976–0,981, sedangkan hasil pengujian konsistensi internal menghasilkan nilai Cronbach's Alpha sebesar 0,984 yang mengindikasikan tingkat reliabilitas sangat tinggi. Analisis karakteristik butir juga menunjukkan bahwa kemampuan soal dalam membedakan kemampuan peserta didik berada pada kategori cukup hingga sangat baik, sementara seluruh butir tergolong memiliki tingkat kesukaran sedang. Hasil analisis kebutuhan juga menunjukkan bahwa guru memiliki kemampuan mengembangkan instrumen secara mandiri, namun terkendala beban kerja, keterbatasan waktu, dan tuntutan penyelesaian materi pembelajaran. Seluruh hasil evaluasi menunjukkan bahwa instrumen yang disusun memiliki kualitas psikometrik yang memadai untuk diterapkan sebagai alat ukur kemampuan berpikir kreatif matematis peserta didik. Kehadiran instrumen ini diharapkan dapat mempermudah guru dalam memperoleh informasi mengenai kemampuan siswa sekaligus mendukung pelaksanaan pembelajaran yang menekankan pengembangan keterampilan berpikir tingkat tinggi.

Kata Kunci: Asesmen Berpikir Kreatif, Open-Ended, Classical Test Theory

ABSTRACT - This study aims to develop an assessment instrument for measuring the mathematical creative thinking skills of eleventh-grade vocational high school (SMK) students through the construction of open-ended questions and to examine the feasibility and validity of the instrument using the Classical Test Theory (CTT) approach. The initial stage of the study focused on conducting a needs analysis as the basis for designing the assessment instrument. To achieve these objectives, the

research employed a Research and Development (R&D) approach, which included constructing the framework and indicators of mathematical creative thinking, developing test items, conducting expert validation, administering a pilot test to 32 students, and analyzing the psychometric characteristics of the items based on CTT, including validity, reliability, discrimination index, and difficulty level. The evaluation results indicated that all test items met the established quality criteria. The item validity coefficients ranged from 0.976 to 0.981, while the internal consistency analysis yielded a Cronbach's Alpha coefficient of 0.984, indicating a very high level of reliability. Furthermore, the item analysis demonstrated that the discrimination indices ranged from fair to very good, and all items were classified as having a moderate level of difficulty. The needs analysis also revealed that teachers possessed the capability to develop assessment instruments independently; however, they encountered constraints related to workload, limited time, and the demands of completing the instructional content. Overall, the findings indicate that the developed instrument possesses adequate psychometric properties and is suitable for assessing students' mathematical creative thinking skills. The availability of this instrument is expected to assist teachers in obtaining accurate information regarding students' creative thinking abilities while supporting the implementation of learning activities that emphasize the development of higher-order thinking skills.

Keywords: *Creative Thinking Assessment, Open-Ended Questions, Classical Test Theory (CTT).*

PENDAHULUAN

Tuntutan pendidikan pada abad ke-21 menempatkan kemampuan berpikir kreatif matematis sebagai salah satu kompetensi kunci yang perlu dikembangkan pada diri setiap peserta didik. Kemampuan ini memungkinkan peserta didik menghasilkan berbagai alternatif penyelesaian masalah, mengembangkan ide-ide baru, serta memberikan solusi yang beragam terhadap suatu permasalahan matematika. Pembelajaran matematika menempatkan kemampuan berpikir kreatif sebagai aspek yang tidak hanya dinilai dari keberhasilan peserta didik memperoleh jawaban yang benar. Kemampuan tersebut juga terlihat dari kecakapan dalam merumuskan beragam strategi penyelesaian, melahirkan ide-ide yang orisinal, serta mengembangkan proses penyelesaian secara terstruktur dan logis. Kemampuan berpikir kreatif perlu mendapat perhatian khusus dalam pembelajaran matematika karena berperan penting dalam membentuk peserta didik yang mampu menyesuaikan diri terhadap perubahan dan menghasilkan solusi yang inovatif. Pada tingkat SMK, kompetensi tersebut menjadi semakin relevan sebagai bekal untuk menghadapi tuntutan dan persaingan di lingkungan kerja.

Namun, kondisi ideal tersebut masih belum sepenuhnya terwujud dalam praktik pembelajaran. Hasil penggalan informasi melalui observasi awal dan wawancara bersama guru matematika di SMK Samudera Indonesia mengindikasikan bahwa kemampuan berpikir kreatif matematis yang dimiliki siswa kelas XI belum berkembang secara optimal. Kemampuan siswa dalam menyelesaikan soal cenderung terbatas pada penerapan prosedur yang telah dipelajari dari guru. Ketika diberikan permasalahan yang dapat diselesaikan melalui lebih dari satu pendekatan, sebagian besar siswa belum mampu menentukan dan menerapkan strategi penyelesaian yang beragam. Ketika diminta mencari alternatif cara lain atau mengembangkan jawaban yang berbeda, sebagian besar siswa cenderung berhenti setelah memperoleh satu jawaban yang dianggap benar. Temuan tersebut mengindikasikan bahwa kemampuan berpikir kreatif matematis peserta didik masih memerlukan penguatan, terutama pada aspek fluency, flexibility, originality, dan elaboration yang menjadi indikator utama dalam menilai kreativitas matematis.

Kondisi rendahnya kemampuan berpikir kreatif matematis dapat dipengaruhi oleh berbagai faktor. Selain karakteristik peserta didik, penggunaan instrumen asesmen yang belum memadai dalam proses pembelajaran juga diduga menjadi salah satu penyebabnya. Temuan hasil pengamatan di lapangan memperlihatkan bahwa sumber belajar yang dimanfaatkan selama proses pembelajaran masih. nformasi yang diperoleh dari kegiatan observasi mengungkapkan bahwa sumber belajar yang digunakan, baik buku paket maupun bahan ajar pendukung, masih didominasi oleh soal-soal rutin yang berorientasi pada penerapan rumus dan pencarian satu jawaban benar. Demikian pula contoh soal yang diberikan guru dalam proses pembelajaran sebagian besar masih bersifat tertutup (close-ended), sehingga peserta didik belum memiliki ruang yang memadai untuk mengeksplorasi berbagai strategi penyelesaian maupun menyampaikan ide-ide yang beragam. Akibatnya, peserta didik terbiasa berpikir prosedural dan kurang terlatih menghadapi permasalahan yang membutuhkan kreativitas matematis.

Keterbatasan instrumen asesmen juga menyebabkan guru belum dapat memperoleh gambaran yang objektif mengenai kemampuan berpikir kreatif yang

dimiliki peserta didik. Penilaian yang dilakukan selama ini lebih berorientasi pada hasil akhir berupa benar atau salahnya jawaban, sedangkan proses berpikir peserta didik, keberagaman strategi penyelesaian, keaslian ide, dan kemampuan mengembangkan solusi belum menjadi aspek yang dinilai secara sistematis. Kondisi tersebut mengakibatkan guru belum memperoleh gambaran secara utuh tentang tingkat kemampuan berpikir kreatif matematis yang dimiliki siswa sehingga penyusunan tindak lanjut pembelajaran yang disesuaikan dengan kebutuhan belajar peserta didik menjadi lebih sulit.

Permasalahan tersebut mencerminkan belum selarasnya pelaksanaan asesmen dengan arah kebijakan kurikulum yang berfokus pada pengembangan keterampilan berpikir tingkat tinggi, sementara praktik penilaian yang diterapkan tetap berorientasi pada penguasaan prosedur. Untuk merespons permasalahan yang telah diidentifikasi, penelitian ini mengembangkan instrumen asesmen yang disusun berdasarkan karakteristik soal open-ended. Melalui soal open-ended, peserta didik didorong menyelesaikan suatu permasalahan menggunakan berbagai strategi yang benar. Pendekatan ini memungkinkan guru memperoleh gambaran yang lebih utuh mengenai tingkat kreativitas berpikir matematis dibandingkan penggunaan soal konvensional. Proses pengembangan dan validasi yang dilaksanakan secara sistematis diperlukan agar instrumen yang dihasilkan memenuhi standar kualitas psikometrik. Melalui proses tersebut, instrumen diharapkan memiliki validitas dan reliabilitas yang memadai serta menghasilkan pengukuran yang tepat terhadap konstruk kemampuan berpikir kreatif matematis.

Berdasarkan uraian tersebut, penelitian ini berfokus pada pengembangan serta validasi instrumen asesmen berpikir kreatif matematis yang disusun menggunakan karakteristik soal bertipe open-ended. Instrumen yang dihasilkan diharapkan mampu memberikan pengukuran terhadap tingkat kreativitas berpikir matematis siswa kelas XI SMK Samudera Indonesia secara lebih akurat. Instrumen yang dihasilkan diharapkan dapat membantu guru memperoleh informasi yang akurat mengenai kemampuan berpikir kreatif matematis peserta didik sekaligus menjadi alternatif asesmen yang mendukung pelaksanaan pembelajaran matematika guna mengembangkan kemampuan berpikir secara lebih kompleks dan kritis.

KAJIAN TEORI

Asesmen berpikir kreatif matematis dilakukan untuk mengidentifikasi kemampuan peserta didik dalam mengembangkan berbagai gagasan, menerapkan alternatif strategi, serta menghasilkan penyelesaian matematika yang menunjukkan unsur kebaruan dan memiliki nilai kebermaknaan. Berbeda dengan asesmen konvensional yang berfokus pada ketepatan jawaban akhir, asesmen berpikir kreatif menekankan bagaimana peserta didik membangun, mengembangkan, dan mengevaluasi berbagai kemungkinan penyelesaian suatu masalah. Hasil penelitian terbaru menunjukkan bahwa asesmen berpikir kreatif memberikan perhatian pada dua aspek utama, yaitu produk yang dihasilkan dan tahapan berpikir yang dilakukan peserta didik ketika mencari penyelesaian terhadap suatu permasalahan (Sa'dijah, C. dkk., 2016).

Secara teoritis, asesmen berpikir kreatif matematis bertumpu pada empat aspek utama kreativitas yang telah menjadi acuan dalam berbagai kajian pendidikan matematika, yaitu kemampuan menghasilkan banyak gagasan (*fluency*), menggunakan beragam pendekatan (*flexibility*), menampilkan kebaruan ide (*originality*), serta mengembangkan gagasan secara terperinci (*elaboration*).

Aspek *fluency* berkaitan dengan kecakapan menghasilkan beragam alternatif pemikiran atau jawaban. Aspek *flexibility* menunjukkan kemampuan beralih dan menyesuaikan penggunaan berbagai pendekatan penyelesaian sesuai karakteristik masalah. Sementara itu, *originality* menggambarkan kemampuan melahirkan solusi yang memiliki unsur kebaruan, sedangkan *elaboration* merujuk pada kemampuan mengembangkan serta menguraikan suatu gagasan secara runtut dan logis. Kajian terbaru menunjukkan bahwa keempat indikator tersebut masih menjadi kerangka yang paling dominan dalam penyusunan rubrik penskoran maupun pengembangan instrumen asesmen berpikir kreatif matematis Sipahi, Y., & Bahar, A. K. (2025).

Bentuk instrumen yang digunakan dalam asesmen berpikir kreatif matematis harus mampu memberikan ruang kepada peserta didik untuk berpikir divergen. Oleh karena itu, soal open-ended direkomendasikan karena memungkinkan adanya berbagai alternatif jawaban dan strategi penyelesaian yang

sama-sama benar.

Melalui penerapan soal semacam ini, fokus penilaian guru tidak terbatas pada ketepatan jawaban akhir. Penilaian juga mencakup cara peserta didik membangun proses penalaran, menunjukkan kreativitas, serta menyampaikan alasan matematis yang mendukung penyelesaian masalah. Penelitian Systematic Literature Review mengenai pendekatan open-ended menunjukkan bahwa implementasi soal terbuka secara berkelanjutan terbukti lebih efektif dalam mengembangkan kemampuan berpikir kreatif matematis daripada penggunaan soal tertutup yang berorientasi pada satu jawaban benar (Dewi & Juandi, 2023. Kwon dkk. (2021) menegaskan bahwa pendekatan *open-ended* berfungsi sebagai katalisator utama dalam menguji dan melejitkan kemampuan berpikir kreatif siswa. Dalam kajian psikometri, kemampuan berpikir kreatif umumnya dinilai berdasarkan empat dimensi utama, yaitu kelancaran (*fluency*), keluwesan (*flexibility*), keaslian (*originality*), dan elaborasi (*elaboration*). Masalah terbuka mendorong siswa untuk menghasilkan beragam alternatif penyelesaian (kelancaran), menerapkan berbagai strategi dalam memecahkan masalah (keluwesan), menemukan cara yang unik atau tidak lazim (keaslian), serta mengembangkan jawaban secara rinci dan sistematis (elaborasi).

Soal open-ended merupakan bentuk instrumen asesmen yang dirancang agar peserta didik memiliki kesempatan menghasilkan beragam alternatif jawaban yang dapat diterima serta menggunakan berbagai strategi dalam menyelesaikan masalah yang dapat dipertanggungjawabkan secara matematis. Berbeda dengan soal tertutup (close-ended) yang umumnya hanya memiliki satu prosedur dan satu jawaban benar, soal open-ended dirancang untuk memberi kesempatan kepada peserta didik mengeksplorasi beragam gagasan, menerapkan beraneka pendekatan penyelesaian, serta mengemukakan alasan berdasarkan pengetahuan dan pengalaman yang dimiliki (Suherman & Vidákovich, 2022; Takahashi, 2021).

Menurut Becker dan Shimada (2019), pendekatan open-ended merupakan strategi instruksional yang menghadapkan siswa pada masalah autentik yang memungkinkan penggunaan beraneka strategi penyelesaian serta menghasilkan lebih dari satu alternatif jawaban yang sah secara matematis (multiple solution methods). Landasan epistemologis dari pendekatan ini adalah teori

konstruktivisme, di mana fokus pembelajaran beralih dari sekadar menghafal algoritma baku menuju pengembangan fleksibilitas kognitif. Melalui interaksi dengan masalah terbuka, siswa didorong untuk mengintegrasikan pengetahuan lama yang telah mereka miliki guna mengonstruksi strategi pemecahan masalah baru yang unik dan kreatif.

Dalam kajian yang dilakukan oleh Sullivan dkk. (2020), masalah *open-ended* diidentifikasi sebagai instrumen krusial dalam mewujudkan diferensiasi pembelajaran (*differentiated instruction*) yang inklusif. Karakteristik utama dari tugas terbuka adalah memiliki "lantai rendah dan atap tinggi" (*low floor, high ceiling*). Artinya, tugas tersebut sangat mudah diakses oleh siswa dengan kemampuan akademik rendah agar mereka dapat mulai bekerja (lantai rendah), namun pada saat yang sama menawarkan ruang eksplorasi yang sangat luas dan menantang bagi siswa berkemampuan tinggi untuk mencapai tingkat berpikir yang lebih kompleks (atap tinggi). Hal ini meminimalkan kecemasan belajar dan meningkatkan partisipasi aktif seluruh siswa di kelas. Berdasarkan analisis teoretis dari Schoenfeld (2022), implementasi pendekatan *open-ended* menuntut pergeseran paradigma peran pendidik dari pusat informasi (*teacher-centered*) menjadi fasilitator dan perancang lingkungan belajar (*student-centered*). Guru tidak lagi berfungsi sebagai pemberi justifikasi "benar atau salah", melainkan sebagai pemandu yang menstimulasi kesadaran metakognitif siswa.

Melalui teknik bertanya produktif (*scaffolding questioning*), guru mengarahkan siswa untuk merefleksikan kembali proses berpikir mereka sendiri, membandingkan efisiensi antarstrategi yang ditemukan, dan menarik kesimpulan konseptual secara mandiri. Menurut Brookhart (2024), tugas-tugas *open-ended* merupakan instrumen asesmen dan pembelajaran yang paling ideal untuk menilai kemampuan berpikir tingkat tinggi (*Higher Order Thinking Skills* atau *HOTS*), terutama pada aspek analisis (C4), evaluasi (C5), dan kreasi (C6). Karena masalah tidak dapat dipecahkan hanya dengan menerapkan rumus hafalan secara mekanis, siswa dipaksa untuk melakukan analisis mendalam terhadap struktur masalah, mengevaluasi validitas dari berbagai alternatif solusi yang mereka temukan, dan akhirnya menciptakan model penyelesaian baru mereka sendiri. soal *open-ended* berfungsi sebagai stimulus yang mendorong peserta didik berpikir divergen.

Melalui soal tersebut, siswa tidak diarahkan untuk mengikuti satu prosedur tertentu, tetapi diberi kebebasan mengembangkan berbagai kemungkinan penyelesaian berdasarkan pemahaman konsep yang dimiliki. Kondisi ini memungkinkan guru memperoleh informasi yang lebih komprehensif mengenai kemampuan siswa dalam menghasilkan berbagai gagasan, menentukan beragam pendekatan penyelesaian, mengembangkan solusi secara logis, serta menyampaikan alasan matematis yang mendukung jawaban mereka. Dengan demikian, soal open-ended tidak sekadar dimanfaatkan untuk menilai hasil belajar, melainkan juga menjadi media yang efektif dalam mengungkap proses berpikir kreatif peserta didik (Sipahi, Y., & Bahar, A. K., 2025).

Dalam bidang psikometri, *Classical Test Theory* (CTT) menjadi salah satu pendekatan yang paling luas dimanfaatkan untuk mendukung pengembangan dan penilaian kualitas instrumen pendidikan maupun psikologi. Dalam *Classical Test Theory* (CTT), diasumsikan bahwa *observed score* yang diperoleh peserta tes tersusun atas dua komponen utama, yaitu skor sesungguhnya (*true score*) serta galat pengukuran (*error score*). Oleh karena itu, kualitas suatu instrumen ditentukan oleh kemampuan skor hasil pengukuran dalam mencerminkan kompetensi yang sesungguhnya dengan tingkat kesalahan pengukuran yang serendah mungkin. Meskipun berkembang berbagai pendekatan modern, CTT masih menjadi metode yang dominan karena prosedurnya relatif sederhana, mudah diterapkan, serta efektif untuk mengevaluasi kualitas instrumen pada tahap pengembangan awal DeMars, C. E. (2018). DeVellis (2022) mengonseptualisasikan *Classical Test Theory* (CTT) sebagai kerangka kerja psikometrik yang mengasumsikan setiap skor empiris yang diperoleh subjek merupakan refleksi linier dari akumulasi skor murni (*true score*) dan galat acak (*random error*). Secara matematis, formulasi ini diwujudkan dalam persamaan $X = T + E$. Karakteristik fundamental dari pendekatan CTT terletak pada orientasinya yang berbasis kelompok (*sample-bound*), di mana estimasi parameter alat ukur diturunkan dari distribusi varians skor tampak di dalam sampel spesifik. Keunggulan taktis dari pendekatan ini adalah efisiensinya dalam pengujian skala dengan ukuran sampel yang relatif moderat, menjadikannya metodologi yang sangat aplikatif dalam pengembangan instrumen psikologi konvensional. Menurut

Crocker dan Algina (2020), fokus utama dari aplikasi CTT adalah pada optimalisasi koefisien reliabilitas instrumen melalui minimalisasi varians galat (σ^2_e). Koefisien reliabilitas dalam paradigma CTT dipandang sebagai proporsi varians skor murni terhadap varians skor tampak. Pendekatan ini mengandalkan metrik konsistensi internal seperti koefisien *Cronbach's Alpha* serta *Standard Error of Measurement* (SEM) untuk mengkuantifikasi derajat ketidakpastian atau fluktuasi skor yang disebabkan oleh galat pengukuran. Dengan meminimalkan SEM, instrumen dapat menghasilkan estimasi kemampuan laten subjek yang lebih stabil dan reproduibel pada kondisi pengujian yang setara. Penfield (2021) memberikan analisis kritis terhadap CTT dengan menyoroti sifat ketergantungan parameter butir pada karakteristik sampel (*sample dependency*).

Dalam model CTT, indeks tingkat kesukaran (proporsi jawaban benar atau *p-value*) dan daya pembeda (korelasi poin-biserial) tidak bersifat invarian, melainkan sangat fluktuatif mengikuti tingkat kemampuan (*ability*) kelompok responden yang diuji. Fenomena ini mengimplikasikan bahwa sebuah butir soal dapat teridentifikasi memiliki daya beda tinggi dan tingkat kesulitan rendah ketika diujikan pada kelompok populasi dengan kapabilitas superior, namun akan menunjukkan metrik sebaliknya jika diujikan pada kelompok dengan kapabilitas inferior. Dalam ranah evaluasi pendidikan operasional, Downing dan Haladyna (2023) menegaskan bahwa CTT tetap menjadi instrumen evaluatif yang paling pragmatis dan efektif. Keunggulan metodologis CTT terletak pada kemampuannya untuk mendeteksi fungsionalitas efektivitas pengecoh (*distractor efficiency*) secara langsung. Pengembang tes dapat dengan cepat mengidentifikasi butir soal yang cacat secara psikometrik melalui kalkulasi indeks diskriminasi tanpa harus memenuhi asumsi matematis yang rigid dan kompleks seperti pada Item Response Theory (IRT). Hal ini menjadikan CTT sebagai standar utama dalam penjaminan mutu asesmen formatif dan sumatif di lembaga pendidikan.

Dalam pengembangan instrumen tes, Classical Test Theory (CTT) berperan sebagai pendekatan untuk menganalisis kualitas setiap butir soal dengan mempertimbangkan aspek validitas butir, reliabilitas tes, tingkat kesukaran, serta daya pembeda. Validitas butir menunjukkan sejauh mana setiap butir mampu mengukur konstruk yang dimaksud, sedangkan reliabilitas menggambarkan

konsistensi hasil pengukuran apabila instrumen digunakan pada kelompok yang memiliki karakteristik serupa. Parameter tingkat kesukaran berfungsi untuk menggambarkan banyaknya peserta didik yang dapat menjawab suatu butir soal secara benar, sedangkan parameter daya pembeda menunjukkan efektivitas butir soal dalam mengidentifikasi perbedaan kemampuan antara peserta didik berkemampuan tinggi dan peserta didik berkemampuan rendah. Kombinasi keempat analisis tersebut memungkinkan peneliti menyeleksi butir yang layak dipertahankan, direvisi, atau dihapus sehingga instrumen yang dihasilkan memiliki kualitas psikometrik yang baik DeMars, C. E. (2018).

METODE PENELITIAN

Pendekatan *Research and Development* (R&D) diterapkan dalam penelitian ini untuk mengembangkan dan memvalidasi instrumen asesmen berpikir kreatif matematis berbasis soal open-ended yang memenuhi persyaratan validitas isi dan validitas empiris. Selain itu, instrumen yang dihasilkan diharapkan memiliki konsistensi pengukuran yang tinggi, karakteristik tingkat kesukaran dan daya pembeda yang memadai, serta mampu memberikan hasil pengukuran yang akurat terhadap kemampuan berpikir kreatif matematis siswa Sekolah Menengah Kejuruan (SMK). Penerapan metode *Research and Development* (R&D) dalam penelitian ini didasarkan pada tujuan untuk menghasilkan instrumen asesmen sekaligus melaksanakan serangkaian tahapan pengembangan, pengujian, evaluasi, dan penyempurnaan secara sistematis hingga diperoleh instrumen yang memiliki kualitas teoritis dan empiris yang dapat dipertanggungjawabkan (Borg et al., 2003; DeVellis & Thorpe, 2021).

Secara konseptual, landasan yang digunakan dalam studi ini adalah pendekatan *Research and Development* (R&D) yang diperkenalkan oleh Borg dan Gall. Menurut mereka, pendekatan tersebut merupakan metode yang berorientasi pada pengembangan suatu produk di bidang pendidikan sekaligus melakukan penilaian terhadap kelayakannya melalui tahapan penelitian yang disusun secara sistematis. Dalam penelitian ini, produk yang dikembangkan bukan berupa media pembelajaran atau perangkat ajar, melainkan instrumen asesmen berpikir kreatif matematis berbasis soal open-ended. Oleh karena itu, prosedur penelitian disesuaikan dengan karakteristik pengembangan instrumen

pendidikan (Borg et al., 2003).

Tahapan pengembangan instrumen mengadopsi kerangka yang dikemukakan oleh DeVellis yang meliputi penentuan konstruk yang akan diukur, penyusunan definisi operasional, pengembangan indikator, penyusunan kisi-kisi instrumen, penyusunan butir soal open-ended, penyusunan rubrik penskoran, validasi isi oleh para ahli, uji coba instrumen, analisis karakteristik butir menggunakan pendekatan *Classical Test Theory* (CTT), revisi instrumen, dan penyusunan instrumen final. Kerangka tersebut dipilih karena memberikan prosedur yang sistematis dalam menghasilkan instrumen yang memiliki kualitas psikometrik yang baik dan sesuai dengan tujuan pengukuran (DeVellis & Thorpe, 2021).

Penelitian dilaksanakan di SMK Samudera Indonesia selama semester genap tahun ajaran 2025/2026. Seluruh peserta didik di sekolah tersebut menjadi sasaran populasi penelitian, sementara uji coba awal (pilot study) melibatkan 32 siswa kelas XI sebagai sampel penelitian. Pemilihan sampel ini bertujuan untuk memperoleh data awal mengenai keterbacaan soal, kejelasan bahasa, kejelasan petunjuk pengerjaan, kesesuaian alokasi waktu, serta karakteristik awal butir soal sebelum instrumen digunakan secara lebih luas (DeVellis & Thorpe, 2021).

Penelitian diawali dengan analisis kebutuhan melalui observasi proses pembelajaran matematika, wawancara dengan guru matematika, analisis kurikulum, analisis buku ajar, serta analisis soal evaluasi yang selama ini digunakan di sekolah. Tahap ini bertujuan mengidentifikasi kebutuhan terhadap instrumen yang mampu mengukur kemampuan berpikir kreatif matematis secara lebih komprehensif. emuan dari analisis kebutuhan selanjutnya digunakan sebagai dasar dalam menyusun konstruk instrumen yang mengacu pada empat dimensi kemampuan berpikir kreatif matematis, meliputi fluency yang berkaitan dengan kelancaran menghasilkan ide, *flexibility* yang menunjukkan kemampuan menggunakan beragam strategi, *originality* yang menggambarkan kebaruan gagasan, serta *elaboration* yang berkaitan dengan pengembangan ide secara terperinci. Selanjutnya, keempat dimensi tersebut diterjemahkan ke dalam kisi-kisi instrumen sebagai acuan dalam merancang butir-butir soal bertipe open-ended (Assessment of Mathematical Creative Thinking: A Systematic Review;

Mathematical Creativity: A Systematic Review of Definitions, Frameworks, and Assessment Practices).

Instrumen yang telah dikembangkan kemudian melalui proses penelaahan oleh sejumlah pakar yang meliputi dosen bidang pendidikan matematika, pakar evaluasi pendidikan, serta praktisi yang mengajar mata pelajaran matematika. Proses penelaahan tersebut bertujuan mengevaluasi kecocokan substansi, konstruksi, penggunaan bahasa, dan keterwakilan setiap indikator terhadap konstruk yang hendak diukur. Evaluasi validitas isi pada setiap butir soal dilakukan dengan memanfaatkan koefisien Aiken's V. Melalui prosedur tersebut, tingkat kesesuaian setiap butir dapat ditentukan berdasarkan penilaian para ahli sebelum instrumen diujicobakan kepada peserta didik (Richard R. Aiken, 1985).

Tahap berikutnya adalah uji coba awal instrumen kepada 32 siswa kelas XI. Data hasil uji coba dianalisis menggunakan pendekatan Classical Test Theory (CTT) yang meliputi analisis validitas butir, reliabilitas internal menggunakan koefisien Alpha Cronbach, indeks tingkat kesukaran, serta daya pembeda. Analisis tersebut bertujuan untuk mengidentifikasi kualitas setiap butir soal sehingga dapat ditentukan butir yang dipertahankan, direvisi, maupun dieliminasi. Hasil analisis kemudian digunakan sebagai dasar penyempurnaan instrumen hingga diperoleh instrumen asesmen berpikir kreatif matematis berbasis soal open-ended yang memenuhi persyaratan kualitas instrumen dan layak digunakan dalam proses pembelajaran maupun penelitian (DeVellis & Thorpe, 2021; Lee J. Cronbach, 1951).

Data penelitian diperoleh melalui observasi, wawancara, dokumentasi, lembar validasi ahli, dan tes menggunakan instrumen yang dikembangkan. Analisis terhadap data kualitatif yang diperoleh melalui kegiatan observasi, wawancara, serta saran validator dilakukan dengan menerapkan model analisis interaktif yang dikembangkan oleh Miles, Huberman, dan Saldaña. Proses analisis tersebut mencakup tahapan kondensasi data, penyajian hasil, serta penyusunan kesimpulan. Adapun data kuantitatif diolah dengan menggunakan koefisien Aiken's V untuk menguji validitas isi, sedangkan pendekatan Classical Test Theory diterapkan untuk mengkaji karakteristik butir yang mencakup validitas butir, reliabilitas, tingkat kesukaran, serta daya pembeda sebagai dasar

dalam menilai mutu psikometrik instrumen hasil pengembangan (Matthew B. Miles et al., 2014; Aiken, 1985; DeVellis & Thorpe, 2021).

HASIL DAN PEMBAHASAN

Instrumen penilaian kemampuan berpikir kreatif matematis yang dikembangkan dalam penelitian ini dirancang dengan berlandaskan hasil identifikasi kebutuhan peserta didik pada tahap awal pelaksanaan penelitian. Hasil analisis menunjukkan bahwa diperlukan suatu instrumen yang mampu mengakomodasi keberagaman ide dan strategi penyelesaian masalah, sehingga dipilih bentuk soal *open-ended* sebagai karakteristik utama tes yang dikembangkan. Soal *open-ended* memberi keleluasaan kepada peserta didik dalam memecahkan masalah melalui beragam alternatif jawaban maupun strategi penyelesaian, sehingga lebih mampu menggambarkan kemampuan berpikir kreatif matematis daripada soal yang hanya memiliki satu jawaban benar.

Instrumen yang dikembangkan berupa seperangkat soal uraian yang dirancang untuk mengukur kemampuan berpikir kreatif matematis berdasarkan indikator yang telah ditetapkan serta diselaraskan dengan tujuan pembelajaran yang ingin dicapai. Tes tersebut dimanfaatkan sebagai instrumen penilaian pada ranah kognitif untuk mengetahui sejauh mana kompetensi yang menjadi target pembelajaran telah dikuasai oleh peserta didik. Penyusunan setiap butir soal dilakukan secara sistematis dengan memperhatikan empat indikator utama kemampuan berpikir kreatif matematis, yaitu *fluency* (kemampuan menghasilkan beragam ide), *flexibility* (kemampuan memilih dan menerapkan berbagai pendekatan penyelesaian), *originality* (kemampuan memunculkan gagasan yang unik dan orisinal), serta *elaboration* (kemampuan mengembangkan dan menjelaskan ide secara rinci). Keempat indikator tersebut menjadi landasan dalam penyusunan soal agar instrumen yang dihasilkan mampu mengukur kemampuan berpikir kreatif matematis secara komprehensif.

Setelah proses penyusunan instrumen selesai, tahap selanjutnya adalah melakukan validasi oleh para ahli (*expert judgment*) sebagai upaya untuk memastikan kualitas instrumen sebelum diujicobakan kepada peserta didik. Validasi dilakukan terhadap tiga aspek utama, yaitu kelayakan isi (*content validity*), kelayakan konstruk (*construct validity*), dan kelayakan bahasa serta

format penyajian. Proses validasi melibatkan tiga orang ahli, yaitu dua dosen pada bidang pendidikan matematika dari perguruan tinggi serta seorang guru mata pelajaran matematika yang berpengalaman dalam pelaksanaan pembelajaran dan pengembangan instrumen evaluasi. Revisi dan penyempurnaan instrumen dilakukan dengan mengacu pada masukan serta rekomendasi yang diberikan oleh para validator. Temuan hasil validasi memperlihatkan bahwa instrumen penilaian kemampuan berpikir kreatif matematis telah memenuhi kriteria kelayakan, sehingga dapat diimplementasikan pada tahap uji coba lapangan. Tahapan selanjutnya ujicoba asesmen berpikir kreatif kepada siswa diperoleh hasil validasi sebagai berikut :

Tabel 1
Validitas Asesmen Berpikir Kreatif

		Correlations					
		P1	P2	P3	P4	P5	TOTAL
P1	Pearson Correlation	1	.943**	.939**	.947**	.943**	.976**
	Sig. (2-tailed)		.000	.000	.000	.000	.000
	N	32	32	32	32	32	32
P2	Pearson Correlation	.943**	1	.949**	.941**	.956**	.979**
	Sig. (2-tailed)	.000		.000	.000	.000	.000
	N	32	32	32	32	32	32
P3	Pearson Correlation	.939**	.949**	1	.958**	.961**	.979**
	Sig. (2-tailed)	.000	.000		.000	.000	.000
	N	32	32	32	32	32	32
P4	Pearson Correlation	.947**	.941**	.958**	1	.973**	.981**
	Sig. (2-tailed)	.000	.000	.000		.000	.000
	N	32	32	32	32	32	32
P5	Pearson Correlation	.943**	.956**	.961**	.973**	1	.985**
	Sig. (2-tailed)	.000	.000	.000	.000		.000
	N	32	32	32	32	32	32
TOTAL	Pearson Correlation	.976**	.979**	.979**	.981**	.985**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	32	32	32	32	32	32

** . Correlation is significant at the 0.01 level (2-tailed).

Hasil uji validitas terhadap lima butir soal yang dianalisis menggunakan SPSS dengan melibatkan 32 responden menunjukkan bahwa koefisien validitas pada setiap butir secara berurutan adalah 0,976; 0,976; 0,976; 0,981; dan 0,981. Seluruh koefisien validitas yang diperoleh termasuk dalam kategori sangat tinggi. Selain itu, seluruh nilai tersebut melampaui nilai pembanding berupa r tabel pada taraf signifikansi 5%, sehingga setiap butir soal dapat dinyatakan telah memenuhi persyaratan validitas. Temuan tersebut mengindikasikan bahwa setiap butir soal telah mampu merepresentasikan konstruk yang diukur secara akurat serta menunjukkan hubungan yang kuat dengan skor total instrumen. Berdasarkan hasil

yang diperoleh, instrumen yang dirancang untuk menilai kemampuan berpikir kreatif dalam matematika dinyatakan memiliki tingkat keabsahan yang memadai. Oleh karena itu, instrumen tersebut layak dimanfaatkan oleh guru sebagai alat penilaian untuk mengevaluasi tingkat kreativitas berpikir matematis peserta didik, baik selama pelaksanaan kegiatan belajar mengajar maupun pada tahap evaluasi hasil belajar. Instrumen yang telah memenuhi kriteria validitas belum sepenuhnya dapat digunakan sebelum diketahui tingkat reliabilitasnya. Setelah setiap butir soal terbukti memenuhi persyaratan validitas, tahap berikutnya adalah pengujian reliabilitas. Pengujian ini bertujuan untuk memastikan bahwa instrumen menunjukkan tingkat keterandalan yang memadai, sehingga data yang dihasilkan dari proses penilaian dapat diyakini keakuratannya.

Tabel 2.
Reabilitas Asesemen Berpikir Kreatif

Reliability Statistics	
Cronbach's Alpha	N of Items
.984	5

Hasil pengujian reliabilitas menunjukkan bahwa nilai koefisien reliabilitas berdasarkan Cronbach's Alpha mencapai 0,984. Koefisien yang diperoleh mengindikasikan bahwa instrumen memiliki tingkat keterandalan yang sangat tinggi, karena nilainya melampaui batas minimum yang ditetapkan sebagai persyaratan reliabilitas. Temuan tersebut menunjukkan bahwa antarbutir dalam instrumen memiliki tingkat kesesuaian yang sangat tinggi dalam merepresentasikan konstruk kreativitas berpikir matematis peserta didik. Dengan demikian, data yang dihasilkan melalui penggunaan instrumen ini cenderung konsisten serta memiliki tingkat kepercayaan yang tinggi ketika diterapkan pada kondisi yang relatif sama.

Instrumen yang telah memenuhi kriteria reliabilitas selanjutnya dianalisis berdasarkan daya pembeda setiap butir soal. Analisis daya pembeda dilaksanakan untuk mengetahui sejauh mana setiap butir soal mampu mengklasifikasikan peserta didik berdasarkan tingkat kemampuannya, sehingga instrumen dapat membedakan secara jelas antara kelompok dengan kemampuan tinggi dan kelompok dengan kemampuan rendah.

Tabel 3.
Daya Pembeda Soal

Aspek yang Dinilai	Nomor soal				
	1	2	3	4	5
Daya Pembeda Asesmen	0,57	0,63	0,30	0,27	0,31
Interpertasi	Sangat Baik	Sangat Baik	Baik	Cukup	Baik

Hasil pengujian daya pembeda pada lima butir soal menunjukkan bahwa nilai indeks daya pembeda yang diperoleh secara berturut-turut adalah 0,57; 0,63; 0,30; 0,27; dan 0,31. Hasil pengelompokan berdasarkan indeks daya pembeda menunjukkan bahwa dua soal memiliki kualitas sangat baik, dua soal berkualitas baik, sedangkan satu soal berada pada kategori cukup. Temuan tersebut mengindikasikan bahwa mayoritas butir soal memiliki daya pembeda yang berada pada kategori baik hingga sangat baik, sehingga mampu mengidentifikasi perbedaan tingkat kemampuan peserta didik secara efektif. Sementara itu, satu butir soal yang berkategori cukup masih mampu membedakan kemampuan siswa, meskipun efektivitasnya belum sebaik butir soal lainnya. Secara umum, instrumen yang disusun menunjukkan kualitas daya pembeda yang memadai. Dengan demikian, setiap butir soal dinilai sesuai untuk dimanfaatkan sebagai alat evaluasi dalam menilai kemampuan berpikir kreatif matematis peserta didik.

Nilai daya pembeda yang baik mencerminkan bahwa setiap butir soal dapat mengidentifikasi perbedaan tingkat kemampuan antarpeserta didik secara efektif. Namun demikian, kualitas butir soal tidak hanya ditentukan oleh daya pembeda, tetapi juga oleh tingkat kesukarannya. Selanjutnya, dilakukan analisis terhadap tingkat kesukaran pada setiap butir soal untuk mengetahui sebaran tingkat kesulitan soal, sehingga kualitas instrumen dapat dievaluasi secara lebih menyeluruh.

Tabel 4.
Tingkat Kesukaran Asesmen Berpikir Kreatif

Aspek yang Dinilai	Nomor soal				
	1	2	3	4	5
Tingkat Kesukran Asesmen	0,49	0,58	0,65	0,65	0,64
Interpertasi	Sedang	Sedang	Sedang	Sedang	Sedang

Hasil pengujian tingkat kesukaran pada lima butir soal menunjukkan bahwa indeks kesukaran yang diperoleh secara berturut-turut adalah 0,49; 0,58; 0,65; 0,65; dan 0,64. Hasil klasifikasi berdasarkan indeks tingkat kesukaran menunjukkan bahwa semua butir soal tergolong dalam kategori sedang. Temuan tersebut mengindikasikan bahwa tingkat kesukaran pada setiap butir soal telah berada pada taraf yang seimbang, sehingga sesuai untuk digunakan dalam menilai kemampuan peserta didik tanpa didominasi oleh soal yang terlalu mudah ataupun terlalu sulit. Klasifikasi tingkat kesukaran pada kategori sedang mencerminkan bahwa instrumen telah dirancang dengan tingkat kesulitan yang seimbang, sehingga peserta didik memiliki peluang yang sama untuk mengungkapkan kemampuan berpikir kreatif matematisnya. Secara keseluruhan, setiap butir soal telah memenuhi standar tingkat kesukaran yang sesuai, sehingga instrumen tersebut dinilai layak dimanfaatkan sebagai alat penilaian untuk mengevaluasi kemampuan berpikir kreatif matematis peserta didik.

Apabila seluruh temuan analisis dipertimbangkan secara terpadu, terlihat bahwa instrumen memperoleh bukti keabsahan yang sangat kuat, tingkat keterandalan yang sangat tinggi dengan koefisien Cronbach's Alpha sebesar 0,984, daya pembeda yang sebagian besar berada pada kategori baik hingga sangat baik, serta tingkat kesukaran yang seimbang. Berdasarkan kondisi tersebut, dapat dinyatakan bahwa instrumen yang disusun telah memenuhi persyaratan sebagai alat ukur yang berkualitas sehingga layak diterapkan, baik untuk keperluan penelitian maupun sebagai perangkat penilaian dalam kegiatan pembelajaran.

Temuan yang diperoleh melalui kegiatan wawancara serta observasi menunjukkan bahwa guru memiliki kompetensi dan potensi yang memadai untuk menyusun instrumen tes berpikir kreatif matematis secara mandiri. Namun, pelaksanaan pengembangan instrumen tersebut belum dapat dilakukan secara optimal karena berbagai kendala yang dihadapi. Beban kerja guru yang cukup tinggi, seperti melaksanakan pembelajaran di kelas, menyusun perangkat pembelajaran melakukan penilaian hasil belajar, serta membimbing siswa, menyebabkan waktu yang tersedia untuk merancang instrumen berpikir kreatif menjadi sangat terbatas. Di sisi lain, guru merupakan pihak yang paling

memahami karakteristik, kemampuan, dan kebutuhan belajar siswa, sehingga seharusnya memiliki peran yang sangat penting dalam mengembangkan instrumen yang sesuai dengan kondisi peserta didik. Kondisi tersebut selaras dengan pandangan M. Ngalim Purwanto (2019) yang menjelaskan bahwa tugas guru mencakup perencanaan, pelaksanaan, dan evaluasi pembelajaran, tidak terbatas hanya pada kegiatan mengajar. Dengan demikian, guru perlu memiliki kemampuan dalam merancang alat evaluasi yang selaras dengan capaian pembelajaran serta memperhatikan karakteristik peserta didik.

Selain keterbatasan waktu, rendahnya minat belajar sebagian siswa juga menjadi salah satu faktor yang mengurangi motivasi guru untuk mengembangkan tes berpikir kreatif. Kondisi tersebut terjadi karena penyelesaian soal berpikir kreatif menuntut kemampuan berpikir tingkat tinggi, proses penalaran yang mendalam, serta alokasi waktu yang lebih panjang daripada penyelesaian soal yang bersifat rutin. Konsekuensinya, masih terdapat peserta didik yang belum mampu menyelesaikan tes dengan hasil yang optimal. Di samping itu, guru harus memastikan seluruh materi ajar tersampaikan dalam batas waktu yang telah ditentukan. Akibatnya, proses pembelajaran yang menekankan pengembangan kemampuan berpikir kreatif kerap belum memperoleh porsi yang optimal.

Atas dasar kondisi tersebut, perlu dikembangkan suatu alat evaluasi untuk menilai kemampuan berpikir kreatif dalam matematika yang bersifat praktis, memiliki tingkat keabsahan dan keterandalan yang baik, serta mudah diterapkan oleh guru. Ketersediaan instrumen tersebut diharapkan dapat mempermudah guru dalam melaksanakan penilaian terhadap kreativitas berpikir peserta didik secara lebih efektif tanpa memberikan tambahan beban kerja yang berarti. Selain itu, instrumen ini diharapkan mampu mendukung terwujudnya pembelajaran yang mendorong berkembangnya kreativitas matematis peserta didik sebagai salah satu kompetensi penting pada pembelajaran abad ke-21.

SIMPULAN

Penelitian ini menghasilkan suatu perangkat asesmen berbentuk soal open-ended untuk mengevaluasi kreativitas berpikir matematis peserta didik kelas XI SMK. Kelayakan perangkat tersebut ditetapkan melalui proses validasi dengan menggunakan pendekatan Classical Test Theory (CTT). Proses penyusunan

instrumen dimulai dengan analisis kebutuhan, kemudian dilanjutkan dengan perumusan indikator kreativitas berpikir dalam matematika yang mencakup aspek fluency, flexibility, originality, dan elaboration. Tahap berikutnya meliputi perancangan soal, penilaian kelayakan oleh para ahli, serta pelaksanaan uji coba yang melibatkan 32 siswa. Berdasarkan analisis kualitas instrumen, kelima soal memperoleh bukti keabsahan yang sangat kuat dengan koefisien berkisar antara 0,976 hingga 0,981. Selain itu, tingkat keterandalan instrumen juga sangat tinggi yang ditunjukkan oleh koefisien Cronbach's Alpha sebesar 0,984. Ditinjau dari aspek daya pembeda, sebagian besar soal berada pada klasifikasi baik hingga sangat baik, sedangkan hasil analisis tingkat kesukaran menempatkan seluruh soal pada kategori sedang. Temuan penelitian mengindikasikan bahwa kualitas psikometrik instrumen telah memenuhi standar yang ditetapkan. Dengan demikian, instrumen tersebut dapat dimanfaatkan sebagai perangkat penilaian untuk mengevaluasi kreativitas berpikir peserta didik dalam pembelajaran matematika.

Temuan dari tahap identifikasi kebutuhan mengindikasikan bahwa guru memiliki kompetensi untuk menyusun perangkat asesmen berpikir kreatif secara mandiri. Meskipun demikian, penerapannya masih menghadapi berbagai kendala, seperti tingginya beban kerja, keterbatasan waktu, tuntutan penyelesaian materi ajar, serta rendahnya minat sebagian peserta didik dalam mengerjakan soal yang memerlukan kemampuan berpikir tingkat tinggi. Oleh karena itu, keberadaan instrumen asesmen berpikir kreatif matematis berbasis soal open-ended yang telah dikembangkan diharapkan dapat membantu guru melaksanakan asesmen secara lebih efektif dan efisien tanpa menambah beban kerja secara signifikan. Selain berfungsi sebagai sarana evaluasi pembelajaran, instrumen ini juga diharapkan dapat menjadi salah satu pilihan asesmen yang mendorong berkembangnya kreativitas berpikir peserta didik dalam matematika sebagai kompetensi yang dibutuhkan pada pembelajaran abad ke-21.

DAFTAR PUSTAKA

Becker, J. P., & Shimada, S. (2019). *The Open-Ended Approach: A New Proposal for Teaching Mathematics*. National Council of Teachers of Mathematics.

- Borg, W. R., Gall, M. D., & Gall, J. P. (2003). *Educational Research: An Introduction* (7th ed.). Allyn & Bacon.
- Brookhart, S. M. (2024). *How to Design Questions and Tasks for Higher-Order Thinking* (2nd ed.). ASCD.
- Crocker, L., & Algina, J. (2020). *Introduction to Classical and Modern Test Theory*. Transatlantic Publications
- DeMars, C. E. (2018). Classical Test Theory and Item Response Theory. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley Handbook of Psychometric Testing*. Wiley. <https://doi.org/10.1002/9781118489772.ch2>
- DeVellis, R. F. (2022). *Scale Development: Theory and Applications* (5th ed.). SAGE Publications.
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale Development: Theory and Applications* (5th ed.). Sage Publications.
- Dewi, N. S., & Juandi, D. (2023). Pengaruh pendekatan open-ended terhadap kemampuan berpikir kreatif matematis: Systematic literature review. *Jurnal Pembelajaran Matematika Inovatif*, 6(3). <https://doi.org/10.22460/jpmi.v6i3.17338>
- Downing, S. M., & Haladyna, T. M. (Eds.). (2023). *Handbook on Test Development*. Routledge.
- Goretzko, D., Siemund, K., & Sterner, P. (2024). Evaluating model fit of measurement models in confirmatory factor analysis. *Educational and Psychological Measurement*, 84(1), 123–144. <https://doi.org/10.1177/00131644231163813>
- Kwon, M., Capraro, M. M., & Capraro, R. M. (2021). Deconstructing Open-Ended Problems: Standardized Test Analysis for Creative Thinking. *Creativity Studies*, 14(2), 345–361.
- M. Ngali Purwanto. (2019). *Prinsip-prinsip dan teknik evaluasi pengajaran*. PT Remaja Rosdakarya
- Penfield, R. D. (2021). *Item Analysis for Educators*. Measurement & Assessment Publishing.
- Sa'dijah, C., Rafiah, H., Gipayana, M., Qohar, A., & Anwar, L. (2016). Asesmen pemecahan masalah open-ended untuk mengukur profil berpikir kreatif matematis siswa berdasar gender. *Sekolah Dasar: Kajian Teori dan Praktik Pendidikan*, 25(2), 147–159. <https://doi.org/10.17977/um009v25i22016p147>
- Schoenfeld, A. H. (2022). *Mathematical Thinking and Problem Solving*. Routledge.
- Sipahi, Y., & Bahar, A. K. (2025). Mathematical creativity: A systematic review of definitions, frameworks, and assessment practices. *Education Sciences*, 15(10), 1348. <https://doi.org/10.3390/educsci15101348>
- Suherman, S., & Vidákovich, T. (2022). Assessment of mathematical creative thinking: A systematic review. *Thinking Skills and Creativity*, 44, 101019. <https://doi.org/10.1016/j.tsc.2022.101019>
- Sullivan, P., Bobis, J., Downton, A., Hughes, S., Livy, S., McCormick, M., & Russo, J. (2020). Threatening or Thriving? How Teachers Provide Challenging Open-Ended Tasks for All Students. *International Journal of Science and Mathematics Education*, 18(7), 1261–1279