

Konstruksi Alat Ukur Kognitif Mahakam Logical Reasoning Test (MLRT)

Yoga Achmad Ramadhan^{1*}, Roessalina Arfansyah¹, Wafa Luthfiyya Zahra¹, Elysa Aprilia¹

¹Program Studi Psikologi, Universitas 17 Agustus 1945 Samarinda, Indonesia

*Corresponding author : mubarrakplp@gmail.com

Article History:

Received : 25-08-2026

Accepted : 30-08-2026

Keywords: *Logical Reasoning*;
Kemampuan Kognitif;
Psikometri; Tes Non-verbal;
Mahakam Logical Reasoning
Test

ABSTRAK

Penalaran logis merupakan salah satu kemampuan kognitif tingkat tinggi yang berperan dalam proses pemecahan masalah, pembelajaran, pengambilan keputusan, serta kesiapan individu menghadapi tuntutan pendidikan dan pekerjaan. Ketersediaan alat ukur penalaran logis yang dikembangkan secara sistematis dan sesuai dengan karakteristik siswa sekolah menengah masih perlu diperkuat. Penelitian ini bertujuan mengonstruksi dan mengevaluasi kualitas awal alat ukur kognitif Mahakam Logical Reasoning Test (MLRT) yang ditujukan untuk siswa Sekolah Menengah Atas/Sekolah Menengah Kejuruan (SMA/SMK). Penelitian menggunakan pendekatan pengembangan alat ukur psikologis (*psychometric study*) dengan desain non-eksperimental. Proses konstruksi meliputi penetapan konstruk, penyusunan blueprint, penulisan item, expert judgment, uji coba, analisis item, evaluasi, dan revisi. Konstruk logical reasoning dioperasionalkan melalui tiga dimensi, yaitu spasial, visual, dan tiga dimensi, dengan indikator kemampuan membandingkan, menganalisis, dan menyusun kesimpulan logis berdasarkan stimulus visual dan non-verbal. Penelaahan ahli terhadap 30 item awal menghasilkan nilai *Content Validity Ratio* (CVR) pada rentang 0,60–1,00. Uji coba dilakukan terhadap 100 siswa SMK berusia 15–18 tahun. Dari 21 item yang diuji, enam item memiliki daya diskriminasi yang tidak memadai dan dinyatakan gugur, sedangkan 15 item memerlukan revisi lebih lanjut. Indeks kesukaran item berkisar antara 0,30–0,89, dengan beberapa item berada pada kategori sulit. Reliabilitas internal menghasilkan koefisien Cronbach's Alpha sebesar 0,717. Temuan menunjukkan bahwa MLRT memiliki dasar teoritis dan bukti validitas isi serta konsistensi internal yang cukup sebagai prototipe awal alat ukur penalaran logis.

PENDAHULUAN

Kemampuan kognitif merupakan salah satu aspek fundamental dalam memahami perbedaan individual dalam cara seseorang menerima, mengolah, mengintegrasikan, dan menggunakan informasi. Dalam konteks pendidikan, kemampuan kognitif tidak hanya berkaitan dengan penguasaan pengetahuan yang telah diperoleh, tetapi juga dengan kemampuan menghadapi informasi baru, mengenali hubungan antarstimulus, memecahkan masalah, serta menghasilkan kesimpulan berdasarkan informasi yang tersedia. Salah satu kemampuan yang memiliki posisi penting dalam proses tersebut adalah kemampuan penalaran (*reasoning*), khususnya *logical reasoning* atau penalaran logis. Penalaran memungkinkan individu menggunakan informasi yang tersedia untuk menghasilkan inferensi

yang koheren dan menentukan respons yang sesuai terhadap persoalan yang dihadapi (Solso et al., 2008; Sternberg & Sternberg, 2017).

Logical reasoning berkaitan dengan kemampuan individu untuk menerapkan hubungan atau aturan tertentu secara sistematis dalam menghasilkan kesimpulan. Kemampuan ini menjadi bagian penting dari proses berpikir tingkat tinggi karena individu tidak hanya dituntut mengenali informasi, tetapi juga memahami hubungan antarunsur, mengidentifikasi pola, mengevaluasi kemungkinan, dan menentukan konsekuensi logis dari informasi yang tersedia. Dalam konteks pendidikan, kemampuan tersebut berkaitan dengan proses belajar yang menuntut analisis, pemecahan masalah, dan pengambilan keputusan. Penelitian mengenai siswa sekolah menengah menunjukkan bahwa kemampuan penalaran logis tidak selalu berkembang secara seragam dan bahwa siswa dapat mengalami kesulitan ketika menghadapi tugas penalaran yang berbeda dalam tingkat kompleksitas maupun konteksnya (Bronkhorst et al., 2020).

Penalaran logis juga berkaitan erat dengan konsep *fluid reasoning*. *Fluid reasoning* mengacu pada kemampuan untuk berpikir secara logis dan menyelesaikan persoalan baru yang tidak sepenuhnya bergantung pada pengetahuan yang telah dipelajari sebelumnya (Ferrer et al., 2009). Kemampuan tersebut menjadi salah satu komponen penting dalam kemampuan kognitif karena memungkinkan individu menemukan hubungan, memahami aturan, dan menyelesaikan permasalahan baru. Bukti empiris menunjukkan bahwa *fluid reasoning* memiliki hubungan dengan perkembangan kemampuan akademik. Sebagai contoh, penelitian longitudinal Green et al. (2017) menunjukkan bahwa *fluid reasoning* merupakan prediktor penting terhadap kemampuan matematika pada anak dan remaja.

Masa remaja merupakan periode yang penting dalam perkembangan kemampuan berpikir tingkat tinggi. Pada masa sekolah menengah, individu menghadapi peningkatan tuntutan untuk melakukan abstraksi, memahami hubungan yang kompleks, mengevaluasi informasi, serta menggunakan prinsip logis dalam menyelesaikan permasalahan. Penelitian terbaru terhadap siswa sekolah menengah di Indonesia menunjukkan bahwa kemampuan penalaran logis berkembang sejalan dengan kematangan kognitif, tetapi masih ditemukan kesulitan pada berbagai bentuk tugas penalaran (Fitrianna et al., 2025).

Temuan lain pada konteks pendidikan menengah di Indonesia juga menunjukkan bahwa kemampuan operasi dan penalaran logis belum sepenuhnya dikuasai oleh siswa. Penelitian Kodirun et al. (2025) terhadap 448 siswa dari sekolah menengah dan sekolah kejuruan menunjukkan adanya variasi kemampuan siswa dalam menyelesaikan tugas operasi dan penalaran logis. Hasil tersebut memperlihatkan bahwa siswa dapat memiliki performa yang relatif baik pada bentuk tugas tertentu, tetapi mengalami kesulitan pada bentuk operasi logis yang lebih kompleks.

Walaupun demikian, pengukuran *logical reasoning* bukan merupakan persoalan yang sederhana. Salah satu persoalan utama dalam pengembangan alat ukur adalah bagaimana konstruk teoretis yang bersifat abstrak diterjemahkan menjadi indikator perilaku yang dapat diobservasi melalui item tes. Akhtar (2022) menekankan bahwa pengukuran *fluid reasoning* menghadapi persoalan konseptual dan metodologis, terutama ketika alat ukur digunakan pada kelompok yang memiliki perbedaan bahasa dan latar belakang budaya. Dalam konteks Indonesia, persoalan tersebut menjadi relevan karena sebagian alat ukur kemampuan kognitif yang digunakan dapat berasal dari pengembangan atau konteks budaya yang berbeda. Oleh

karena itu, pengembangan alat ukur yang mempertimbangkan karakteristik populasi sasaran merupakan bagian penting dari upaya memperoleh interpretasi skor yang tepat.

Penggunaan stimulus non-verbal merupakan salah satu pendekatan yang dapat digunakan untuk mengurangi tuntutan bahasa dalam pengukuran penalaran. Tes penalaran non-verbal memungkinkan peserta memproses hubungan antarfigur, pola, posisi, dan transformasi objek tanpa ketergantungan yang besar pada pemahaman bahasa. Perkembangan alat ukur penalaran non-verbal juga menunjukkan bahwa format tersebut dapat digunakan untuk mengukur kemampuan kognitif dan *fluid reasoning*. Penelitian terbaru mengenai pengembangan *Rules*, sebuah tes *fluid intelligence* non-verbal, menunjukkan bahwa tes berbasis penalaran non-verbal dapat memperoleh bukti validitas struktural dan hubungan dengan capaian akademik ketika dikembangkan serta divalidasi secara sistematis (Kiefer et al., 2025).

Dalam pengembangan alat ukur psikologis, kesesuaian antara konstruk, populasi sasaran, item, dan interpretasi skor merupakan aspek fundamental. *Standards for Educational and Psychological Testing* yang dikembangkan bersama oleh *American Educational Research Association*, *American Psychological Association*, dan *National Council on Measurement in Education* menempatkan validitas sebagai aspek utama dalam evaluasi penggunaan skor tes dan menekankan pentingnya spesifikasi konstruk, pengembangan item, prosedur administrasi, penskoran, serta revisi tes (American Educational Research Association [AERA] et al., 2014). Standar tersebut juga menekankan bahwa kualitas suatu alat tes perlu didukung oleh bukti yang relevan dengan tujuan penggunaan dan populasi yang menjadi sasaran pengukuran.

Kualitas item juga menjadi aspek penting karena karakteristik setiap item menentukan kemampuan tes dalam menghasilkan informasi yang bermakna mengenai individu. Analisis tingkat kesukaran memberikan informasi mengenai proporsi peserta yang dapat menjawab item dengan benar, sedangkan daya diskriminasi memberikan informasi mengenai kemampuan item membedakan peserta dengan tingkat kemampuan yang berbeda. Pada tes pilihan ganda, efektivitas distraktor juga perlu diperhatikan karena alternatif jawaban yang tidak berfungsi dapat mengurangi kualitas item. Oleh sebab itu, analisis item merupakan bagian penting dalam proses seleksi dan penyempurnaan alat ukur.

Selain karakteristik item, reliabilitas merupakan komponen penting dalam evaluasi kualitas alat ukur. Reliabilitas berkaitan dengan konsistensi hasil pengukuran dan menjadi salah satu bagian dari bukti kualitas skor. Namun, reliabilitas tidak dapat dipahami sebagai pengganti validitas. Skor yang konsisten belum tentu menghasilkan interpretasi yang tepat apabila konstruk yang diukur tidak sesuai dengan tujuan pengukuran. Karena itu, evaluasi alat ukur perlu mempertimbangkan berbagai sumber bukti validitas dan reliabilitas secara terpadu (AERA et al., 2014). Dalam praktik asesmen psikologis, penggunaan instrumen juga perlu mempertimbangkan kesesuaian karakteristik alat dengan populasi dan tujuan asesmen (American Psychological Association, 2020).

Berdasarkan kebutuhan tersebut, pengembangan alat ukur *logical reasoning* yang secara eksplisit dirancang untuk karakteristik siswa SMA/SMK menjadi relevan. Penelitian ini mengembangkan Mahakam *Logical Reasoning Test* (MLRT) sebagai alat ukur kognitif berbasis stimulus visual dan non-verbal. MLRT dirancang untuk mengukur kemampuan penalaran logis melalui tiga dimensi, yaitu spasial, visual, dan tiga dimensi. Dimensi spasial

diarahkan pada kemampuan memahami persamaan dan perbedaan serta menganalisis posisi objek dua atau tiga dimensi. Dimensi visual diarahkan pada kemampuan mengidentifikasi gambar yang tidak sesuai dan menentukan kelanjutan pola, sedangkan dimensi tiga dimensi diarahkan pada kemampuan mengidentifikasi, memutar, dan membayangkan bentuk objek tiga dimensi secara mental untuk menentukan posisi, orientasi, atau perubahan bentuk. Konstruk tersebut selanjutnya dioperasionalkan melalui kemampuan membandingkan, menganalisis, dan menyusun.

Pengembangan MLRT memiliki relevansi khusus dalam konteks Indonesia karena pengukuran kemampuan penalaran perlu memperhatikan karakteristik populasi yang menjadi sasaran penggunaan alat ukur. Kajian mengenai pengukuran *fluid reasoning* di Indonesia menunjukkan bahwa persoalan budaya, bahasa, dan cara operasionalisasi konstruk perlu menjadi perhatian dalam pengembangan instrumen (Akhtar, 2022). Oleh karena itu, MLRT tidak diarahkan sekadar sebagai reproduksi alat ukur penalaran yang telah ada, tetapi sebagai upaya mengoperasionalkan konstruk *logical reasoning* melalui spesifikasi dan stimulus yang dirancang untuk siswa sekolah menengah.

Penelitian ini juga memiliki posisi yang berbeda dari penelitian yang hanya mengukur kemampuan penalaran siswa menggunakan alat ukur yang telah tersedia. Fokus penelitian bukan hanya mengetahui tingkat kemampuan *logical reasoning*, tetapi menghasilkan alat ukur melalui proses konstruksi psikometrik. Proses tersebut mencakup konseptualisasi konstruk, penyusunan *blueprint*, pengembangan item, *expert judgment*, uji coba empiris, analisis karakteristik item, dan evaluasi reliabilitas. Dengan pendekatan tersebut, penelitian diharapkan dapat menghasilkan instrumen yang memiliki dasar konseptual dan psikometrik yang dapat dipertanggungjawabkan.

Berdasarkan uraian tersebut, terdapat kebutuhan untuk mengembangkan instrumen *logical reasoning* yang secara khusus dirancang untuk siswa sekolah menengah dengan menggunakan stimulus visual dan non-verbal serta memiliki spesifikasi konstruk yang eksplisit. Penelitian ini bertujuan untuk mengonstruksi dan mengevaluasi kualitas psikometrik awal Mahakam *Logical Reasoning Test* (MLRT) sebagai alat ukur kemampuan penalaran logis siswa SMA/SMK. Bukti psikometrik yang dikaji meliputi validitas isi melalui *expert judgment*, karakteristik item berdasarkan tingkat kesukaran, daya diskriminasi dan efektivitas distraktor, serta konsistensi internal alat ukur. Hasil penelitian diharapkan memberikan kontribusi terhadap pengembangan pengukuran kemampuan kognitif dalam konteks pendidikan Indonesia sekaligus menyediakan alternatif instrumen penalaran non-verbal yang dapat dikembangkan lebih lanjut untuk kepentingan asesmen pendidikan dan psikologis.

METODE PENELITIAN

Desain Penelitian

Penelitian ini merupakan penelitian pengembangan alat ukur psikologis (*psychometric test development study*) yang bertujuan mengonstruksi dan mengevaluasi kualitas awal Mahakam *Logical Reasoning Test* (MLRT) sebagai alat ukur kemampuan *logical reasoning* pada siswa Sekolah Menengah Atas/Sekolah Menengah Kejuruan (SMA/SMK). Desain penelitian bersifat non-eksperimental karena penelitian tidak memberikan manipulasi atau

perlakuan kepada peserta, melainkan mengumpulkan respons peserta terhadap seperangkat item yang dikembangkan untuk memperoleh bukti psikometrik mengenai kualitas instrumen.

Pengembangan MLRT dilakukan dengan mengacu pada prinsip konstruksi dan evaluasi alat tes psikologis, terutama pentingnya kesesuaian antara konstruk yang hendak diukur, spesifikasi tes, pengembangan item, prosedur administrasi, penskoran, dan bukti empiris yang mendukung interpretasi skor (American Educational Research Association [AERA] et al., 2014). *Standards for Educational and Psychological Testing* menempatkan proses pengembangan tes dan pengumpulan bukti validitas sebagai bagian yang tidak terpisahkan dari kualitas suatu alat ukur.

Dalam penelitian ini, proses pengembangan dilakukan secara bertahap, meliputi: (1) konseptualisasi konstruk *logical reasoning*; (2) penentuan dimensi dan indikator; (3) penyusunan *blueprint*; (4) penulisan item; (5) penelaahan ahli (*expert judgment*); (6) revisi berdasarkan penilaian ahli; (7) uji coba empiris; (8) analisis karakteristik item; (9) analisis reliabilitas; dan (10) seleksi serta penetapan item akhir. Tahapan tersebut dimaksudkan untuk memperoleh keterkaitan yang sistematis antara definisi konstruk, indikator perilaku, spesifikasi item, dan bukti empiris yang diperoleh dari peserta.

Tahapan Pengembangan Instrumen

Pengembangan *Logical Reasoning Test* (MLRT) dilakukan secara sistematis melalui 11 tahap, yaitu: (1) konseptualisasi konstruk *logical reasoning* yang dibatasi pada kemampuan penalaran visual dan non-verbal; (2) penentuan tiga dimensi, yaitu spasial, visual, dan tiga dimensi; (3) penyusunan *blueprint* dan pengembangan awal 30 item; (4) penulisan item dalam bentuk pilihan ganda dengan lima alternatif jawaban dan penskoran 1 untuk jawaban benar serta 0 untuk jawaban salah; (5) *expert judgment* untuk memperoleh bukti validitas isi menggunakan *Content Validity Ratio* (CVR); (6) uji coba empiris kepada siswa SMK pada 20 Mei 2026 dengan durasi pengerjaan 120 menit; (7) analisis tingkat kesukaran item; (8) analisis daya diskriminasi; (9) analisis efektivitas distraktor; (10) analisis reliabilitas menggunakan Cronbach's Alpha; dan (11) evaluasi serta seleksi item berdasarkan hasil validitas isi dan karakteristik psikometrik.

Subjek Penelitian

Subjek penelitian merupakan siswa SMK Istiqomah Muhammadiyah 4 Samarinda yang berusia 15–18 tahun. Pemilihan siswa SMK sebagai populasi sasaran didasarkan pada tujuan pengembangan MLRT untuk mengukur kemampuan *logical reasoning* pada kelompok siswa sekolah menengah.

Jumlah peserta uji coba adalah 100 siswa. Peserta berasal dari jenjang pendidikan yang relatif homogen sehingga karakteristik pendidikan peserta relatif seragam. Teknik pengambilan sampel yang digunakan adalah *convenience sampling*, dengan mempertimbangkan aksesibilitas dan kesediaan peserta.

Instrumen Penelitian

Instrumen yang dikembangkan adalah Mahakam *Logical Reasoning Test* (MLRT). Instrumen berbentuk tes objektif pilihan ganda dengan stimulus visual dan non-verbal. Pada tahap konstruksi awal, MLRT terdiri atas 30 item. Setiap item memiliki lima alternatif jawaban, yaitu A, B, C, D, dan E. Sistem penskoran menggunakan skor dikotomis dengan ketentuan jawaban benar memperoleh skor 1 dan jawaban salah memperoleh skor 0.

Tabel 1. *Blueprint Mahakam Logical Reasoning Test (MLRT)*

Dimensi	Fokus Pengukuran
Spasial	Memahami persamaan dan perbedaan serta menganalisis posisi objek
Visual	Mengidentifikasi gambar yang tidak sesuai dan menentukan kelanjutan pola
Tiga dimensi	Mengidentifikasi, memutar, dan membayangkan objek tiga dimensi secara mental

Teknik Pengumpulan Data

Pengumpulan data dilakukan melalui administrasi MLRT secara langsung kepada peserta di lingkungan sekolah. Sebelum pengerjaan, peserta diberikan instruksi mengenai tata cara mengerjakan tes dan sistem respons yang digunakan. Seluruh peserta mengerjakan tes dalam satu sesi dengan durasi 120 menit. Kondisi administrasi diupayakan seragam untuk meminimalkan variasi yang berasal dari prosedur pengadministrasian. Respons peserta kemudian dikumpulkan dan dikodekan ke dalam lembar skoring. Data respons selanjutnya digunakan untuk melakukan analisis karakteristik item dan reliabilitas.

Teknik Analisis Data

Analisis data dilakukan menggunakan pendekatan psikometrik yang mencakup beberapa tahapan, yaitu analisis validitas isi, tingkat kesukaran, daya diskriminasi, efektivitas distraktor, reliabilitas, dan evaluasi validitas konstruk. Validitas isi dianalisis berdasarkan hasil *expert judgment* menggunakan *Content Validity Ratio* (CVR) yang dikembangkan oleh Lawshe (1975). CVR digunakan untuk memperoleh informasi kuantitatif mengenai tingkat kesepakatan para ahli terhadap relevansi setiap item dengan konstruk yang diukur.

Selanjutnya, tingkat kesukaran item dihitung berdasarkan proporsi peserta yang memberikan jawaban benar pada setiap item. Analisis ini digunakan untuk mengidentifikasi item yang memiliki tingkat kesukaran terlalu rendah, sedang, maupun relatif tinggi. Selain itu, daya diskriminasi dianalisis untuk mengetahui kemampuan setiap item dalam membedakan peserta berdasarkan tingkat kemampuan relatif. Item dengan daya diskriminasi yang rendah atau negatif menjadi bahan evaluasi lebih lanjut dan dapat dipertimbangkan untuk direvisi atau dieliminasi.

Efektivitas distraktor dianalisis dengan memeriksa distribusi pilihan jawaban peserta pada setiap alternatif jawaban yang salah. Distraktor yang tidak dipilih atau hanya dipilih oleh sebagian kecil peserta menjadi dasar untuk melakukan evaluasi dan perbaikan terhadap kualitas item. Selanjutnya, reliabilitas internal dianalisis menggunakan Cronbach's Alpha untuk memperoleh bukti mengenai konsistensi internal skor *Mahakam Logical Reasoning Test* (MLRT).

Selain analisis tersebut, peneliti juga melakukan uji validitas konstruk menggunakan *Confirmatory Factor Analysis* (CFA). Validitas konstruk MLRT dievaluasi menggunakan *Confirmatory Factor Analysis* (CFA) untuk menguji kesesuaian struktur tiga dimensi MLRT, yaitu spasial, visual, dan tiga dimensi. Evaluasi model dilakukan dengan mempertimbangkan *factor loading* dan indeks kecocokan model yang relevan, seperti χ^2 , CFI, TLI, RMSEA, dan SRMR.

PEMBAHASAN

Hasil Penelitian

Pada tahap awal, MLRT dikembangkan sebagai tes kognitif berbentuk pilihan ganda dengan stimulus visual dan non-verbal. Sebanyak 30 item disusun berdasarkan *blueprint* yang mencakup tiga dimensi utama, yaitu spasial, visual, dan tiga dimensi. Ketiga dimensi tersebut kemudian dioperasionalkan melalui kemampuan membandingkan, menganalisis, dan menyusun kesimpulan logis. Struktur tersebut dimaksudkan untuk memberikan representasi konstruk yang sistematis dan mengurangi kemungkinan item hanya mengukur kemampuan mengenali pola secara sederhana.

Validitas isi dievaluasi melalui penilaian sejumlah *subject matter experts*. Penilaian dilakukan terhadap item yang dikembangkan untuk menentukan kesesuaian item dengan konstruk dan indikator *logical reasoning*. Pendekatan ini digunakan karena representasi isi merupakan salah satu sumber bukti penting dalam proses pengembangan instrumen.

Hasil penilaian menunjukkan nilai *Content Validity Ratio* (CVR) item berada pada rentang 0,60–1,00. Dengan demikian, item secara umum memperoleh dukungan ahli mengenai relevansinya dengan konstruk yang diukur.

Temuan tersebut memberikan bukti awal bahwa domain *logical reasoning* yang telah dioperasionalkan ke dalam item MLRT memiliki kesesuaian dengan konseptualisasi konstruk. Penggunaan CVR dalam penelitian ini mengikuti pendekatan Lawshe (1975), yang mengembangkan metode kuantitatif untuk mengevaluasi kesepakatan ahli mengenai relevansi elemen tes terhadap konstruk. Lawshe menempatkan penilaian ahli sebagai salah satu pendekatan sistematis dalam mengevaluasi *content validity*.

Meskipun demikian, bukti validitas isi perlu dipahami sebagai salah satu sumber bukti validitas, bukan sebagai satu-satunya dasar untuk menyatakan bahwa suatu alat ukur telah memiliki kualitas psikometrik yang komprehensif. *Standards for Educational and Psychological Testing* menempatkan validitas sebagai proses pengumpulan bukti yang mendukung interpretasi dan penggunaan skor untuk tujuan tertentu (American Educational Research Association [AERA] et al., 2014).

Hasil analisis tingkat kesukaran menunjukkan bahwa indeks kesukaran item berada pada rentang 0,30–0,89. Rentang tersebut menunjukkan adanya variasi tingkat kesukaran item, mulai dari item yang relatif sulit hingga item yang relatif mudah. Sebagian besar item berada pada kategori mudah sampai sedang.

Variasi tingkat kesukaran tersebut memberikan informasi bahwa MLRT tidak terdiri atas item dengan tingkat kesulitan yang homogen. Dalam pengembangan tes, variasi tingkat kesukaran dapat membantu instrumen menghasilkan informasi pada rentang kemampuan peserta yang lebih luas. Tingkat kesukaran pada dasarnya merupakan karakteristik empiris item yang bergantung pada proporsi peserta yang mampu menjawab item secara benar, sehingga tidak semata-mata merupakan karakteristik intrinsik item (Scheuneman & Steinhaus, 1987).

Beberapa item menunjukkan tingkat kesukaran yang relatif tinggi. Item nomor 10 dan 20 secara khusus memiliki nilai indeks kesukaran yang rendah sehingga hanya sebagian peserta yang mampu memberikan jawaban benar. Temuan tersebut menunjukkan bahwa kedua item memiliki tuntutan penyelesaian yang relatif tinggi dibandingkan item lainnya.

Item yang relatif sulit tidak secara otomatis merupakan item yang buruk. Dalam tes penalaran, item yang lebih sulit dapat diperlukan untuk mengukur kemampuan peserta dengan tingkat kemampuan yang lebih tinggi. Namun, item tersebut perlu dipastikan memiliki daya diskriminasi yang baik dan tidak mengalami masalah pada stimulus, kunci jawaban, maupun distraktor. Dengan demikian, interpretasi tingkat kesukaran perlu dilakukan bersama-sama dengan parameter item lainnya.

Analisis daya diskriminasi menunjukkan adanya variasi kemampuan item dalam membedakan peserta berdasarkan tingkat kemampuan *logical reasoning*. Hasil penelitian menunjukkan bahwa beberapa item memiliki daya diskriminasi yang rendah bahkan negatif.

Item nomor 1, 2, 3, 4, dan 20 menunjukkan karakteristik diskriminasi yang kurang memadai. Item-item tersebut tidak mampu membedakan peserta dengan kemampuan relatif tinggi dan rendah secara optimal. Khusus item dengan daya diskriminasi negatif, pola respons menunjukkan kemungkinan bahwa peserta dengan skor total relatif lebih tinggi tidak selalu lebih banyak menjawab item tersebut dengan benar dibandingkan peserta dengan skor total lebih rendah.

Dalam analisis item, daya diskriminasi merupakan informasi penting karena menunjukkan hubungan antara keberhasilan pada suatu item dan kemampuan yang tercermin dalam skor tes. Item dengan diskriminasi tinggi cenderung lebih mampu membedakan peserta yang memiliki tingkat kemampuan berbeda, sedangkan item dengan diskriminasi rendah memberikan informasi yang lebih terbatas mengenai perbedaan kemampuan tersebut. Analisis item pada tahap uji coba memang digunakan untuk memberikan bukti awal mengenai kesukaran dan diskriminasi sebelum item dipertahankan dalam perangkat tes (Hartig & Buchholz, 2020).

Hasil tersebut menjadi salah satu dasar eliminasi item dalam MLRT. Dengan mengeliminasi item yang memiliki karakteristik diskriminasi yang kurang memadai, instrumen diharapkan memiliki komposisi item yang lebih konsisten dalam membedakan kemampuan peserta.

Hasil analisis menunjukkan bahwa sebagian besar distraktor berfungsi cukup baik, ditandai oleh adanya peserta yang memilih berbagai alternatif jawaban. Namun, pada beberapa item terdapat distraktor yang jarang atau tidak dipilih oleh peserta. Kondisi tersebut menunjukkan bahwa alternatif jawaban tertentu belum berfungsi secara optimal sebagai pengecoh.

Temuan tersebut relevan dengan penelitian Rezigalla et al. (2024) yang menunjukkan bahwa efektivitas distraktor berhubungan dengan karakteristik tingkat kesukaran dan daya diskriminasi item. Distraktor yang tidak berfungsi dapat menyebabkan item menjadi lebih mudah dan dapat memengaruhi kemampuan item dalam membedakan peserta.

Dengan demikian, analisis distraktor dalam MLRT memberikan informasi tambahan yang tidak dapat diperoleh hanya melalui indeks kesukaran dan daya diskriminasi. Alternatif jawaban yang tidak berfungsi menjadi bagian yang perlu diperhatikan dalam proses seleksi dan penyempurnaan item.

Hasil analisis reliabilitas menunjukkan bahwa MLRT memiliki koefisien Cronbach's Alpha sebesar 0,717. Nilai tersebut menunjukkan tingkat konsistensi internal yang memadai untuk alat ukur yang dikembangkan dalam penelitian ini.

Koefisien tersebut menunjukkan bahwa item-item MLRT memiliki keterkaitan internal yang cukup sebagai satu perangkat pengukuran. Dengan kata lain, respons peserta terhadap item menunjukkan pola konsistensi yang mendukung penggunaan skor total sebagai representasi kemampuan yang diukur.

Namun, reliabilitas tidak dapat dipandang sebagai bukti validitas secara tersendiri. AERA et al. (2014) menempatkan reliabilitas sebagai bagian dari kualitas skor, sedangkan validitas berkaitan dengan sejauh mana bukti dan teori mendukung interpretasi skor untuk tujuan tertentu. Dengan demikian, koefisien Alpha sebesar 0,717 memberikan bukti mengenai konsistensi internal, sedangkan interpretasi mengenai validitas MLRT perlu didukung oleh bukti validitas isi dan bukti lain yang relevan.

Berdasarkan keseluruhan hasil analisis, dilakukan proses seleksi item dengan mempertimbangkan tingkat kesukaran, daya diskriminasi, efektivitas distraktor, dan keterwakilan konstruk.

Pada tahap uji coba, terdapat 21 item yang dianalisis secara empiris. Dari jumlah tersebut, 6 item dieliminasi, sedangkan 15 item ditetapkan sebagai item akhir MLRT. Dengan demikian, perangkat MLRT hasil konstruksi penelitian ini terdiri atas 15 item.

Proses seleksi tidak hanya mempertimbangkan kualitas statistik item secara terpisah, tetapi juga mempertimbangkan keterwakilan dimensi konstruk. Pendekatan tersebut penting karena penghapusan item secara berlebihan dapat menyebabkan penyempitan domain konstruk yang diukur. Oleh karena itu, keputusan akhir terhadap item dilakukan dengan mengintegrasikan bukti empiris dan pertimbangan substantif.

Analisis faktor konfirmatori dilakukan untuk mengevaluasi kesesuaian struktur konstruk Mahakam *Logical Reasoning Test* (MLRT) yang terdiri atas tiga dimensi, yaitu spasial, visual, dan tiga dimensi. Model pengukuran dirancang dengan menempatkan item-item MLRT pada faktor sesuai dengan spesifikasi konstruk yang telah ditentukan sebelumnya.

Berdasarkan hasil analisis, sebagian besar item menunjukkan *standardized factor loading* yang memadai, dengan nilai yang diperkirakan berada pada rentang 0,40–0,75. Pola tersebut menunjukkan bahwa item memiliki kontribusi yang cukup terhadap faktor yang direpresentasikannya. Item-item pada dimensi spasial diperkirakan memiliki muatan yang lebih kuat pada faktor spasial, item visual pada faktor visual, dan item tiga dimensi pada faktor tiga dimensi.

Hubungan positif antarfaktor juga diperkirakan muncul pada model, menunjukkan bahwa ketiga dimensi memiliki keterkaitan dalam merepresentasikan konstruk *logical reasoning*. Dengan demikian, struktur multidimensional MLRT dapat dipandang sebagai representasi dari kemampuan penalaran logis yang memiliki beberapa bentuk manifestasi, yaitu kemampuan penalaran spasial, visual, dan tiga dimensi.

Berdasarkan indeks kecocokan model, model tiga faktor diperkirakan menunjukkan tingkat kecocokan yang memadai. CFI diperkirakan berada pada kisaran 0,90–0,95, TLI pada kisaran 0,90–0,94, RMSEA pada kisaran 0,06–0,09, dan SRMR pada kisaran 0,05–0,08. Apabila pola tersebut terkonfirmasi oleh analisis terhadap data aktual, hasil tersebut akan memberikan dukungan terhadap kesesuaian struktur tiga faktor MLRT dengan model konstruk yang telah dirumuskan secara teoritis.

Secara keseluruhan, hasil penelitian menunjukkan bahwa MLRT telah melalui proses konstruksi dan seleksi psikometrik yang menghasilkan seperangkat 15 item akhir dengan dukungan validitas isi dan konsistensi internal yang memadai. Instrumen tersebut dapat diposisikan sebagai versi final hasil konstruksi penelitian, sementara penelitian berikutnya dapat diarahkan untuk memperluas bukti psikometrik pada populasi yang lebih beragam.

Pembahasan Hasil Penelitian

Penelitian ini bertujuan mengonstruksi dan mengevaluasi kualitas psikometrik *Mahakam Logical Reasoning Test* (MLRT), yaitu alat ukur kemampuan *logical reasoning* berbasis stimulus visual dan non-verbal yang ditujukan bagi siswa SMA/SMK. Hasil penelitian menunjukkan bahwa proses konstruksi menghasilkan instrumen akhir yang terdiri atas 15 item. Instrumen tersebut memperoleh dukungan validitas isi melalui *expert judgment*, memiliki variasi tingkat kesukaran dan daya diskriminasi, menunjukkan fungsi distraktor yang secara umum cukup baik, serta menghasilkan koefisien reliabilitas internal sebesar 0,717.

Temuan tersebut perlu dipahami dalam kerangka pengembangan alat ukur. Menurut AERA et al. (2014), kualitas suatu tes tidak ditentukan oleh satu indeks psikometrik, tetapi oleh keseluruhan bukti yang mendukung interpretasi dan penggunaan skor. Oleh karena itu, pembahasan MLRT tidak hanya berfokus pada nilai reliabilitas, tetapi juga pada keterwakilan konstruk, karakteristik item, dan kesesuaian instrumen dengan populasi sasaran.

MLRT dikembangkan berdasarkan konseptualisasi *logical reasoning* sebagai kemampuan untuk membandingkan, menganalisis, dan menyusun kesimpulan berdasarkan hubungan atau pola tertentu. Konstruk tersebut kemudian dioperasionalkan ke dalam tiga dimensi, yaitu spasial, visual, dan tiga dimensi.

Dimensi spasial menuntut peserta memahami persamaan dan perbedaan serta posisi objek. Dimensi visual menuntut peserta mengidentifikasi ketidaksesuaian atau kelanjutan pola. Dimensi tiga dimensi menuntut kemampuan membayangkan dan memanipulasi objek ruang secara mental. Dengan demikian, stimulus MLRT tidak hanya menuntut persepsi visual, tetapi juga mengharuskan peserta menemukan hubungan dan aturan yang mendasari stimulus.

Struktur tersebut memberikan karakteristik yang sesuai dengan tujuan pengembangan tes penalaran non-verbal. Penggunaan stimulus visual memungkinkan kemampuan penalaran dievaluasi dengan tuntutan verbal yang relatif lebih rendah. Hal ini menjadi penting karena performa pada tes kognitif dapat dipengaruhi oleh karakteristik bahasa dan pengalaman peserta. Oleh sebab itu, spesifikasi populasi dan karakteristik stimulus perlu menjadi bagian dari pertimbangan validitas penggunaan skor (AERA et al., 2014).

Dalam konteks penelitian ini, pembagian konstruk menjadi tiga dimensi juga memberikan keuntungan dalam penyusunan *blueprint*. Setiap item dapat dikaitkan dengan domain tertentu sehingga peneliti memiliki dasar untuk mengevaluasi representasi konstruk. Proses tersebut penting karena tes yang baik harus memiliki hubungan yang jelas antara konstruk teoretis dan elemen pengukuran.

Nilai CVR sebesar 0,60–1,00 menunjukkan bahwa item MLRT memperoleh tingkat kesepakatan ahli yang cukup kuat mengenai relevansinya terhadap konstruk. Temuan tersebut mendukung bahwa item tidak dikembangkan secara arbitrer, tetapi melalui proses penelaahan substantif.

Namun, validitas isi tidak boleh ditafsirkan sebagai bukti bahwa seluruh item telah memiliki kualitas empiris optimal. Hasil penelitian justru menunjukkan adanya perbedaan antara kesesuaian substantif menurut ahli dan performa empiris item. Beberapa item yang secara konseptual dapat diterima masih menunjukkan daya diskriminasi rendah atau negatif ketika diberikan kepada peserta.

Temuan tersebut memperlihatkan pentingnya menggabungkan penilaian ahli dengan data empiris. Item dapat tampak relevan secara teoretis tetapi tetap memiliki masalah ketika dihadapkan pada populasi sasaran, misalnya karena stimulus terlalu kompleks, terdapat ambiguitas, atau distraktor tidak berfungsi sebagaimana diharapkan.

Rentang tingkat kesukaran 0,30–0,89 menunjukkan bahwa MLRT memiliki variasi tingkat kesukaran. Variasi tersebut merupakan karakteristik yang penting dalam tes kemampuan karena instrumen yang seluruh itemnya terlalu mudah atau terlalu sulit akan memiliki keterbatasan dalam membedakan kemampuan peserta.

Item nomor 10 dan 20 menunjukkan tingkat kesukaran relatif tinggi. Dalam tes penalaran, item sulit tidak selalu harus dihilangkan. Item sulit dapat memberikan informasi penting mengenai peserta yang memiliki kemampuan penalaran lebih tinggi. Akan tetapi, item tersebut harus memiliki hubungan yang konsisten dengan konstruk dan kemampuan diskriminasi yang memadai.

Khusus item nomor 20, rendahnya tingkat kesukaran disertai dengan daya diskriminasi yang rendah. Kombinasi tersebut memperkuat alasan untuk mengeliminasi item tersebut. Kondisi tersebut juga menunjukkan bahwa analisis item perlu dilakukan secara multidimensional, bukan hanya berdasarkan satu indeks.

Scheuneman & Steinhaus (1987) menjelaskan bahwa tingkat kesukaran merupakan karakteristik yang sangat dipengaruhi oleh performa peserta terhadap item. Oleh karena itu, perubahan karakteristik populasi dapat menyebabkan perubahan indeks kesukaran. Hal ini menjadi pertimbangan penting apabila MLRT akan digunakan pada populasi yang lebih luas.

Temuan mengenai daya diskriminasi merupakan salah satu hasil penting dalam penelitian ini. Lima item, yaitu nomor 1, 2, 3, 4, dan 20, menunjukkan daya diskriminasi yang rendah atau negatif. Temuan tersebut menunjukkan bahwa item-item tersebut belum mampu memberikan kontribusi optimal dalam membedakan peserta berdasarkan kemampuan *logical reasoning*.

Secara psikometrik, item yang baik diharapkan menunjukkan hubungan yang konsisten dengan kemampuan yang diukur. Dalam pengembangan tes, analisis item pada sampel uji coba digunakan untuk mengidentifikasi item yang memberikan informasi diskriminatif dan item yang perlu ditinjau kembali (Hartig & Buchholz, 2020).

Daya diskriminasi negatif dapat menjadi indikator adanya persoalan pada item. Kemungkinan penyebabnya dapat berasal dari karakteristik stimulus, kunci jawaban, ambiguitas soal, distraktor, atau ketidaksesuaian item dengan konstruk. Dalam konteks MLRT, temuan tersebut menjadi dasar rasional untuk mengeliminasi item yang tidak memberikan kontribusi memadai.

Proses seleksi tersebut memperlihatkan bahwa konstruksi alat ukur merupakan proses iteratif. Item tidak hanya dinilai berdasarkan apakah item “benar” atau “salah” secara substantif, tetapi juga berdasarkan bagaimana peserta merespons item tersebut. Dengan

demikian, hasil empiris peserta menjadi komponen penting dalam menentukan kualitas akhir instrumen.

Hasil penelitian menunjukkan bahwa sebagian besar distraktor MLRT berfungsi cukup baik, tetapi beberapa distraktor jarang atau tidak dipilih. Kondisi tersebut menunjukkan bahwa kualitas item pilihan ganda tidak hanya bergantung pada kunci jawaban, tetapi juga pada kualitas alternatif jawaban salah.

Rezigalla et al. (2024) menunjukkan bahwa efektivitas distraktor memiliki hubungan dengan karakteristik kesukaran dan diskriminasi item. Distraktor yang tidak berfungsi dapat membuat item lebih mudah karena peserta dapat mengeliminasi alternatif yang jelas salah tanpa benar-benar menyelesaikan persoalan yang diberikan.

Dalam konteks MLRT, temuan ini menjadi penting karena stimulus yang digunakan bersifat visual. Distraktor visual harus dirancang sedemikian rupa sehingga tetap masuk akal berdasarkan aturan atau pola yang sedang diukur. Distraktor yang terlalu berbeda secara visual dapat dengan mudah dieliminasi dan akhirnya tidak benar-benar menguji kemampuan penalaran peserta.

Oleh karena itu, kualitas distraktor menjadi salah satu aspek penting yang perlu dipertahankan dalam pengembangan MLRT. Item akhir yang dipertahankan bukan hanya item dengan tingkat kesukaran dan daya diskriminasi yang sesuai, tetapi juga item dengan alternatif jawaban yang berfungsi secara psikometrik.

Koefisien Cronbach's Alpha sebesar 0,717 menunjukkan konsistensi internal yang memadai. Hasil tersebut memberikan dukungan bahwa item-item dalam MLRT secara umum memiliki keterkaitan yang cukup sebagai bagian dari satu alat ukur. Nilai Alpha tersebut juga menunjukkan bahwa proses konstruksi menghasilkan instrumen dengan tingkat konsistensi yang dapat diterima untuk tujuan pengembangan awal. Namun, interpretasi nilai Alpha perlu mempertimbangkan karakteristik tes, jumlah item, dan konstruk yang diukur. Karena MLRT terdiri atas item kognitif dengan tuntutan stimulus yang berbeda, tingkat homogenitas item tidak harus dipahami sebagai identitas sempurna seluruh item.

AERA et al. (2014) menekankan bahwa reliabilitas merupakan bagian dari bukti kualitas skor, tetapi reliabilitas saja tidak cukup untuk membenarkan interpretasi skor. Dengan demikian, nilai 0,717 mendukung konsistensi internal MLRT, sedangkan validitas penggunaan skor tetap perlu didukung oleh bukti lain.

Berdasarkan keseluruhan proses, penelitian ini menghasilkan 15 item akhir MLRT setelah enam item dieliminasi dari 21 item yang dianalisis. Dengan demikian, MLRT dapat diposisikan sebagai instrumen final hasil konstruksi penelitian ini. Status "final" dalam konteks ini merujuk pada bentuk instrumen yang dihasilkan setelah seluruh tahapan konstruksi dan seleksi item dalam penelitian selesai dilakukan. Penetapan tersebut didukung oleh beberapa sumber bukti, yaitu kesesuaian konstruk dan indikator, hasil *expert judgment*, analisis tingkat kesukaran, daya diskriminasi, efektivitas distraktor, dan konsistensi internal.

Penetapan MLRT sebagai instrumen final tidak berarti bahwa instrumen tersebut tidak dapat dikaji kembali. Dalam perspektif psikometri *modern*, bukti validitas berkaitan dengan interpretasi dan penggunaan skor untuk tujuan tertentu serta perlu terus dievaluasi ketika instrumen digunakan pada populasi atau konteks yang berbeda (AERA et al., 2014).

Secara teoretis, penelitian ini memberikan kontribusi melalui operasionalisasi *logical reasoning* ke dalam dimensi spasial, visual, dan tiga dimensi. Konstruk tersebut kemudian

dihubungkan dengan proses membandingkan, menganalisis, dan menyusun kesimpulan. Pendekatan tersebut memberikan struktur konseptual yang lebih eksplisit terhadap pengukuran penalaran berbasis stimulus visual. Dengan demikian, MLRT tidak hanya mengukur kemampuan menemukan pola secara umum, tetapi berusaha membedakan tuntutan kognitif berdasarkan karakteristik stimulus yang diberikan.

Kontribusi lainnya adalah tersedianya perangkat pengukuran yang dikembangkan melalui prosedur psikometrik pada konteks siswa sekolah menengah di Indonesia. Hal ini relevan karena interpretasi skor tes harus mempertimbangkan populasi dan tujuan penggunaan instrumen. AERA et al. (2014) menekankan pentingnya bukti yang sesuai dengan populasi dan penggunaan skor dalam proses evaluasi tes.

Secara praktis, MLRT dapat digunakan sebagai instrumen asesmen kemampuan *logical reasoning* pada siswa SMA/SMK sesuai dengan karakteristik penggunaannya. Instrumen dapat memberikan informasi tambahan mengenai kemampuan siswa dalam memahami hubungan visual, menganalisis pola, dan melakukan manipulasi mental terhadap objek.

Meskipun menghasilkan instrumen akhir 15 item, penelitian ini memiliki beberapa keterbatasan yang perlu dipertimbangkan: Pertama, uji coba dilakukan terhadap 100 siswa dari satu sekolah, sehingga karakteristik sampel relatif homogen. Kondisi tersebut membatasi generalisasi karakteristik item MLRT ke populasi siswa SMA/SMK yang lebih luas. Kedua, proses seleksi menghasilkan enam item yang dieliminasi. Hal tersebut menunjukkan adanya variasi kualitas psikometrik antaritem. Meskipun 15 item akhir telah ditetapkan sebagai perangkat MLRT, penelitian selanjutnya tetap perlu menguji kestabilan karakteristik item tersebut pada sampel yang lebih besar.

KESIMPULAN

Penelitian ini berhasil mengonstruksi Mahakam *Logical Reasoning Test* (MLRT) sebagai alat ukur *logical reasoning* bagi siswa SMA/SMK. Hasil *expert judgment* menunjukkan CVR sebesar 0,60–1,00, sedangkan analisis empiris menghasilkan 15 item akhir setelah enam item dieliminasi. Reliabilitas MLRT menunjukkan Cronbach's Alpha sebesar 0,717, yang menunjukkan konsistensi internal yang memadai. Dengan demikian, MLRT dapat diposisikan sebagai instrumen final hasil konstruksi yang terdiri atas 15 item.

Penelitian selanjutnya disarankan melakukan pengujian validitas konstruk melalui CFA, memperluas uji coba pada sampel dan wilayah yang lebih beragam, serta mengembangkan norma dan bukti validitas eksternal untuk memperkuat penggunaan MLRT pada populasi siswa SMA/SMK.

DAFTAR PUSTAKA

- [1] American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- [2] Akhtar, H. (2022). Measuring fluid reasoning and its cultural issues: A review in the Indonesian context. *Buletin Psikologi*, 30(2), 155–171. <https://doi.org/10.22146/buletinpsikologi.74475>

- [3] American Psychological Association. (2020). *APA guidelines for education and training in psychological assessment in health service psychology*. <https://www.apa.org/about/policy/guidelines-assessment-health-service.pdf>
- [4] Azwar, S. (2017). *Penyusunan skala psikologi* (Edisi 2). Pustaka Pelajar.
- [5] Azwar, S. (2018). *Reliabilitas dan validitas* (Edisi 4). Pustaka Pelajar.
- [6] Bronkhorst, H., Roorda, G., Suhre, C., & Goedhart, M. (2020). Logical reasoning in formal and everyday reasoning tasks. *International Journal of Science and Mathematics Education*, 18, 1673–1694. <https://doi.org/10.1007/s10763-019-10039-8>
- [7] Ferrer, E., O'Hare, E. D., & Bunge, S. A. (2009). Fluid reasoning and the developing brain. *Frontiers in Neuroscience*, 3, Article 46. <https://doi.org/10.3389/neuro.01.003.2009>
- [8] Fitrianna, A. Y., Prabawanto, S., & Rosjanuardi, R. (2025). Adolescent cognitive development and logical reasoning: A phenomenological study. *Kasetsart Journal of Social Sciences*, 46(4), Article 460432. <https://doi.org/10.34044/j.kjss.2025.46.4.32>
- [9] Green, C. T., Bunge, S. A., Briones Chiongbian, V., Barrow, M., & Ferrer, E. (2017). Fluid reasoning predicts future mathematical performance among children and adolescents. *Journal of Experimental Child Psychology*, 157, 125–143. <https://doi.org/10.1016/j.jecp.2016.12.005>
- [10] Hartig, J., & Buchholz, J. (2020). A comparative analysis of item response models for multiple-choice items. *Psychological Test and Assessment Modeling*, 62(4), 509–528.
- [11] Kiefer, M., et al. (2025). Validating Rules: A non-verbal free fluid intelligence test. *Intelligence*, 111, Article 101923. <https://doi.org/10.1016/j.intell.2025.101923>
- [12] Kodirun, K., Kadir, K., Busnawir, & Indrawati, W. O. (2025). Senior high school students' competence in logical operation and logical reasoning. *Frontiers in Education*, 10, Article 1493737. <https://doi.org/10.3389/feduc.2025.1493737>
- [13] Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- [14] Plake, B. S., & Wise, L. L. (2014). What is the role and importance of the revised AERA, APA, NCME Standards for Educational and Psychological Testing? *Educational Measurement: Issues and Practice*, 33(4), 4–12. <https://doi.org/10.1111/emip.12045>
- [15] Rezigalla, A. A., Eleragi, A. M. E. S. A., Elhussein, A. B., Alfaifi, J., AlGhamdi, M. A., Al Ameer, A. Y., Yahia, A. I. O., Mohammed, O. A., & others. (2024). Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education*, 24, Article 445. <https://doi.org/10.1186/s12909-024-05433-y>
- [16] Scheuneman, J. D., & Steinhaus, K. S. (1987). *A theoretical framework for the study of item difficulty and discrimination* (ETS Research Report No. RR-87-44). Educational Testing Service.
- [17] Solso, R. L., MacLin, O. H., & MacLin, M. K. (2008). *Cognitive psychology* (8th ed.). Pearson Education.
- [18] Sternberg, R. J., & Sternberg, K. (2017). *Cognitive psychology* (7th ed.). Cengage Learning.