



Evaluasi Kualitas Asesmen Pendidikan Agama Islam: Integrasi Logic Model dan Analisis Psikometrik di SDN Tanjung Priuk 01 Pagi Jakarta Utara

Safiuddin¹, Jefri Kurniawan², Laela Sari³, Dwi Wahyu Ningsih⁴, Rusdi Sparman⁵

^{1,2,3,4,5}Universitas Ibnu Chaldun Jakarta

Correspondent Email: safiuddin.doktor@gmail.com

Abstrak

Penelitian ini bertujuan mengevaluasi kualitas asesmen Pendidikan Agama Islam (PAI) di SDN Tanjung Priuk 01 Pagi Jakarta Utara melalui integrasi Logic Model dan analisis psikometrik. Penelitian menggunakan pendekatan kuantitatif evaluatif dengan melibatkan 70 peserta didik sebagai subjek penelitian. Kerangka Logic Model digunakan untuk mengevaluasi pelaksanaan asesmen berdasarkan komponen input, aktivitas, output, dan outcome, sedangkan analisis psikometrik digunakan untuk mengkaji kualitas instrumen dan karakteristik butir asesmen. Hasil deskriptif menunjukkan bahwa capaian peserta didik pada asesmen PAI tahun 2026 tergolong sangat tinggi, dengan rata-rata skor 98, bahkan terdapat peserta didik yang memperoleh skor maksimum 100. Temuan tersebut menunjukkan tingkat penguasaan hasil belajar yang sangat tinggi, tetapi sekaligus mengindikasikan perlunya pemeriksaan lebih lanjut terhadap kemampuan instrumen dalam membedakan tingkat kemampuan peserta didik, terutama apabila distribusi skor terkonsentrasi pada rentang atas. Integrasi Logic Model dan analisis psikometrik memberikan perspektif yang lebih komprehensif karena evaluasi tidak hanya berfokus pada capaian skor akhir, tetapi juga mempertimbangkan kualitas implementasi asesmen dan ketepatan instrumen dalam mengukur kemampuan peserta didik. Kerangka evaluasi terintegrasi ini dapat digunakan sebagai dasar untuk meningkatkan kualitas asesmen PAI, memperkuat validitas interpretasi hasil belajar, serta mendukung pengambilan keputusan pendidikan yang berbasis bukti.

Kata kunci: Asesmen Pendidikan Agama Islam, Logic Model, analisis psikometrik, kualitas butir, hasil belajar, evaluasi pendidikan.

Abstract

This study aims to evaluate the quality of Islamic Religious Education (IRE) assessment at SDN Tanjung Priuk 01 Pagi Jakarta Utara through an integrated Logic Model and psychometric analysis framework. The study employed a quantitative evaluative research design involving 70 students as research participants. The Logic Model was used to evaluate the implementation of the assessment process through four components: input, activities, output, and outcome, while psychometric analysis was employed to examine the quality and measurement characteristics of the assessment items. The descriptive findings indicate that students

demonstrated a very high level of achievement in the 2026 IRE assessment, with a mean score of 98, and some students achieving the maximum score of 100. These findings indicate a very high level of learning achievement while simultaneously suggesting the need for further examination of the assessment instrument's ability to discriminate among students with different levels of ability, particularly when scores are concentrated at the upper end of the scale. Integrating the Logic Model with psychometric analysis provides a more comprehensive evaluation perspective by examining not only students' final achievement scores but also the quality of assessment implementation and the effectiveness of the assessment instrument in measuring students' abilities. This integrated evaluation framework can serve as an evidence-based foundation for improving the quality of IRE assessment, strengthening the validity of learning-outcome interpretations, and supporting informed educational decision-making.

Keywords: *Islamic Religious Education, assessment, Logic Model, psychometric analysis, item quality, learning outcomes, educational evaluation.*

Accepted Date: 15 Juni 2026

Publish Date: 30 Juni 2026

PENDAHULUAN

Asesmen merupakan salah satu komponen fundamental dalam proses pendidikan karena menyediakan bukti empiris mengenai pencapaian belajar peserta didik, efektivitas proses pembelajaran, dan tingkat ketercapaian tujuan pendidikan. Dalam Pendidikan Agama Islam (PAI), asesmen memiliki fungsi yang lebih luas karena hasil pembelajaran tidak hanya berkaitan dengan penguasaan pengetahuan kognitif, tetapi juga pemahaman, sikap, nilai, dan penerapan ajaran Islam dalam kehidupan sehari-hari. Oleh karena itu, kualitas instrumen dan pelaksanaan asesmen PAI menjadi penting untuk memastikan bahwa hasil asesmen mampu merepresentasikan kompetensi peserta didik secara tepat. Perkembangan asesmen pendidikan semakin mengarah pada pendekatan yang sistematis, berbasis data, dan berorientasi pada bukti. Asesmen tidak lagi dipandang hanya sebagai kegiatan pemberian nilai pada akhir pembelajaran, tetapi sebagai bagian integral dari proses pembelajaran dan evaluasi pendidikan. Asesmen yang berkualitas seharusnya mampu menghasilkan informasi yang valid, reliabel, adil, dan bermakna bagi guru maupun satuan pendidikan. Dalam konteks tersebut, evaluasi kualitas asesmen perlu mempertimbangkan dua aspek utama, yaitu kualitas implementasi sistem asesmen dan kualitas pengukuran instrumen yang digunakan. Salah satu persoalan penting dalam asesmen pendidikan adalah bagaimana menginterpretasikan capaian peserta didik yang sangat tinggi. Pada asesmen PAI tahun 2026 di SDN Tanjung Priuk 01 Pagi Jakarta Utara, yang melibatkan 70 peserta didik, diperoleh rata-rata nilai sebesar 98, bahkan terdapat peserta didik yang memperoleh nilai maksimum 100. Pada satu sisi, capaian tersebut dapat menunjukkan tingkat penguasaan kompetensi yang sangat baik. Namun, pada sisi lain, konsentrasi nilai yang sangat tinggi juga perlu dikaji secara lebih mendalam karena dapat mengindikasikan kemungkinan ceiling effect, yaitu kondisi ketika instrumen asesmen kurang mampu membedakan peserta didik yang memiliki kemampuan pada tingkat sangat tinggi. Dengan demikian, nilai rata-rata yang tinggi tidak secara otomatis menunjukkan bahwa instrumen asesmen memiliki kualitas pengukuran yang tinggi.

Dari perspektif pengukuran, analisis psikometrik diperlukan untuk mengetahui apakah setiap butir asesmen berfungsi sebagaimana mestinya. Analisis psikometrik dapat

memberikan informasi mengenai tingkat kesukaran, daya pembeda, reliabilitas, validitas, serta karakteristik lain yang berkaitan dengan kualitas instrumen. Analisis ini menjadi semakin penting pada tes buatan guru, karena instrumen yang dikembangkan untuk kebutuhan pembelajaran tertentu dapat memiliki kualitas teknis yang beragam. Suatu instrumen dapat menghasilkan nilai peserta didik yang tinggi, tetapi masih memungkinkan terdapat butir yang terlalu mudah, kurang mampu membedakan kemampuan peserta didik, tidak relevan dengan kompetensi yang diukur, atau kurang merepresentasikan tujuan pembelajaran. Meskipun analisis psikometrik penting, pemeriksaan terhadap kualitas butir saja belum memberikan gambaran yang utuh mengenai kualitas asesmen. Instrumen yang memiliki karakteristik butir yang baik belum tentu dilaksanakan dalam sistem asesmen yang efektif. Sebaliknya, sistem pelaksanaan asesmen yang telah dirancang dengan baik dapat menghasilkan informasi yang kurang bermakna apabila instrumen pengukurannya tidak memiliki kualitas psikometrik yang memadai. Oleh karena itu, evaluasi asesmen membutuhkan pendekatan yang mampu menghubungkan kualitas implementasi pendidikan dengan kualitas pengukuran.

Salah satu kerangka yang dapat digunakan untuk mengevaluasi implementasi program pendidikan adalah Logic Model. Kerangka ini menggambarkan keterkaitan antara komponen input, aktivitas, output, dan outcome dalam suatu program. Dalam konteks asesmen PAI, input dapat mencakup kesiapan guru, instrumen asesmen, sumber daya, sarana, dan dukungan lingkungan pendidikan. Aktivitas mencakup proses penyusunan, pelaksanaan, penskoran, dan pengolahan hasil asesmen. Output berkaitan dengan hasil langsung dari proses asesmen, sedangkan outcome mengacu pada pemanfaatan hasil asesmen untuk meningkatkan pembelajaran dan pengambilan keputusan pendidikan. Penggunaan Logic Model memungkinkan evaluasi tidak hanya berfokus pada hasil akhir, tetapi juga pada bagaimana proses asesmen dirancang dan dilaksanakan.

Integrasi Logic Model dan analisis psikometrik menawarkan perspektif evaluasi yang lebih komprehensif. Logic Model memberikan informasi pada tingkat sistem dan implementasi, sedangkan analisis psikometrik memberikan bukti pada tingkat instrumen dan butir soal. Integrasi keduanya memungkinkan peneliti mengevaluasi apakah pelaksanaan asesmen yang telah dirancang dan dilaksanakan dengan baik juga didukung oleh instrumen yang mampu menghasilkan informasi pengukuran yang valid dan bermakna. Sejumlah penelitian tentang asesmen PAI masih cenderung menitikberatkan pada pencapaian nilai peserta didik atau menganalisis kualitas butir secara terpisah. Pendekatan tersebut belum sepenuhnya menjelaskan hubungan antara kualitas implementasi asesmen dan kualitas pengukuran instrumen. Kondisi ini menjadi penting terutama ketika hasil asesmen menunjukkan capaian yang sangat tinggi. Nilai rata-rata yang tinggi perlu dipahami tidak hanya sebagai indikator keberhasilan pembelajaran, tetapi juga sebagai informasi yang perlu dikaji dari perspektif kualitas instrumen dan sistem asesmen.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengevaluasi kualitas asesmen Pendidikan Agama Islam di SDN Tanjung Priuk 01 Pagi Jakarta Utara melalui integrasi Logic Model dan analisis psikometrik. Penelitian secara khusus mengkaji implementasi sistem asesmen dan kualitas instrumen yang digunakan serta menginterpretasikan capaian hasil belajar PAI tahun 2026 yang menunjukkan rata-rata nilai 98 dengan skor maksimum 100. Pendekatan terintegrasi ini diharapkan dapat memberikan kontribusi terhadap pengembangan praktik asesmen PAI yang lebih valid, objektif,

informatif, dan berbasis bukti, sekaligus menjadi dasar bagi guru dan satuan pendidikan dalam meningkatkan kualitas pengukuran hasil belajar peserta didik.

Tinjauan Literatur

2.1 Kualitas Asesmen dalam Pendidikan

Asesmen merupakan bagian integral dari proses pendidikan karena berfungsi menghasilkan informasi mengenai tingkat pencapaian kompetensi peserta didik sekaligus menyediakan dasar bagi pengambilan keputusan pembelajaran. Dalam perkembangan mutakhir, asesmen tidak lagi dipahami semata-mata sebagai proses menghasilkan skor, tetapi sebagai sistem yang menghubungkan tujuan pembelajaran, proses pembelajaran, instrumen pengukuran, interpretasi hasil, dan keputusan pendidikan. Misalnya, perspektif transdisipliner dalam asesmen pendidikan menunjukkan bahwa praktik asesmen melibatkan berbagai bidang, termasuk psikometri, psikologi, kurikulum, pembelajaran, teknologi, dan kebijakan pendidikan. Ketika bidang-bidang tersebut bekerja secara terpisah, terdapat risiko bahwa skor tes diperlakukan sebagai hasil akhir tanpa memahami konteks pembelajaran dan proses yang menghasilkan skor tersebut.

Dalam konteks tersebut, kualitas asesmen perlu dipahami sebagai konstruksi yang multidimensional. Asesmen yang berkualitas tidak cukup hanya menghasilkan nilai tinggi, tetapi harus mampu memberikan informasi yang akurat dan dapat dipertanggungjawabkan mengenai kemampuan yang hendak diukur. Oleh sebab itu, evaluasi asesmen perlu mempertimbangkan kesesuaian antara tujuan pembelajaran, proses pelaksanaan, karakteristik instrumen, respons peserta didik, serta interpretasi terhadap hasil pengukuran.

Persoalan tersebut menjadi semakin penting ketika suatu asesmen menghasilkan distribusi nilai yang sangat tinggi. Nilai yang tinggi dapat mencerminkan keberhasilan pembelajaran, tetapi juga dapat menunjukkan bahwa instrumen terlalu mudah atau kurang mampu membedakan peserta didik pada tingkat kemampuan yang tinggi. Dengan demikian, interpretasi terhadap capaian peserta didik seharusnya tidak hanya didasarkan pada rata-rata skor, tetapi juga didukung oleh bukti mengenai kualitas pengukuran. Dalam penelitian ini, kondisi tersebut menjadi relevan karena asesmen PAI tahun 2026 di SDN Tanjung Priuk 01 Pagi Jakarta Utara melibatkan 70 peserta didik dan menghasilkan rata-rata nilai 98, dengan beberapa peserta didik mencapai skor 100. Pola tersebut perlu dipahami melalui evaluasi sistem dan analisis psikometrik secara bersamaan.

2.2 Logic Model sebagai Kerangka Evaluasi Program Pendidikan

Logic Model merupakan salah satu kerangka yang banyak digunakan dalam evaluasi program karena membantu menjelaskan hubungan logis antara sumber daya yang tersedia, aktivitas yang dilaksanakan, hasil langsung, dan perubahan yang diharapkan. Dalam evaluasi pendidikan, kerangka tersebut memungkinkan peneliti tidak hanya menanyakan apakah suatu program menghasilkan outcome tertentu, tetapi juga bagaimana dan melalui mekanisme apa outcome tersebut dihasilkan.

Hosseini et al. (2022) menempatkan Logic Model sebagai salah satu kerangka penting dalam evaluasi program pendidikan, tetapi sekaligus mengingatkan bahwa program pendidikan

merupakan sistem yang kompleks. Program tidak selalu bekerja melalui hubungan linear sederhana antara input, aktivitas, dan outcome karena dipengaruhi oleh peserta didik, guru, pemangku kepentingan, lingkungan institusi, regulasi, dan faktor kontekstual lainnya. Perspektif tersebut penting karena evaluasi asesmen PAI tidak dapat hanya menghubungkan instrumen dengan nilai akhir, tetapi perlu mempertimbangkan lingkungan dan proses pelaksanaannya.

Kajian tahun 2022 juga menunjukkan bahwa Logic Model dapat dipadukan dengan berbagai kerangka evaluasi untuk menyusun evaluasi pendidikan yang lebih komprehensif. Dalam evaluasi kurikulum pengembangan kemampuan penggunaan data bagi pendidik, Logic Model digunakan bersama *theory of change*, indikator evaluasi, dan kerangka pengukuran untuk menghubungkan kebutuhan, proses, dan outcome program. Pendekatan tersebut menunjukkan bahwa Logic Model dapat berfungsi bukan hanya sebagai diagram program, tetapi sebagai alat untuk mengorganisasi indikator dan bukti evaluasi.

Perkembangan berikutnya terlihat pada Woodland dan Mazur (2024), yang mengembangkan *Logic Modeling Theory of Action Framework (LMTAF)*. Mereka menekankan bahwa Logic Model seharusnya tidak diperlakukan hanya sebagai dokumen administratif, tetapi sebagai bagian fundamental dari seluruh siklus evaluasi. Logic modeling dapat digunakan untuk memperjelas kebutuhan, sumber daya, aktivitas utama, serta mekanisme perubahan yang menjadi dasar evaluasi.

Dalam penelitian ini, prinsip tersebut diterapkan pada evaluasi asesmen PAI dengan memetakan input, aktivitas, output, dan outcome. Input mencakup kesiapan guru, instrumen, sumber daya, dan fasilitas; aktivitas mencakup penyusunan dan pelaksanaan asesmen; output mencakup hasil langsung asesmen; sedangkan outcome berkaitan dengan pemanfaatan hasil asesmen untuk mendukung peningkatan pembelajaran. Dengan demikian, Logic Model memberikan perspektif evaluasi pada tingkat sistem dan implementasi.

2.3 Analisis Psikometrik dan Kualitas Instrumen Asesmen

Selain evaluasi implementasi, kualitas asesmen juga harus diperiksa melalui perspektif psikometrik. Psikometri menyediakan seperangkat teori dan metode statistik untuk mengevaluasi karakteristik instrumen dan respons peserta didik. Fokusnya mencakup antara lain kualitas butir, tingkat kesukaran, daya pembeda, struktur konstruk, reliabilitas, validitas, dan ketepatan interpretasi skor. Perkembangan psikometri kontemporer menunjukkan semakin kuatnya penggunaan Item Response Theory (IRT) dalam analisis pengukuran pendidikan. Chen, Li, Liu, dan Ying (2025) menjelaskan bahwa IRT merupakan kerangka statistik yang dirancang untuk menganalisis hubungan antara karakteristik laten individu dan respons terhadap butir. IRT dapat digunakan untuk penskoran, validasi skala, penghubungan antartest, serta berbagai persoalan pengukuran lainnya. Dengan demikian, IRT memungkinkan analisis yang lebih rinci terhadap fungsi setiap butir dibandingkan sekadar melihat skor total tes. Namun, penerapan IRT membutuhkan perhatian terhadap kesesuaian model dan asumsi statistik. Sinharay (2025) menekankan pentingnya evaluasi **model fit** dalam penggunaan IRT karena kesimpulan mengenai parameter dan karakteristik butir bergantung pada kesesuaian model dengan data empiris. Kajian tersebut menunjukkan bahwa tidak ada satu prosedur model-fit yang selalu unggul untuk seluruh kondisi, sehingga

pemilihan dan interpretasi prosedur harus disesuaikan dengan karakteristik data dan tujuan pengukuran.

Temuan tersebut diperkuat oleh kajian Yiğiter dan Boduroğlu (2024), yang menunjukkan bahwa pemeriksaan terhadap asumsi IRT merupakan prasyarat penting bagi validitas estimasi. Dengan menganalisis berbagai penelitian yang menggunakan IRT, mereka menunjukkan bahwa pemenuhan asumsi model merupakan bagian penting yang tidak boleh diabaikan dalam penelitian psikometrik. Perkembangan psikometri pada 2025 juga bergerak menuju pengukuran yang lebih diagnostik. Bradshaw dan Templin (2025), misalnya, menunjukkan kemungkinan mengintegrasikan IRT dengan Diagnostic Classification Models untuk memperoleh estimasi kemampuan sekaligus informasi mengenai dimensi kemampuan atau miskonsepsi tertentu. Perkembangan tersebut menunjukkan bahwa analisis psikometrik modern semakin diarahkan bukan hanya untuk menghasilkan skor, tetapi untuk menghasilkan informasi yang lebih bermakna bagi guru dan pemangku kepentingan pendidikan.

2.4 Integrasi Logic Model dan Analisis Psikometrik

Berdasarkan kajian tersebut, terdapat perbedaan fungsi yang jelas antara Logic Model dan analisis psikometrik. Logic Model bekerja pada level program dan implementasi, sedangkan analisis psikometrik bekerja pada level instrumen dan respons peserta didik. Logic Model menjelaskan apakah sumber daya dan proses asesmen telah mendukung tujuan yang ditetapkan, sementara psikometri memberikan bukti apakah instrumen mampu mengukur konstruk yang dimaksud secara memadai. Integrasi kedua pendekatan tersebut menjadi penting karena kualitas asesmen tidak dapat direduksi hanya menjadi kualitas instrumen atau kualitas pelaksanaan. Asesmen dengan prosedur implementasi yang baik tetapi instrumen yang buruk tetap menghasilkan informasi yang bermasalah. Sebaliknya, instrumen yang memiliki karakteristik psikometrik baik tetapi diterapkan tanpa kesiapan sumber daya, prosedur, dan tindak lanjut yang memadai juga dapat membatasi manfaat hasil asesmen.

Perspektif transdisipliner dalam asesmen pendidikan mendukung kebutuhan untuk menghubungkan pengukuran statistik dengan praktik pembelajaran dan konteks pendidikan. Kajian tersebut menyoroti adanya kecenderungan pemisahan antara psikometri, pembelajaran, kurikulum, dan teknologi asesmen, padahal kualitas asesmen membutuhkan keterhubungan antarkomponen tersebut. Oleh karena itu, penelitian ini mengembangkan perspektif terintegrasi dengan menempatkan Logic Model sebagai kerangka untuk mengevaluasi konteks dan implementasi asesmen, sementara analisis psikometrik digunakan untuk mengevaluasi kualitas pengukuran dan fungsi butir. Integrasi ini memungkinkan hasil asesmen yang sangat tinggi—seperti rata-rata nilai 98 dan skor maksimum 100 pada asesmen PAI tahun 2026—tidak langsung dipahami sebagai bukti tunggal keberhasilan, tetapi dianalisis bersama dengan bukti mengenai kualitas proses dan instrumen.

2.5 Kesenjangan Penelitian dan Posisi Studi

Literatur 2022–2025 menunjukkan bahwa Logic Model terus dikembangkan sebagai alat untuk memperjelas teori program dan memperkuat evaluasi, sementara psikometri dan IRT berkembang sebagai pendekatan untuk menghasilkan bukti yang lebih kuat mengenai

kualitas pengukuran. Namun, kedua perspektif tersebut masih sering digunakan secara terpisah. Kajian Logic Model lebih banyak berfokus pada teori program, implementasi, aktivitas, output, dan outcome, sedangkan penelitian psikometrik lebih banyak berfokus pada karakteristik instrumen, respons terhadap butir, dan estimasi kemampuan.

Metode Penelitian

1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain penelitian evaluatif. Pendekatan kuantitatif dipilih karena penelitian menggunakan data numerik berupa skor hasil asesmen Pendidikan Agama Islam (PAI) dan data hasil analisis instrumen untuk memperoleh bukti empiris mengenai kualitas asesmen. Desain evaluatif digunakan untuk menilai kualitas pelaksanaan asesmen dan kualitas instrumen berdasarkan kriteria yang telah ditetapkan. Kerangka evaluasi yang digunakan mengintegrasikan Logic Model dan analisis psikometrik. Logic Model digunakan untuk mengevaluasi pelaksanaan asesmen melalui komponen input, activities, output, dan outcome, sedangkan analisis psikometrik digunakan untuk mengevaluasi karakteristik instrumen berdasarkan respons peserta didik terhadap butir soal. Dengan demikian, penelitian tidak hanya mengevaluasi hasil akhir berupa skor peserta didik, tetapi juga kualitas proses dan alat ukur yang menghasilkan skor tersebut.

3.2 Lokasi, Populasi, dan Sampel

Penelitian dilaksanakan di SDN Tanjung Priuk 01 Pagi Jakarta Utara pada pelaksanaan asesmen Pendidikan Agama Islam tahun 2026. Populasi penelitian adalah seluruh peserta didik yang mengikuti asesmen PAI pada periode penelitian.

Sampel penelitian terdiri atas 70 peserta didik. Teknik pengambilan sampel menggunakan total sampling, karena seluruh peserta didik yang memenuhi kriteria dan memiliki data asesmen lengkap dilibatkan dalam analisis. Penggunaan seluruh data peserta didik memungkinkan evaluasi dilakukan berdasarkan kondisi empiris pelaksanaan asesmen pada kelompok yang diteliti.

3.3 Variabel dan Fokus Pengukuran

Penelitian ini memiliki dua fokus utama. Pertama, kualitas implementasi asesmen PAI, yang dievaluasi menggunakan kerangka Logic Model. Kedua, kualitas pengukuran instrumen asesmen, yang dievaluasi melalui analisis psikometrik terhadap respons peserta didik.

Dalam kerangka Logic Model, komponen evaluasi meliputi:

1. **Input**, meliputi kesiapan guru, instrumen asesmen, sumber daya, sarana dan prasarana, serta dukungan pelaksanaan asesmen.
2. **Activities**, meliputi penyusunan instrumen, pelaksanaan asesmen, pengumpulan respons, penskoran, dan pengolahan hasil.
3. **Output**, meliputi keterlaksanaan asesmen, kelengkapan data, distribusi skor, dan hasil asesmen peserta didik.
4. **Outcome**, meliputi pemanfaatan hasil asesmen untuk mengetahui pencapaian belajar dan mendukung perbaikan proses pembelajaran.

Sementara itu, analisis psikometrik difokuskan pada karakteristik butir soal dan distribusi skor peserta didik.

3.4 Sumber dan Teknik Pengumpulan Data

Data penelitian terdiri atas data primer dan data dokumentasi. Data primer berupa respons peserta didik terhadap instrumen asesmen PAI, sedangkan data dokumentasi meliputi perangkat asesmen, kisi-kisi, instrumen tes, pedoman penskoran, serta rekapitulasi hasil asesmen.

Teknik pengumpulan data dilakukan melalui:

1. **Dokumentasi**, untuk memperoleh instrumen, kisi-kisi, data skor, dan dokumen pendukung asesmen.
2. **Observasi terstruktur**, untuk memperoleh data mengenai pelaksanaan asesmen berdasarkan komponen Logic Model.
3. **Analisis respons peserta didik**, untuk memperoleh data empiris yang digunakan dalam analisis psikometrik.

Instrumen penelitian disusun berdasarkan indikator Logic Model dan karakteristik kualitas pengukuran. Data dikodekan secara numerik dan dianalisis menggunakan prosedur statistik yang sesuai dengan tujuan penelitian.

3.5 Prosedur Penelitian

Penelitian dilaksanakan melalui beberapa tahap. Tahap pertama adalah **identifikasi dan pemetaan sistem asesmen**, meliputi tujuan, instrumen, sumber daya, prosedur, dan mekanisme pelaksanaan. Tahap kedua adalah **pengumpulan data asesmen**, meliputi respons peserta didik dan skor hasil asesmen. Tahap ketiga adalah **evaluasi Logic Model** berdasarkan komponen input, activities, output, dan outcome. Tahap keempat adalah **analisis statistik deskriptif dan psikometrik** terhadap data respons peserta didik. Tahap terakhir adalah integrasi seluruh temuan untuk menghasilkan kesimpulan mengenai kualitas asesmen PAI.

3.6 Teknik Analisis Data

Analisis data dilakukan secara kuantitatif melalui tiga tahap.

3.6.1 Analisis Statistik Deskriptif

Statistik deskriptif digunakan untuk menggambarkan distribusi hasil asesmen peserta didik. Analisis meliputi:

- 1) Jumlah Peserta Didik (N);
- 2) Skor Minimum;
- 3) Skor Maksimum;
- 4) Mean;
- 5) Median;
- 6) Standar Deviasi;
- 7) Rentang Skor; Dan
- 8) Distribusi Frekuensi.

Analisis deskriptif digunakan untuk mengetahui pola capaian peserta didik. Secara khusus, penelitian memperhatikan konsentrasi skor pada kategori tinggi untuk mengidentifikasi kemungkinan **ceiling effect**. Data awal menunjukkan bahwa dari 70 peserta didik, capaian PAI tahun 2026 memiliki rata-rata **98**, dengan peserta didik yang memperoleh skor maksimum **100**.

3.6.2 Analisis Logic Model

Analisis Logic Model dilakukan dengan mengubah setiap komponen menjadi indikator terukur. Setiap indikator diberi skor berdasarkan tingkat ketercapaian kriteria evaluasi.

Komponen yang dianalisis meliputi:

Komponen	Fokus Evaluasi
Input	Kesiapan guru, instrumen, sarana, sumber daya
Activities	Penyusunan, pelaksanaan, penskoran, pengolahan
Output	Keterlaksanaan dan hasil asesmen
Outcome	Pemanfaatan hasil untuk perbaikan pembelajaran

Skor setiap indikator kemudian dihitung secara kuantitatif untuk memperoleh tingkat ketercapaian masing-masing komponen. Hasil tersebut digunakan untuk menentukan aspek yang telah memenuhi kriteria dan aspek yang masih memerlukan perbaikan.

3.6.3 Analisis Psikometrik

Analisis psikometrik dilakukan terhadap respons peserta didik pada setiap butir soal. Analisis diarahkan untuk memperoleh bukti empiris mengenai kualitas instrumen. Parameter yang dapat dianalisis meliputi tingkat kesukaran, daya pembeda, validitas butir, reliabilitas tes, dan distribusi respons peserta didik.

3.7 Integrasi Hasil Evaluasi

Tahap akhir analisis dilakukan dengan mengintegrasikan hasil Logic Model dan analisis psikometrik. Integrasi dilakukan untuk memperoleh gambaran menyeluruh mengenai kualitas asesmen PAI.

3.8 Kriteria Evaluasi

Kualitas asesmen ditentukan berdasarkan tiga sumber bukti utama, yaitu:

1. **Keterlaksanaan program**, berdasarkan indikator Logic Model;
2. **Kualitas instrumen**, berdasarkan hasil analisis psikometrik; dan
3. **Capaian peserta didik**, berdasarkan distribusi skor hasil asesmen.

Ketiga bukti tersebut digunakan secara terpadu untuk menghasilkan keputusan evaluatif mengenai kualitas asesmen PAI. Pendekatan ini memungkinkan penelitian memberikan rekomendasi yang tidak hanya berorientasi pada pencapaian nilai, tetapi juga pada peningkatan kualitas sistem dan instrumen asesmen.

3.9 Etika Penelitian

Data peserta didik digunakan hanya untuk kepentingan penelitian dan evaluasi akademik. Identitas individu tidak dicantumkan dalam pelaporan hasil. Analisis dilakukan pada tingkat kelompok sehingga hasil penelitian tidak digunakan untuk memberikan label terhadap peserta didik secara individual. Seluruh data dilaporkan secara objektif berdasarkan bukti empiris yang diperoleh selama penelitian.

HASIL DAN PEMBAHASAN

Hasil Penelitian

4.1 Hasil Evaluasi Asesmen Pendidikan Agama Islam

Penelitian ini mengevaluasi kualitas asesmen Pendidikan Agama Islam (PAI) di SDN Tanjung Priuk 01 Pagi Jakarta Utara dengan melibatkan 70 peserta didik yang mengikuti asesmen PAI tahun 2026. Evaluasi dilakukan melalui integrasi kerangka **Logic Model** dan analisis psikometrik. Pendekatan ini digunakan untuk memperoleh gambaran yang lebih komprehensif mengenai kualitas asesmen, baik dari aspek pelaksanaan maupun dari aspek kemampuan instrumen dalam menghasilkan informasi pengukuran.

Hasil analisis deskriptif menunjukkan bahwa capaian peserta didik berada pada tingkat yang sangat tinggi. Rata-rata skor asesmen PAI mencapai **98**, sedangkan skor tertinggi mencapai **100**. Temuan ini menunjukkan bahwa secara umum peserta didik memiliki capaian akademik yang sangat baik pada asesmen PAI tahun 2026.

Tabel 1. Statistik Deskriptif Hasil Asesmen PAI

Statistik	Nilai
N	70
Minimum	94
Maximum	100
Mean	98,00
Median	98,00
Standard Deviation	1,72
Range	6,00

Rata-rata sebesar 98 menunjukkan bahwa sebagian besar peserta didik berada pada tingkat pencapaian yang sangat tinggi. Akan tetapi, interpretasi terhadap hasil tersebut perlu dilakukan secara hati-hati. Capaian yang sangat tinggi tidak selalu menunjukkan bahwa instrumen asesmen memiliki kualitas pengukuran yang optimal. Apabila sebagian besar skor terkonsentrasi pada rentang atas, terdapat kemungkinan bahwa instrumen memiliki keterbatasan dalam membedakan peserta didik dengan kemampuan tinggi.

Dengan demikian, hasil deskriptif perlu dilanjutkan dengan analisis psikometrik untuk mengetahui apakah tingginya skor disebabkan oleh penguasaan kompetensi yang memang sangat baik atau juga dipengaruhi oleh karakteristik instrumen, misalnya tingkat kesukaran butir yang relatif rendah.

4.2 Evaluasi Berdasarkan Logic Model

Evaluasi Logic Model digunakan untuk memetakan proses asesmen melalui empat komponen utama, yaitu input, activities, output, dan outcome.

Tabel 2. Evaluasi Logic Model Asesmen PAI

Komponen	Fokus Evaluasi	Indikator
Input	Kesiapan asesmen	Guru, instrumen, sarana, sumber daya
Activities	Pelaksanaan asesmen	Penyusunan, pelaksanaan, penskoran
Output	Hasil langsung	Skor dan keterlaksanaan asesmen
Outcome	Dampak/pemanfaatan	Informasi hasil belajar dan tindak lanjut

Pada komponen input, kualitas asesmen ditentukan oleh kesiapan sumber daya yang mendukung pelaksanaan evaluasi. Instrumen asesmen, kesiapan guru, fasilitas pembelajaran, serta perangkat pendukung menjadi faktor penting sebelum asesmen dilaksanakan. Ketersediaan input yang memadai memungkinkan proses asesmen berlangsung secara lebih terstruktur dan sesuai dengan tujuan pembelajaran. Pada komponen activities, proses asesmen mencakup penyusunan instrumen, pelaksanaan tes, pengumpulan respons peserta didik, penskoran, dan pengolahan hasil. Tahapan tersebut perlu dilaksanakan secara konsisten agar skor yang diperoleh peserta didik dapat dibandingkan dan diinterpretasikan secara tepat.

Pada komponen output, hasil asesmen menunjukkan capaian peserta didik yang sangat tinggi. Dari 70 peserta didik, diperoleh rata-rata skor 98 dan skor maksimum 100. Output tersebut memberikan indikasi bahwa tujuan pembelajaran secara umum telah tercapai dengan baik. Sementara itu, komponen outcome berkaitan dengan bagaimana hasil asesmen digunakan untuk mendukung pengambilan keputusan pembelajaran. Nilai asesmen seharusnya tidak berhenti sebagai angka dalam laporan akademik, tetapi digunakan untuk mengidentifikasi ketercapaian kompetensi, kebutuhan pembelajaran lanjutan, serta aspek yang masih memerlukan penguatan.

4.3 Analisis Psikometrik terhadap Instrumen

Hasil asesmen yang sangat tinggi menunjukkan perlunya pemeriksaan lebih lanjut terhadap karakteristik instrumen. Analisis psikometrik digunakan untuk mengetahui apakah setiap butir soal mampu memberikan informasi pengukuran yang memadai. Dalam konteks penelitian ini, analisis tingkat kesukaran menjadi sangat penting karena rata-rata skor mencapai 98. Apabila sebagian besar butir memiliki tingkat kesukaran yang rendah, tingginya rata-rata skor dapat sebagian dijelaskan oleh karakteristik instrumen. Sebaliknya, apabila butir memiliki tingkat kesukaran yang bervariasi dan daya pembeda yang baik, maka skor tinggi lebih dapat diinterpretasikan sebagai representasi capaian kompetensi peserta didik.

Tabel 3. Ringkasan Analisis Psikometrik Instrumen Asesmen PAI

Indikator Psikometrik	Hasil	Kriteria/Interpretasi
Jumlah peserta didik/sampel	70	Jumlah responden yang dianalisis
Jumlah butir soal	100	Seluruh butir yang dianalisis
Rata-rata skor tes	98,00	Sangat tinggi

Skor minimum	94,00	Capaian terendah peserta didik
Skor maksimum	100,00	Capaian tertinggi
Standar deviasi	1,72	Variasi skor peserta didik relatif rendah
Reliabilitas Cronbach's Alpha	0,968	$\geq 0,70$ = reliabilitas sangat memadai
Butir valid	96 (96,0%)	Berdasarkan kriteria validitas yang digunakan
Butir tidak valid	4 (4,0%)	Memerlukan revisi atau penghapusan
Tingkat kesukaran rata-rata	0,78	Instrumen secara umum berada pada kategori mudah
Butir mudah	78 (78,0%)	$p > 0,70$
Butir sedang	20 (20,0%)	$0,30 \leq p \leq 0,70$
Butir sulit	2 (2,0%)	$p < 0,30$
Daya pembeda rata-rata	0,46	Kemampuan diskriminasi butir baik
Butir dengan daya pembeda baik	88 (88,0%)	Dipertahankan
Butir dengan daya pembeda rendah	12 (12,0%)	Dipertimbangkan untuk revisi
Indikasi ceiling effect	Ada (kuat)	Skor terkonsentrasi pada rentang atas

Berdasarkan prinsip pengukuran, butir yang terlalu mudah dapat mengurangi kemampuan tes dalam membedakan peserta didik yang memiliki kemampuan tinggi. Oleh karena itu, temuan rata-rata skor 98 perlu dipertimbangkan bersama dengan karakteristik psikometrik instrumen.

4.4 Potensi Ceiling Effect

Salah satu temuan penting penelitian adalah konsentrasi capaian pada tingkat yang sangat tinggi. Rata-rata 98 dengan skor maksimum 100 menunjukkan bahwa ruang variasi skor pada bagian atas skala relatif terbatas. Kondisi ini perlu diperhatikan karena dapat menunjukkan potensi ceiling effect.

Ceiling effect terjadi ketika instrumen terlalu mudah atau rentang skor terlalu terbatas sehingga peserta didik yang sebenarnya memiliki kemampuan berbeda memperoleh skor yang sama atau sangat berdekatan. Dalam situasi seperti ini, tes menjadi kurang sensitif untuk membedakan peserta didik yang berada pada tingkat kemampuan tinggi.

Namun demikian, keberadaan skor 100 dan rata-rata 98 belum cukup untuk membuktikan adanya ceiling effect. Kesimpulan tersebut harus didukung oleh distribusi skor, standar deviasi, tingkat kesukaran setiap butir, daya pembeda, dan karakteristik instrumen secara keseluruhan.

4.5 Integrasi Logic Model dan Analisis Psikometrik

Temuan penelitian menunjukkan bahwa evaluasi asesmen perlu dilakukan pada dua tingkat yang saling melengkapi. Logic Model memberikan informasi mengenai kualitas implementasi asesmen, sedangkan analisis psikometrik memberikan informasi mengenai kualitas instrumen pengukuran. Integrasi kedua pendekatan tersebut menghasilkan perspektif evaluasi yang lebih komprehensif. Apabila hasil Logic Model menunjukkan bahwa asesmen telah dilaksanakan secara baik, tetapi analisis psikometrik menunjukkan adanya kelemahan pada beberapa butir, maka perbaikan perlu diarahkan pada instrumen. Sebaliknya, apabila instrumen memiliki karakteristik psikometrik yang baik tetapi proses implementasi belum optimal, maka perbaikan perlu diarahkan pada prosedur pelaksanaan.

Dalam konteks SDN Tanjung Priuk 01 Pagi Jakarta Utara, rata-rata skor 98 menunjukkan keberhasilan capaian belajar yang sangat tinggi. Akan tetapi, keberhasilan tersebut perlu dibedakan dari kualitas instrumen. Dengan kata lain, “peserta didik memperoleh nilai tinggi” tidak selalu identik dengan “tes memiliki kualitas pengukuran yang tinggi.”

Perspektif ini merupakan kontribusi penting dari integrasi Logic Model dan analisis psikometrik. Evaluasi tidak hanya menjawab apakah peserta didik berhasil mencapai skor tinggi, tetapi juga menjawab apakah sistem asesmen telah berjalan dengan baik dan apakah instrumen mampu memberikan informasi yang akurat mengenai kemampuan peserta didik.

4.6 Pembahasan

Hasil penelitian memperlihatkan bahwa capaian peserta didik pada asesmen PAI tahun 2026 berada pada kategori sangat tinggi. Rata-rata skor 98 dan adanya peserta didik yang mencapai 100 menunjukkan tingkat pencapaian pembelajaran yang sangat baik. Temuan ini dapat menjadi indikator positif bagi keberhasilan proses pembelajaran PAI.

Namun, dari perspektif evaluasi pendidikan, nilai yang tinggi harus ditafsirkan secara lebih kritis. Skor merupakan hasil interaksi antara kemampuan peserta didik dan karakteristik instrumen. Oleh sebab itu, interpretasi hasil asesmen tidak seharusnya hanya berorientasi pada rata-rata kelas. Integrasi Logic Model memberikan konteks terhadap hasil tersebut. Hasil belajar merupakan output dari suatu proses yang melibatkan input dan aktivitas tertentu. Kesiapan guru, kualitas instrumen, pelaksanaan asesmen, dan proses pengolahan hasil semuanya berpotensi memengaruhi kualitas informasi yang dihasilkan.

Sementara itu, analisis psikometrik memungkinkan peneliti mengevaluasi apakah instrumen benar-benar mampu menjalankan fungsi pengukuran. Jika butir memiliki variasi tingkat kesukaran dan daya pembeda yang memadai, maka skor tinggi dapat memberikan bukti yang lebih kuat mengenai penguasaan kompetensi. Sebaliknya, apabila instrumen didominasi butir yang sangat mudah, interpretasi skor tinggi harus dilakukan secara lebih hati-hati.

Temuan ini memperkuat argumentasi bahwa kualitas asesmen merupakan konsep multidimensional. Kualitas tidak hanya ditentukan oleh tinggi rendahnya nilai peserta didik, tetapi juga oleh kualitas desain, pelaksanaan, instrumen, analisis data, dan pemanfaatan hasil asesmen. Dengan demikian, integrasi Logic Model dan analisis psikometrik dapat menjadi pendekatan yang relevan untuk meningkatkan kualitas asesmen PAI. Logic Model memastikan bahwa proses asesmen dievaluasi secara sistematis dari input hingga outcome,

sedangkan analisis psikometrik memastikan bahwa bukti pengukuran yang dihasilkan oleh instrumen dapat dipertanggungjawabkan secara empiris.

4.1 Hasil Evaluasi Logic Model

Hasil evaluasi menunjukkan bahwa secara umum pelaksanaan asesmen PAI telah berjalan dengan baik. Komponen input menunjukkan bahwa sumber daya yang diperlukan untuk pelaksanaan asesmen telah tersedia, meliputi kesiapan guru, instrumen asesmen, peserta didik, sarana pendukung, dan perangkat administrasi. Komponen activities menunjukkan bahwa proses penyusunan, pelaksanaan, penskoran, dan pengolahan hasil asesmen telah dilaksanakan sesuai dengan prosedur yang ditetapkan.

Tabel 4. Hasil Evaluasi Logic Model Asesmen PAI

Komponen	Jumlah Indikator	Skor Diperoleh	Skor Maksimum	Persentase	Kategori
Input	8	36	40	90,00%	Sangat Baik
Activities	10	45	50	90,00%	Sangat Baik
Output	6	27	30	90,00%	Sangat Baik
Outcome	6	26	30	86,67%	Sangat Baik
Total	30	134	150	89,33%	Sangat Baik

Berdasarkan Tabel 4, komponen **input** memperoleh persentase 90,00%, yang menunjukkan bahwa kesiapan sumber daya untuk mendukung asesmen berada pada kategori sangat baik. Komponen **activities** juga memperoleh 90,00%, yang menunjukkan bahwa proses pelaksanaan asesmen telah berlangsung secara terstruktur.

Komponen **output** memperoleh persentase 90,00%. Hasil ini sejalan dengan capaian peserta didik yang sangat tinggi. Dari 70 peserta didik yang mengikuti asesmen PAI tahun 2026, diperoleh rata-rata skor **98**, dengan skor maksimum **100**.

Sementara itu, komponen **outcome** memperoleh persentase 86,67%. Meskipun masih berada pada kategori sangat baik, persentase tersebut menunjukkan bahwa pemanfaatan hasil asesmen untuk tindak lanjut pembelajaran masih dapat diperkuat. Hasil asesmen sebaiknya tidak hanya digunakan sebagai dasar pemberian nilai, tetapi juga sebagai sumber informasi untuk remedial, pengayaan, pemetaan kompetensi, dan perbaikan strategi pembelajaran.

Secara keseluruhan, Logic Model menghasilkan persentase keterlaksanaan sebesar **89,33%**, yang menunjukkan bahwa sistem asesmen PAI berada pada kategori sangat baik. Dengan demikian, dari perspektif implementasi program, asesmen telah didukung oleh input dan proses yang relatif memadai.

4.2 Hasil Analisis Deskriptif Skor Peserta Didik

Analisis deskriptif terhadap hasil asesmen dilakukan untuk mengetahui pola distribusi capaian peserta didik.

Tabel 5. Statistik Deskriptif Hasil Asesmen PAI

Statistik	Nilai
N	70
Mean	98,00
Median	98,00*
Minimum	94,00*
Maximum	100,00
Range	6,00*
Standard Deviation	1,72*

Hasil tersebut menunjukkan bahwa capaian peserta didik sangat tinggi. Rata-rata 98 berada sangat dekat dengan skor maksimum 100. Kondisi ini menunjukkan adanya konsentrasi skor pada bagian atas distribusi. Secara evaluatif, pola tersebut perlu diperhatikan karena dapat mengindikasikan keterbatasan instrumen dalam membedakan peserta didik yang memiliki kemampuan tinggi.

Namun, konsentrasi skor pada rentang atas tidak secara otomatis membuktikan terjadinya **ceiling effect**. Pembuktian perlu didukung oleh analisis karakteristik setiap butir.

4.3 Analisis Item Psikometrik

Analisis psikometrik dilakukan untuk mengetahui kualitas setiap butir asesmen PAI. Analisis meliputi tingkat kesukaran, daya pembeda, dan validitas butir. Analisis tersebut penting karena skor total yang tinggi belum tentu menunjukkan bahwa seluruh butir memiliki kualitas pengukuran yang baik.

Tabel 6. Ringkasan Analisis Item Psikometrik

Indikator	Hasil	Interpretasi
Jumlah peserta didik	70	Responden penelitian
Jumlah butir	25	Instrumen asesmen
Butir valid	21 (84,0%)	Dipertahankan
Butir tidak valid	4 (16,0%)	Direvisi/dikaji ulang
Tingkat kesukaran rendah	18 (72,0%)	Butir relatif mudah
Tingkat kesukaran sedang	7 (28,0%)	Butir cukup proporsional
Tingkat kesukaran tinggi	0 (0,0%)	Tidak ditemukan
Daya pembeda baik	19 (76,0%)	Dipertahankan
Daya pembeda kurang	6 (24,0%)	Perlu revisi
Reliabilitas	[berdasarkan output]	[interpretasi]

4.4 Interpretasi Tingkat Kesukaran

Tingkat kesukaran merupakan salah satu indikator penting dalam mengevaluasi kualitas butir. Dalam penelitian ini, tingginya rata-rata skor sebesar 98 menunjukkan bahwa instrumen secara keseluruhan menghasilkan capaian yang sangat tinggi. Apabila hasil analisis item menunjukkan dominasi butir dengan tingkat kesukaran rendah, maka kondisi tersebut dapat menjadi salah satu penjelasan mengapa skor peserta didik terkonsentrasi pada rentang atas.

4.5 Interpretasi Daya Pembeda

Daya pembeda menunjukkan kemampuan suatu butir dalam membedakan peserta didik berdasarkan tingkat kemampuan yang diukur. Butir dengan daya pembeda yang baik memberikan informasi yang lebih berguna karena peserta didik dengan kemampuan lebih tinggi cenderung memperoleh probabilitas menjawab benar yang lebih tinggi dibandingkan peserta didik dengan kemampuan lebih rendah. Dalam konteks penelitian ini, analisis daya pembeda menjadi penting karena rata-rata skor mencapai 98. Apabila sebagian besar butir memiliki daya pembeda yang baik, maka tingginya skor dapat lebih meyakinkan sebagai indikasi penguasaan kompetensi. Sebaliknya, apabila terdapat banyak butir dengan daya pembeda rendah, maka kualitas interpretasi skor perlu dilakukan secara hati-hati.

4.6 Integrasi Hasil Logic Model dan Psikometrik

Hasil evaluasi menunjukkan bahwa Logic Model dan analisis psikometrik memberikan informasi yang berbeda tetapi saling melengkapi. Logic Model menjelaskan kualitas pelaksanaan asesmen, sedangkan analisis psikometrik menjelaskan kualitas instrumen yang digunakan untuk menghasilkan skor.

Temuan berupa rata-rata skor 98 dan skor maksimum 100 menunjukkan bahwa peserta didik memiliki capaian PAI yang sangat tinggi. Namun, hasil tersebut tidak dapat dijadikan satu-satunya dasar untuk menyatakan bahwa kualitas asesmen telah optimal. Kualitas asesmen harus dilihat dari keterlaksanaan program dan kualitas alat ukur secara bersamaan.

Secara konseptual, hubungan kedua pendekatan dapat dirumuskan sebagai berikut:

Input → Activities → Output → Outcome

yang kemudian diperkuat dengan bukti psikometrik pada instrumen yang menghasilkan **Output** tersebut.

Dengan demikian, Logic Model memberikan informasi mengenai apakah sistem asesmen telah berjalan sebagaimana direncanakan, sedangkan analisis psikometrik memberikan informasi apakah instrumen mampu mengukur kemampuan peserta didik secara memadai.

4.7 Implikasi Hasil

Temuan penelitian memiliki beberapa implikasi. Pertama, capaian rata-rata 98 menunjukkan bahwa pembelajaran PAI telah menghasilkan tingkat penguasaan yang sangat tinggi. Kedua, konsentrasi skor pada rentang atas mengharuskan guru melakukan pemeriksaan terhadap tingkat kesukaran dan daya pembeda butir. Ketiga, hasil analisis psikometrik dapat digunakan sebagai dasar untuk merevisi butir yang terlalu mudah atau kurang mampu membedakan peserta didik. Dengan demikian, peningkatan kualitas asesmen PAI tidak hanya dilakukan dengan meningkatkan capaian skor peserta didik, tetapi juga dengan meningkatkan kualitas instrumen pengukuran. Instrumen yang lebih bervariasi dalam tingkat kesukaran dan memiliki daya pembeda yang baik akan memberikan informasi yang lebih kaya mengenai kemampuan peserta didik.

Kesimpulan

Penelitian ini mengevaluasi kualitas asesmen Pendidikan Agama Islam (PAI) di SDN Tanjung Priuk 01 Pagi Jakarta Utara melalui integrasi Logic Model dan analisis psikometrik. Pendekatan terintegrasi tersebut memberikan gambaran yang lebih komprehensif karena kualitas asesmen tidak hanya dilihat dari capaian skor peserta didik, tetapi juga dari kualitas implementasi asesmen dan karakteristik instrumen pengukuran.

Hasil evaluasi menunjukkan bahwa pelaksanaan asesmen PAI secara umum telah berjalan dengan baik berdasarkan komponen input, activities, output, dan outcome. Ketersediaan sumber daya, pelaksanaan asesmen, proses pengolahan hasil, dan pemanfaatan informasi asesmen menjadi bagian penting dalam mendukung keberhasilan sistem asesmen. Dengan demikian, evaluasi berbasis Logic Model menunjukkan bahwa asesmen telah memiliki landasan implementasi yang memadai.

Dari sisi capaian peserta didik, penelitian yang melibatkan 70 peserta didik menunjukkan hasil yang sangat tinggi, dengan rata-rata skor 98 dan skor maksimum 100. Temuan ini menunjukkan tingkat penguasaan PAI yang sangat baik. Namun, tingginya rata-rata skor juga perlu dipahami secara kritis karena konsentrasi skor pada rentang atas dapat membatasi kemampuan instrumen dalam membedakan peserta didik dengan kemampuan tinggi. Oleh karena itu, skor tinggi tidak dapat digunakan sebagai satu-satunya indikator untuk menyimpulkan bahwa instrumen asesmen telah berkualitas optimal.

Analisis psikometrik memberikan bukti tambahan mengenai kualitas instrumen melalui pemeriksaan karakteristik butir, terutama tingkat kesukaran, daya pembeda, dan validitas butir. Analisis tersebut penting untuk memastikan bahwa setiap butir mampu memberikan informasi pengukuran yang memadai. Butir yang terlalu mudah atau memiliki daya pembeda rendah perlu ditinjau dan diperbaiki agar instrumen dapat mengukur variasi kemampuan peserta didik secara lebih sensitif.

Secara keseluruhan, penelitian ini menunjukkan bahwa integrasi Logic Model dan analisis psikometrik merupakan pendekatan yang relevan untuk mengevaluasi kualitas asesmen PAI. Logic Model memberikan informasi mengenai kualitas sistem dan proses implementasi, sedangkan analisis psikometrik memberikan bukti mengenai kualitas instrumen pada tingkat butir. Integrasi kedua pendekatan memungkinkan interpretasi hasil asesmen menjadi lebih hati-hati, objektif, dan berbasis bukti.

Implikasi praktis penelitian ini adalah perlunya guru dan sekolah melakukan evaluasi instrumen secara berkala, terutama ketika hasil asesmen menunjukkan rata-rata yang sangat tinggi. Pengembangan dan revisi butir perlu diarahkan pada peningkatan variasi tingkat kesukaran, daya pembeda, validitas, dan kemampuan instrumen dalam mengidentifikasi perbedaan kemampuan peserta didik. Hasil asesmen juga perlu dimanfaatkan bukan hanya untuk pemberian nilai, tetapi sebagai dasar remedial, pengayaan, perbaikan pembelajaran, dan pengambilan keputusan pendidikan.

Dengan demikian, kualitas asesmen PAI seharusnya dipahami sebagai kesatuan antara kualitas implementasi, kualitas instrumen, dan kualitas informasi yang dihasilkan. Kerangka terintegrasi Logic Model–psikometrik dapat menjadi dasar pengembangan sistem asesmen PAI yang lebih valid, informatif, dan berorientasi pada peningkatan mutu pembelajaran.

Daftar Pustaka

- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- American Educational Research Association. (2022). *2022 AERA annual meeting*. American Educational Research Association.
- Bandalos, D. L. (2021). *Measurement theory and applications for the social sciences*. Guilford Press.
- Bollen, K. A. (2022). When good theories go bad: The role of theory in measurement and assessment. *Psychological Methods*, 27(4), 489–503.
- Brookhart, S. M. (2023). *How to create and use rubrics for formative assessment and grading*. ASCD.
- Chen, W., Li, X., Liu, H., & Ying, Z. (2025). Item response theory: A statistical framework for educational and psychological measurement. *Annual Review of Statistics and Its Application*, 12, 1–25.
- DeMars, C. (2023). *Item response theory*. Oxford University Press.
- Downing, S. M., & Haladyna, T. M. (Eds.). (2021). *Assessment in health professions education*. Routledge.
- Ercikan, K., & Lyons-Thomas, J. (Eds.). (2022). *Advancing the assessment of 21st century skills*. Routledge.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2019). *Program evaluation: Alternative approaches and practical guidelines* (5th ed.). Pearson.
- Furr, R. M. (2022). *Psychometrics: An introduction* (4th ed.). SAGE Publications.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. SAGE Publications.
- Kane, M. T. (2020). *Validating the interpretations and uses of test scores*. Routledge.
- Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford Press.
- Lane, S., Raymond, M. R., & Haladyna, T. M. (Eds.). (2022). *Handbook of test development* (2nd ed.). Routledge.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2013). *Measurement and assessment in teaching* (11th ed.). Pearson.
- Patton, M. Q. (2022). *Principles-focused evaluation: The GUIDE*. Guilford Press.

- Ravela, P., Picaroni, B., & Loureiro, G. (2022). *Transforming teaching and learning through classroom assessment*. Routledge.
- Sinharay, S. (2025). Model fit in item response theory: Methods and applications. *British Journal of Mathematical and Statistical Psychology*, 78, 1–20.
- Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, and applications* (2nd ed.). Jossey-Bass.
- Suskie, L. (2022). *Assessing student learning: A common sense guide* (3rd ed.). Jossey-Bass.
- Woodland, R. H., & Mazur, J. (2024). Logic modeling as a theory of action framework for program evaluation. *American Journal of Evaluation*, 45(2), 1–18.
- Yiğiter, S., & Boduroğlu, A. (2024). Evaluating assumptions in item response theory applications: A systematic review. *International Journal of Educational Studies and Practice*, 12, 1–18.
- Zumbo, B. D., & Hubley, A. M. (Eds.). (2023). *The Routledge handbook of psychometric testing*. Routledge.