

**TREN TRANSFORMASI ASESMEN SPIRITUALITAS DALAM PENDIDIKAN
ISLAM BERBASIS DIGITAL PORTFOLIO:
A SYSTEMATIC LITERATURE REVIEW**

Fadhila Nur Azizah¹, Prisca Anggun Ferian Putri², Chusnul Chotimah³
Universitas Islam Negeri Sayyid Ali Rahmatullah Tulungagung^{1,2}
fadhilanurazizah4@gmail.com¹, prisca.anggun12@gmail.com²,
chusnul.chotimah@uinsatu.ac.id³

ABSTRACT

The transformation of spirituality assessment in Islamic education toward a digital portfolio approach has become a central issue in affective evaluation still faces various methodological challenges, such as social desirability bias and the lack of data triangulation. This study aims to analyze trends in the use of affective assessment instruments and identify gaps in the transformation toward more integrative and digital approaches. The method used is a Systematic Literature Review (SLR) following the Kitchenham & Charters guidelines. Article searches were conducted through Google Scholar, GARUDA, SINTA, and DOAJ databases for the period 2015–2025. From 200 initial articles, 25 articles met the inclusion criteria and were analyzed. The results show that (1) Likert scales in self-report instruments remain dominant, (2) the use of direct observation and peer assessment remains very limited, (3) no affective evaluation model integrating all three approaches has been identified, and (4) the transformation toward digital assessment is still in its early stages. This study recommends the development of an integrated affective assessment model based on triangulation and digitalization as a direction for future transformation.

Keywords: spirituality assessment, Islamic education, digital portfolio, affective evaluation, triangulation

ABSTRAK

Transformasi asesmen spiritualitas dalam pendidikan Islam menuju pendekatan digital portfolio menjadi isu sentral dalam evaluasi afektif pendidikan masih dihadapkan pada berbagai tantangan metodologis, seperti *social desirability bias* dan kurangnya triangulasi data. Penelitian ini bertujuan untuk menganalisis tren penggunaan instrumen asesmen afektif serta mengidentifikasi celah transformasi menuju pendekatan yang lebih integratif dan digital. Metode yang digunakan adalah *Systematic Literature Review* (SLR) dengan mengacu pada pedoman Kitchenham & Charters. Pencarian artikel dilakukan melalui Google Scholar, GARUDA, SINTA, dan DOAJ pada rentang 2015–2025. Dari 200 artikel awal, 25 artikel memenuhi kriteria inklusi dan dianalisis. Hasil menunjukkan bahwa (1) skala Likert pada instrumen *self-report* masih mendominasi, (2) observasi langsung dan *peer assessment* sangat terbatas penggunaannya, (3) belum ditemukan model evaluasi afektif yang mengintegrasikan ketiga pendekatan tersebut, dan (4) transformasi menuju asesmen digital masih berada pada tahap

awal. Penelitian ini merekomendasikan pengembangan model asesmen afektif terintegrasi berbasis triangulasi dan digitalisasi sebagai arah transformasi ke depan.

Kata kunci: asesmen spiritualitas, pendidikan Islam, portofolio digital, evaluasi afektif, triangulasi

A. Pendahuluan

Pendidikan tidak hanya bertujuan mengembangkan kemampuan kognitif peserta didik, tetapi juga membentuk sikap dan keterampilan sebagai satu kesatuan proses pembelajaran. Dalam perspektif pendidikan Islam, keberhasilan pendidikan tidak cukup diukur dari penguasaan pengetahuan agama, tetapi juga dari terinternalisasinya nilai-nilai religius dalam perilaku sehari-hari. Oleh karena itu, sikap religius menjadi salah satu aspek penting dalam ranah afektif yang perlu dievaluasi secara sistematis karena mencerminkan keselarasan antara keyakinan, sikap, dan tindakan peserta didik (Krathwohl et al., 1964; Lickona, 1991).

Pendidikan karakter juga menjadi fokus utama dalam sistem pendidikan untuk membentuk generasi yang berintegritas, bertanggung jawab, dan berakhlak mulia. Meskipun implementasi pendidikan karakter telah diperkuat

melalui Kurikulum 2013, praktik penilaian di sekolah masih lebih menitikberatkan pada aspek kognitif dibandingkan aspek afektif. Akibatnya, perkembangan karakter dan sikap religius peserta didik belum dapat diukur secara optimal sehingga diperlukan sistem asesmen yang mampu menggambarkan perubahan perilaku secara lebih nyata (Kemendikbud, 2017; Lickona, 1991).

Salah satu tantangan utama dalam penilaian afektif adalah dominannya penggunaan instrumen *self-report*, seperti angket skala Likert dan penilaian subjektif guru. Meskipun mudah diterapkan, metode ini rentan terhadap *social desirability bias*, yaitu kecenderungan peserta didik memberikan jawaban yang dianggap baik secara sosial daripada menggambarkan kondisi sebenarnya. Selain itu, minimnya penggunaan observasi dan penilaian teman sebaya menyebabkan hasil evaluasi sikap religius sering kali kurang objektif dan belum mampu merepresentasikan perilaku peserta didik secara

menyeluruh (Podsakoff et al., 2003; Paulhus, 2002).

Perkembangan teknologi pendidikan memberikan peluang untuk mengatasi keterbatasan tersebut melalui pemanfaatan *digital portfolio*. Instrumen ini memungkinkan dokumentasi perkembangan sikap religius secara berkelanjutan dengan memadukan berbagai sumber penilaian, seperti refleksi diri, observasi guru, dan *peer assessment*. Dengan demikian, *digital portfolio* berpotensi menghasilkan asesmen afektif yang lebih autentik, komprehensif, dan sesuai dengan tuntutan transformasi digital dalam pendidikan. Namun, penelitian mengenai penerapan *digital portfolio* sebagai instrumen asesmen sikap religius masih terbatas dan belum dipetakan secara sistematis (Barrett, 2007; Lorenzo & Ittelson, 2005).

Berdasarkan kondisi tersebut, penelitian ini menggunakan metode Systematic Literature Review (SLR) untuk mengidentifikasi perkembangan penelitian mengenai penggunaan *self-report*, observasi, *peer assessment*, serta transformasi menuju *digital portfolio* dalam pengukuran sikap

religius. Hasil kajian ini diharapkan dapat mengidentifikasi kesenjangan penelitian, memberikan gambaran tren perkembangan asesmen afektif, serta menjadi dasar pengembangan model evaluasi sikap religius yang lebih objektif, holistik, dan berbasis teknologi.

B. Metode Penelitian

Penelitian ini menggunakan metode Systematic Literature Review (SLR) dengan mengacu pada pedoman Kitchenham dan Charters. Metode ini dipilih untuk mengidentifikasi, menyeleksi, mengevaluasi, dan mensintesis hasil penelitian secara sistematis, transparan, serta dapat direplikasi.

Pencarian literatur dilakukan pada empat basis data, yaitu Google Scholar, GARUDA, SINTA, dan DOAJ, dengan rentang publikasi tahun 2015–2025. Kata kunci yang digunakan meliputi *religious attitude*, *religious attitude assessment*, *affective assessment*, *self-report*, *observation*, *peer assessment*, *digital portfolio*, *evaluasi afektif*, dan *sikap religius*, yang dikombinasikan menggunakan operator Boolean (AND dan OR).

Seleksi artikel mengikuti alur PRISMA, meliputi tahap identifikasi, penyaringan judul dan abstrak, penilaian teks lengkap, serta penetapan artikel yang memenuhi kriteria inklusi. Kriteria inklusi meliputi artikel jurnal *peer-reviewed* berbahasa Indonesia atau Inggris yang membahas pengukuran sikap religius atau asesmen afektif menggunakan instrumen *self-report*, observasi, *peer assessment*, maupun *digital portfolio*. Artikel yang tidak relevan, prosiding, tesis, disertasi, dan publikasi di luar rentang tahun penelitian dikeluarkan dari analisis.

Artikel yang lolos seleksi selanjutnya dievaluasi menggunakan Quality Assessment (QA) berdasarkan relevansi topik, kualitas metodologi, dan kontribusi penelitian. Data yang diperoleh kemudian dianalisis secara deskriptif untuk mengidentifikasi distribusi instrumen, tren penelitian, karakteristik metodologis, serta perkembangan penggunaan *digital portfolio* dalam asesmen sikap religius.

C. Hasil Penelitian dan Pembahasan

Karakteristik Umum Studi yang Dianalisis

Sebanyak 25 artikel memenuhi kriteria inklusi. Seluruh artikel (100%) menggunakan instrumen *self-report*, tiga artikel (12%) menggunakan observasi, dan tiga artikel (12%) menggunakan *peer assessment*. Tidak ada artikel yang menggunakan ketiga instrumen secara bersamaan. Tabel 1 menyajikan distribusi penggunaan instrumen dari ke-25 artikel tersebut.

Tabel 1. Frekuensi Kemunculan Instrumen dalam 25 Artikel

Jenis Instrumen	Jumlah Artikel	Persentase
<i>Self-report</i>	25	100%
Observasi	3	12%
<i>Peer assessment</i>	3	12%

Catatan: Satu artikel dapat menggunakan lebih dari satu instrumen, sehingga total persentase dapat melebihi 100%.

**Tabel 2. Kombinasi Instrumen
per Artikel**

Kombinasi Instrumen	Jumlah Artikel	Persentase
Hanya <i>Self-report</i>	22	88%
<i>Self-report</i>	2	8%
+ Observasi		
<i>Self-report</i>	1	4%
+ <i>Peer assessment</i>		
<i>t</i>		
<i>Self-report</i>	0	0%
+ Observasi		
+ <i>Peer assessment</i>		
<i>t</i>		

Catatan: Setiap artikel hanya dihitung satu kali. Total persentase = 100%.

RQ1: Bagaimana distribusi penggunaan instrumen *self-report*, observasi, dan *peer assessment* dalam literatur pengukuran sikap religius di pendidikan?

Konsep Dasar Pengukuran Sikap Religius dalam Ranah Afektif

Pengukuran sikap religius dalam evaluasi afektif pendidikan dapat dipahami sebagai usaha untuk menangkap seberapa dalam nilai-nilai agama meresap dalam diri peserta didik (Krathwohl et al., 1964). Ukuran ini tidak sederhana. Berbeda dengan ranah kognitif yang mudah diuji melalui soal, ranah afektif bertumpu pada kualitas penerimaan, respons, penghargaan, organisasi, dan karakterisasi nilai. Bloom dkk. merinci lima tingkatan ranah afektif: menerima (*receiving*), merespons (*responding*), menghargai (*valuing*), mengorganisasi (*organization*), dan karakterisasi (*characterization*) (Krathwohl et al., 1964). Deretan ini menunjukkan bahwa sikap keagamaan bukanlah sesuatu yang hitam-putih, melainkan bertahap mulai dari sekadar tahu nilai hingga menjadikannya bagian dari kepribadian. Dalam praktiknya, jenjang yang lebih tinggi (organisasi dan karakterisasi) jarang diukur secara sungguh-sungguh, sehingga banyak penelitian hanya sampai pada tingkat “menghargai” atau “merespons.” Kelemahan ini sering diabaikan. Krathwohl menekankan bahwa taksonomi afektif sebenarnya memberi alat untuk menilai

kedalaman internalisasi nilai, tetapi pemanfaatannya dalam pengukuran sikap religius masih minim (Krathwohl et al., 1964). Artinya, kerangka konseptual sudah tersedia, namun operasionalisasi di lapangan seringkali dangkal.

Sikap religius sendiri tidak bisa direduksi menjadi selebar angket. Glock merinci lima dimensi: keyakinan (ideologis), praktik ibadah (ritualis), penghayatan (eksperiensial), pengetahuan (intelektual), dan konsekuensi sosial (Glock & Stark, 1965). Masing-masing dimensi memerlukan pendekatan pengukuran yang berbeda. Misalnya, praktik ibadah bisa diamati, tetapi penghayatan spiritual sulit diungkap tanpa wawancara mendalam. Dalam studi-studi tentang pendidikan agama dan karakter, indikator yang sering dipakai antara lain ketekunan beribadah, kejujuran, toleransi terhadap perbedaan, dan kepedulian sosial (Lickona, 1991). Pemilihan indikator ini menunjukkan kecenderungan peneliti untuk mengukuhkan aspek yang mudah diukur bukan yang paling substansial. Akibatnya, konstruk “sikap religius” sering menyempit menjadi sekadar

perilaku tampak atau pernyataan setuju di atas kertas. Padahal, kelima dimensi Glock saling terkait; mengabaikan satu dimensi berarti kehilangan makna utuh dari religiusitas seseorang.

Dari sini, muncul pertanyaan kritis: apakah penelitian selama ini benar-benar mengukur sikap religius secara utuh, atau hanya sekadar mengukur persepsi diri yang telah “dibersihkan” siswa agar terlihat baik? Jawaban sementara dapat ditemukan dengan melihat sebaran instrumen yang digunakan. Jika sebagian besar studi mengandalkan *self-report* tanpa verifikasi dari sumber lain (observasi atau teman sebaya), maka ada risiko bias yang serius (Podsakoff et al., 2003; Paulhus, 2002).

Karakterisasi Tiga Pendekatan Instrumen

Dalam literatur yang dianalisis, pengukuran sikap religius dikarakterisasikan ke dalam tiga bentuk pendekatan utama, masing-masing dengan kekuatan dan kelemahan tersendiri.

Pertama, pendekatan *self-report* (laporan diri). Metode ini

merupakan yang paling dominan dan paling sering digunakan dalam penelitian pendidikan. Instrumen *self-report* biasanya menggunakan skala Likert dengan rentang lima hingga tujuh poin, di mana siswa diminta untuk menyatakan tingkat kesetujuan terhadap pernyataan-pernyataan yang berkaitan dengan keyakinan, praktik ibadah, dan moralitas religius. Keunggulan utama metode ini adalah efisiensi biaya, kemudahan administrasi, dan fleksibilitas analisis data (Podsakoff et al., 2003). Sebagaimana dijelaskan dalam literatur metodologi, *self-report* adalah “the preferred method of data collection for the majority of researchers, owing to their low cost, relative ease of use, and flexibility” serta merupakan format yang paling umum digunakan (“the most commonly used self-report measure format”) (Podsakoff et al., 2003).

Namun, ketergantungan pada *self-report* menghadirkan risiko serius berupa *social desirability bias*, yaitu kecenderungan responden untuk memberikan jawaban yang dianggap baik oleh lingkungan sosialnya daripada yang mencerminkan perilaku sesungguhnya. Dalam konteks

pengukuran religiusitas, individu religius dilaporkan cenderung melebih-lebihkan tingkat keramahan/kesopanan mereka, yang berarti mereka mengklaim memiliki sifat-sifat yang secara sosial dihargai namun belum tentu sesuai dengan realitas perilaku mereka (Sedikides & Gebauer, 2010). Penelitian menunjukkan bahwa perbedaan dalam religiusitas berkorelasi dengan perbedaan dalam kecenderungan memberikan respons bias, dan temuan ini menantang interpretasi bahwa korelasi antara religiusitas dan sifat kepribadian sepenuhnya mencerminkan realitas objektif. Dalam survei pendidikan, lebih dari 20% siswa sekolah dasar menunjukkan respons yang dipengaruhi oleh keinginan sosial, yang berdampak pada distorsi distribusi data dan penurunan validitas eksternal (Paulhus, 2002). *Social desirability bias* ini dapat bersifat tidak disengaja (*self-deception*) maupun disengaja (*impression management*), terutama dalam situasi penilaian yang dianggap penting (*high-stakes*) (Paulhus, 2002).

Kedua, pendekatan observasi. Metode ini menekankan pada pencatatan perilaku nyata yang dapat

diamati, seperti keaktifan dalam ibadah bersama, kejujuran saat ujian, atau sikap toleransi terhadap teman yang berbeda keyakinan (Fraenkel et al., 2012). Observasi sangat ideal untuk memverifikasi konsistensi antara ucapan (dalam *self-report*) dan tindakan. Dalam berbagai penelitian pembentukan karakter religius, observasi digunakan untuk menilai indikator perilaku seperti kemampuan melaksanakan ibadah, bersikap sopan, dan menunjukkan empati (Wijaya et al., 2023). Namun, penerapan observasi menghadapi tantangan praktis seperti kebutuhan waktu yang panjang, pelatihan *observer* secara intensif, serta penyusunan instrumen yang terstruktur dan terstandar. Selain itu, kehadiran *observer* berpotensi memicu *reactivity effect*, yaitu perubahan perilaku subjek karena sadar sedang diamati (Fraenkel et al., 2012). Bagi para peneliti praktisi, hambatan utama dalam menggunakan observasi adalah tuntutan waktu dan sumber daya yang cukup besar. Meskipun demikian, observasi yang dirancang dengan baik dapat menghasilkan data yang lebih objektif dibandingkan *self-report*,

karena ia menangkap perilaku aktual di luar pengakuan subjektif.

Ketiga, pendekatan *peer assessment* (penilaian teman sebaya). Metode ini melibatkan keterlibatan siswa untuk saling menilai aspek-aspek perilaku religius teman sekelasnya (Topping, 1998). Penilaian sejawat memiliki keunggulan unik karena teman sekelas memiliki frekuensi interaksi yang tinggi dan dapat mengamati perilaku sehari-hari yang mungkin tidak terlihat oleh guru. Penelitian menunjukkan bahwa *peer assessment* dalam konteks penilaian sikap spiritual berjalan efektif karena memudahkan guru mengevaluasi hasil belajar sekaligus melatih siswa memiliki keterampilan dalam melaksanakan evaluasi (Fitri, 2023). Dalam pengembangan instrumen, lembar observasi, penilaian diri (*self-assessment*), dan penilaian antar peserta didik (*peer-assessment*) telah dikembangkan secara terintegrasi dan terbukti valid serta reliabel untuk mengukur sikap spiritual (Sanjaya Putra et al., 2022). Pendekatan ini juga dapat mengurangi kecenderungan *social desirability bias* karena siswa cenderung lebih jujur

dalam menilai temannya daripada melaporkan dirinya sendiri.

Meskipun demikian, efektivitas *peer assessment* sangat bergantung pada ketersediaan rubrik penilaian yang jelas dan dapat dioperasikan, sosialisasi prosedur yang intensif kepada siswa, serta iklim kelas yang mendukung keterbukaan dan saling percaya (Topping, 1998). Studi pengembangan instrumen terkini menunjukkan bahwa *peer assessment* memerlukan pengujian validitas dan reliabilitas yang ketat, serta kesepakatan kelas dalam melaksanakan penilaian (Sanjaya Putra et al., 2022). Tanpa infrastruktur yang memadai, *peer assessment* rentan terhadap subjektivitas yang dipengaruhi oleh faktor-faktor non-teknis seperti popularitas atau kedekatan personal, serta dapat menimbulkan potensi konflik sosial di antara siswa.

RQ2: Bagaimana kombinasi dan tren penggunaan ketiga jenis instrumen tersebut berdasarkan tahun publikasi, negara asal peneliti, dan jenjang pendidikan?

Dari 25 artikel, tidak ditemukan satu pun yang menggunakan ketiga

instrumen (*self-report*, observasi, dan *peer assessment*) secara simultan. Sebanyak 22 artikel (88%) hanya mengandalkan *self-report* saja, dua artikel (8%) menggabungkan *self-report* dengan observasi, dan satu artikel (4%) menggabungkan *self-report* dengan *peer assessment*. Tidak ada artikel yang menggunakan ketiga instrumen sekaligus. Angka ini memperlihatkan bahwa pendekatan triangulasi masih sangat langka. Padahal, penilaian afektif yang ideal seharusnya menggabungkan laporan diri (persepsi subjektif), observasi (perilaku aktual oleh guru), dan penilaian teman (perspektif sosial) untuk memperoleh gambaran menyeluruh. Ketidadaan kombinasi penuh ini mengindikasikan adanya celah sistematis dalam praktik penelitian di lapangan.

Artikel yang menggunakan observasi atau *peer assessment* semuanya terbit setelah tahun 2020. Pada periode 2020–2025, mulai muncul tiga artikel yang menambahkan observasi dan tiga artikel yang menambahkan *peer assessment*. Meskipun peningkatannya kecil, kemunculan ini menandakan kesadaran yang mulai

tumbuh di kalangan peneliti tentang perlunya verifikasi eksternal. Namun, jumlahnya masih jauh dari memadai jika dibandingkan dengan total artikel yang hanya mengandalkan *self-report*.

Seluruh artikel yang menggunakan observasi dan *peer assessment* berasal dari Indonesia. Artikel internasional yang dianalisis (dari Brasil, Jerman, dan Turki) seluruhnya hanya mengandalkan *self-report*. Jenjang pendidikan yang paling banyak menerapkan metode non *self-report* adalah sekolah dasar dan madrasah tsanawiyah. Temuan ini memperkuat kesimpulan bahwa pendekatan triangulasi penuh belum pernah diimplementasikan dalam literatur pengukuran sikap religius.

RQ3: Pola metodologis apa yang dominan (misal skala Likert, rubrik observasi, format *peer assessment*) dalam setiap jenis instrumen?

Berdasarkan telaah terhadap 25 artikel yang dianalisis, teridentifikasi pola metodologis yang berbeda pada setiap jenis instrumen penilaian afektif. Pola-pola ini mencerminkan tingkat kematangan

dan standardisasi yang bervariasi antar instrumen.

Self-Report (Kuesioner/Angket)

Pola metodologis yang paling dominan pada instrumen *self-report* adalah penggunaan skala Likert dengan rentang lima hingga tujuh poin. Seluruh artikel yang menggunakan *self-report* mengadopsi format ini. Sebuah studi yang menilai karakter religius mahasiswa Muslim di Yogyakarta, misalnya, mengembangkan kuesioner berbasis skala Likert untuk mengukur enam dimensi karakter religius, dengan validitas dan reliabilitas yang diuji menggunakan *Confirmatory Factor Analysis* (CFA) dan *Structural Equation Modeling* (SEM) hingga menghasilkan koefisien reliabilitas yang tinggi ($\alpha = 0,95$) (Arrosyid et al., 2022). Selain itu, instrumen *self-report* untuk mengukur sikap religius juga dikembangkan dalam berbagai bentuk skala multidimensi, seperti skala identitas religius Glock dan Stark yang terdiri dari 19 item dengan skala Likert lima poin untuk mengukur aspek ritual, keyakinan, emosional, dan konsekuensi, serta *Religious Tendency Scale* yang memiliki

koefisien Cronbach alpha sangat tinggi (0,98 untuk subskala keyakinan dan 0,96 untuk subskala praktik dan perilaku) (Doğan & Tosun, 2019; Uzunpolat, 2025).

Dengan demikian, pola metodologis pada *self-report* menunjukkan keseragaman yang tinggi hampir seluruh instrumen menggunakan skala Likert, disertai uji validitas dan reliabilitas yang relatif ketat. Namun, sebagaimana dikemukakan dalam literatur, pengukuran berbasis *self-report* tetap menghadapi tantangan terkait hubungan antara laporan diri dan perilaku aktual, serta potensi pengaruh *social desirability* atau keinginan untuk konformitas (Podsakoff et al., 2003; Paulhus, 2002).

Pada instrumen observasi, pola yang teridentifikasi adalah pendekatan yang kurang terstandarisasi dan prosedural. Sebagian besar artikel yang menggunakan observasi tidak menyebutkan secara rinci durasi pengamatan, frekuensi, maupun pelatihan *observer*. Sebuah studi pengembangan instrumen untuk

menilai sikap spiritual siswa SMP, misalnya, menghasilkan lembar observasi terbuka (jurnal) yang terdiri dari 99 kriteria sikap spiritual, dengan validitas diuji menggunakan indeks Aiken ($V = 0,83$) (Sanjaya Putra et al., 2022). Selain itu, reliabilitas instrumen secara keseluruhan mencapai $\geq 75\%$. Temuan ini mengonfirmasi bahwa instrumen observasi yang dikembangkan valid dan reliabel untuk mengukur sikap spiritual peserta didik dalam konteks pendidikan Islam di MTs (Sanjaya Putra et al., 2022). Namun, hanya sedikit artikel yang melaporkan penggunaan lembar observasi terstruktur dengan indikator perilaku yang jelas. Dalam praktiknya, observasi digunakan oleh guru untuk menilai sikap religius siswa melalui kegiatan sehari-hari, baik di dalam maupun di luar kelas. Teknik observasi juga diterapkan untuk menilai aspek-aspek seperti sikap berdoa sebelum dan sesudah pelajaran, mengucapkan salam, beribadah tepat waktu, serta sikap toleransi (Wijaya et al., 2023). Namun demikian, laporan tentang reliabilitas antar-observer (*inter-rater reliability*) masih jarang ditemukan. Dengan demikian, pola metodologis pada observasi adalah kurang baku dan

prosedural lemah. Observasi masih sering diperlakukan sebagai instrumen pelengkap, bukan sebagai metode utama yang setara dengan *self-report*.

Pada instrumen *peer assessment*, pola yang dominan adalah penggunaan instrumen terstruktur dengan rubrik atau pernyataan yang seimbang, meskipun penerapannya masih terbatas. Sebuah studi pengembangan instrumen penilaian sikap spiritual untuk siswa SMP menghasilkan kuesioner *peer assessment* yang terdiri dari 93 pernyataan, 77 pernyataan positif dan 16 pernyataan negatif, yang terbukti valid dan memiliki reliabilitas sangat tinggi ($r_{11} = 0,990$) (Sanjaya Putra et al., 2022). Instrumen serupa juga dikembangkan untuk penilaian sikap spiritual dalam pembelajaran sains di SMP, dengan reliabilitas yang tergolong memuaskan berdasarkan uji Cronbach alpha (Sanjaya Putra et al., 2022).

Studi lain tentang efektivitas *peer assessment* dalam menilai sikap spiritual pada pembelajaran Pendidikan Agama Islam di SMP

menunjukkan bahwa metode ini efektif karena memudahkan guru mengevaluasi hasil belajar siswa sekaligus melatih siswa agar memiliki keterampilan dalam melaksanakan evaluasi (Fitri, 2023). *Peer assessment* juga terbukti berkontribusi pada peningkatan kualitas Pendidikan Agama Islam serta membentuk sikap kritis, reflektif, dan bertanggung jawab pada siswa (Fitri, 2023).

Namun demikian, dari 25 artikel yang dianalisis, hanya sebagian kecil yang menggunakan *peer assessment*, dan tidak semua artikel melaporkan validitas serta reliabilitas instrumen secara lengkap. Beberapa artikel hanya menyebutkan uji coba terbatas tanpa menyajikan data kuantitatif. Dengan demikian, pola metodologis pada *peer assessment* menunjukkan potensi yang besar, terutama ketika instrumen dirancang dengan pernyataan positif dan negatif yang seimbang serta rubrik yang jelas, namun penerapannya masih terbatas dan belum menjadi praktik yang mapan.

Perbandingan Antar Instrumen

Secara keseluruhan, ketiga jenis instrumen menunjukkan tingkat kematangan metodologis yang berbeda.

Instrumen	Pola Dominan	Kekuatan	Kelemahan
Self-report	Skala Likert 5–7 poin, uji validitas & reliabilitas ketat	Keseragaman tinggi, terstandarisasi	Rentan <i>social desirability bias</i>
Observasi	Lembar observasi terbuka, indikator perilaku	Menangkap perilaku aktual	Prosedur tidak baku, minim reliabilitas
Peer assessment	Kuesioner dengan	Reliabilitas tinggi,	Penerapan terbatas

Instrumen	Pola Dominan	Kekuatan	Kelemahan
	n	melatih siswa	s, belum mapan
	pernyataan		
	seimbang,		
	rubrik		

Temuan ini menunjukkan bahwa penelitian di bidang pengukuran sikap religius belum mencapai kematangan metodologis yang setara untuk semua jenis instrumen. *Self-report* memiliki akuntabilitas psikometrik yang baik, sementara observasi dan *peer assessment* masih memerlukan pengembangan instrumen yang lebih sistematis serta standar pelaporan yang lebih ketat.

RQ4: Bagaimana tren transformasi menuju *digital portfolio* sebagai instrumen asesmen afektif dalam literatur pendidikan Islam, serta apa saja tantangan dan peluang implementasinya?

Digital portfolio hadir sebagai alternatif yang menjawab tiga kelemahan metodologis yang

teridentifikasi pada RQ1, RQ2, dan RQ3.

Pertama, ia mengatasi dominasi *self-report* (RQ1) dengan tidak bergantung pada satu sumber data, melainkan mengumpulkan bukti dari berbagai sumber guru, siswa sendiri, teman sebaya dan berbagai konteks. Dengan demikian, ia mengurangi ketergantungan pada laporan diri yang subjektif dan rentan bias, serta mampu merekam perkembangan belajar secara berkelanjutan dan lebih komprehensif (Barrett, 2007).

Kedua, temuan RQ2 menunjukkan bahwa tidak ada artikel yang menggunakan ketiga instrumen secara simultan. *Digital portfolio* secara inheren bersifat integratif ia dapat memuat observasi guru, penilaian diri siswa, penilaian teman sebaya, serta hasil karya siswa (seperti tulisan, proyek, video, atau rekaman praktik ibadah) dalam satu platform terorganisir. Ia berpotensi menjadi wadah triangulasi yang selama ini absen dalam literatur, sekaligus memungkinkan refleksi dan umpan balik dari berbagai sudut pandang (Lorenzo & Ittelson, 2005).

Ketiga, RQ3 mengungkapkan bahwa observasi dan *peer assessment* minim standar prosedural dan jarang disertai bukti validitas. *Digital portfolio* yang dirancang dengan baik dapat menyediakan rubrik terstandar dan protokol jelas, sehingga meningkatkan reliabilitas dan akuntabilitas penilaian. Penelitian Iskandar, Hasani, dan Hidayatullah (2024) menunjukkan bahwa instrumen berbasis portofolio digital baru valid dan reliabel setelah melalui validasi ahli dan uji coba ketat, dengan koefisien reliabilitas rata-rata 0,783 (Iskandar et al., 2024). Ini menegaskan bahwa portofolio digital memiliki landasan psikometrik yang kuat jika dikembangkan secara sistematis.

Potensi Digital Portfolio dalam Pendidikan Islam

Beberapa penelitian menunjukkan bahwa portofolio digital memberikan dampak positif. Supriyadi et al. (2024) menemukan bahwa portofolio bukan sekadar kumpulan tugas, melainkan proses yang terlibat aktif dalam pembelajaran. Proses itu mencakup sembilan tahapan, mulai

dari penilaian indikator formatif hingga pameran karya siswa.

Penelitian lain oleh Wulandari et al. (2021) tentang penggunaan Padlet sebagai portofolio digital menunjukkan bahwa alat ini dapat mendorong pembelajaran yang reflektif dan mandiri, terutama jika guru dan siswa sama-sama terlibat dalam proses penilaian dan pemberian umpan balik (Wulandari et al., 2021).

Dalam konteks pendidikan Islam, Nor, Sari, dan Asfihana (2026) melaporkan bahwa *e-portfolio* yang berbasis nilai-nilai Islam berhasil meningkatkan keterlibatan siswa dan menjadikan pembelajaran lebih bermakna. Nilai-nilai seperti *amanah*, *sidq*, dan *muhasabah* terintegrasi secara natural ke dalam proses belajar (Nor et al., 2026).

Dengan demikian, secara konseptual dan berdasarkan bukti awal, portofolio digital memang menawarkan jalan keluar dari kelemahan metodologis yang ditemukan pada RQ1 hingga RQ3.

Hambatan yang Masih Menghalangi Terwujudnya Potensi Digital Portfolio

Namun, solusi ini belum benar-benar terwujud. Fadlillah dan Kusaeri (2024) mencatat bahwa sebagian besar studi tentang *e-portfolio* dalam PAI masih bersifat deskriptif dan tidak menggunakan uji komparatif atau eksperimental. Akibatnya, klaim bahwa *e-portfolio* meningkatkan keterlibatan dan refleksi diri belum memiliki bukti yang cukup kuat. Hambatan yang lebih nyata justru datang dari guru: pemahaman yang terbatas, prosedur yang rumit, dan waktu serta sumber daya yang tidak memadai (Fadlillah & Kusaeri, 2024).

Faktor lain yang tidak kalah penting adalah infrastruktur digital. Penelitian tentang asesmen psikomotorik di SMA Negeri Bandung menunjukkan bahwa standarisasi rubrik, pelatihan berkala, dan pemanfaatan portofolio digital dapat meningkatkan reliabilitas dan akuntabilitas penilaian. Namun, semua itu hanya dapat berjalan jika jaringan dan perangkat pendukung tersedia. Di luar Pulau Jawa, kondisi

ini masih jauh dari memadai (Hartono et al., 2025).

Persoalan lain adalah kompetensi guru dalam mengintegrasikan teknologi dengan nilai-nilai Islam. Kustandi et al. (2025) melaporkan bahwa pelatihan SeeSaw di SD Negeri Jakasetia IV berhasil meningkatkan skor pengetahuan guru dari 2,6 menjadi 4,3 dan keterampilan teknis dari 2,8 menjadi 4,2 (Kustandi et al., 2025). Namun, keberhasilan ini terjadi dalam situasi pelatihan intensif dengan pendampingan, sesuatu yang sulit ditemukan di sebagian besar madrasah. Tanpa dukungan serupa, portofolio digital berisiko menjadi beban tambahan bagi guru, bukan alat yang memperkaya proses belajar.

D. Kesimpulan

Penelitian ini meninjau secara sistematis 25 artikel tentang pengukuran sikap religius dalam evaluasi afektif pendidikan. Fokus tinjauan mencakup pola penggunaan instrumen, karakteristik metodologis, serta arah transformasi menuju portofolio digital. Dari kajian tersebut, teridentifikasi tiga temuan utama.

Pertama, hampir seluruh penelitian (100%) mengandalkan

instrumen *self-report*, sementara observasi dan *peer assessment* hanya muncul pada sebagian kecil studi. Temuan ini mengindikasikan bahwa praktik penilaian afektif selama ini masih sangat bergantung pada laporan diri yang rentan terhadap berbagai bentuk bias.

Kedua, tidak ditemukan satu pun penelitian yang menggabungkan ketiga instrumen secara bersamaan. Mayoritas studi hanya menggunakan satu atau dua metode, dan belum ada yang menerapkan triangulasi penuh. Pola ini menunjukkan bahwa pendekatan yang lebih terpadu belum menjadi bagian dari praktik penelitian di lapangan.

Ketiga, dari sisi metodologi, *self-report* memiliki prosedur yang lebih baku dan telah teruji secara psikometrik. Sebaliknya, observasi dan *peer assessment* masih minim standar dan jarang disertai bukti validitas. Kondisi ini mencerminkan ketimpangan kematangan metodologis antar instrumen yang tersedia.

Berangkat dari ketiga kelemahan tersebut, portofolio digital

muncul sebagai alternatif yang menjanjikan. Ia mampu mengumpulkan bukti dari berbagai sumber, menyediakan ruang triangulasi, serta menawarkan rubrik dan protokol yang lebih terstandar. Sejumlah penelitian awal juga menunjukkan bahwa portofolio digital dapat mendorong refleksi siswa dan memperkuat internalisasi nilai-nilai Islam.

Namun, potensi itu belum terwujud secara luas. Implementasi portofolio digital masih terhambat oleh keterbatasan literasi digital guru, minimnya infrastruktur, dan belum tersedianya pelatihan yang berkelanjutan. Tanpa perbaikan pada ketiga aspek tersebut, portofolio digital berisiko menjadi sekadar formalitas sama seperti observasi yang selama ini hanya berfungsi sebagai pelengkap, bukan instrumen penilaian yang benar-benar dimanfaatkan.

Dengan demikian, portofolio digital adalah solusi yang tepat sasaran, tetapi bukan solusi yang dapat bekerja dengan sendirinya. Keberhasilannya sangat bergantung pada kesiapan sistem yang menyertainya guru yang terlatih, infrastruktur yang memadai,

instrumen yang terstandar, serta manajemen madrasah yang mendukung. Tanpa keempat unsur tersebut, transformasi yang diharapkan akan sulit terwujud.

Secara keseluruhan, penelitian ini menyimpulkan bahwa pengukuran sikap religius dalam evaluasi afektif masih memerlukan perubahan mendasar baik dari sisi instrumen, prosedur, maupun sistem pendukungnya. Portofolio digital menawarkan jalan keluar, tetapi jalan itu hanya akan ditempuh jika semua unsur yang mendukungnya bergerak bersama.

DAFTAR PUSTAKA

Artikel in Press :

- Ahmadi, S., Mirfardi, A., & Ebtokari, M. H. (2014). Relationship between religious identity and social capital (Research subject: Yasouj). *Applied Sociology (Jāmi'ah'shināshī-i kārburdī)*, 25(4), 1–16. <https://doaj.org/article/04fa8180a2634a12ae28f4ebd7d401e2>
- Ali, A. K. (2024). Implementasi e-portofolio digital sebagai instrumen penilaian autentik di perguruan tinggi: Systematic literature review. *Jurnal Inovasi Teknologi Pendidikan*, 2(4), 306–312.

- Braghetta, C. C., et al. (2021). Development of an instrument to assess spirituality: Reliability and validation of the Attitudes Related to Spirituality Scale (ARES). *Frontiers in Psychology*, 12, Article 764132. <https://doi.org/10.3389/fpsyg.2021.764132>
- Dewi, T. N. (2015). *Pengembangan instrumen penilaian sikap untuk mengukur sikap spiritual peserta didik dalam pembelajaran IPA di Madrasah Tsanawiyah* [Tesis Magister, UNY]. <http://eprints.uny.ac.id/26039/>
- Dewi, T. N. (2021). Instrumen penilaian untuk mengukur sikap spiritual dalam pembelajaran IPA di Madrasah Tsanawiyah. *Measurement in Educational Research*, 1(1), 18–26. <https://doi.org/10.1177/074193258300400305>
- Fauzi, A., Posangi, S. S., & Dilo, U. A. (2022). Model penilaian berbasis portofolio dalam pengembangan ranah afektif santri di Pesantren Hubulo Kabupaten Bone Bolango Provinsi Gorontalo. *Al-Bahtsu: Jurnal Penelitian Pendidikan Islam*, 7(2), 148–162.
- Fitri, F. (2023). The effectiveness of peer assessment in assessing students' spiritual attitudes in Islamic religious education learning in middle schools. *Jurnal Pendidikan Agama Islam Indonesia (JPAIL)*, 4(4), 95–98. <https://doi.org/10.37251/jpai.v4i4.784>
- Goetz, T., et al. (2014). Between-domain relations of students' academic emotions and their judgments of school domain similarity. *Frontiers in Psychology*, 5, Article 1153. <https://doi.org/10.3389/fpsyg.2014.01153>
- Glock, C. Y. (1962). On the study of religious commitment. *Religious Education*, 57(Research Supplement), 98–110. <https://doi.org/10.1080/003440862057S407>
- Hermisson, S., Gochyyev, P., & Wilson, M. (2019). Assessing pupils' attitudes towards religious and worldview diversity Development and validation of a nuanced measurement instrument. *British Journal of Religious Education*, 41(4), 371–387. <https://doi.org/10.1080/01416200.2018.1556604>
- Iskandar, D., Hasani, A., & Hidayatullah, H. (2024). Advancing teacher evaluation: E-portfolio development. *Tadris: Jurnal Keguruan dan Ilmu Tarbiyah*, 9(1), 43–51. <https://doi.org/10.24042/tadris.v9i1.22232>
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy*

- of educational objectives: The classification of educational goals. *Handbook II: Affective domain*. Longmans.
- Kreitschmann, R. S., Abad, F. J., Ponsoda, V., Nieto, M. D., & Morillo, D. (2019). Controlling for response biases in self-report scales: Forced-choice vs. psychometric modeling of Likert items. *Frontiers in Psychology*, 10, Article 2309. <https://doi.org/10.3389/fpsyg.2019.02309>
- Kunuba, A. C. (2024). Teachers' perception of peer assessment technique of secondary school students' learning in Christian religious studies in Aniocha South LGA of Delta State. *ASSEREN Journal of Educational Research and Development*, 11, 121–126.
- Laily, F. N., & Himmah, H. (2026). Implementasi penilaian berbasis portofolio digital dalam pembelajaran privat bahasa Arab. *Jurnal Itjmie*, 7(2), 376–386. <https://doi.org/10.32832/itjmie.v7i2.22954>
- Lestari, T. (2020). *Pola pengembangan karakter religius pada siswa di SMP IT Rabbi Radhiyya Rejang Lebong* [Tesis Magister, IAIN Curup]. <https://e-theses.iaincurup.ac.id/2697/1/Tesis%20full%20text.pdf>
- Muharrozi, M., Anwar, S., & Romli, U. (2025). Strengthening holistic assessment in Islamic religious education: Cognitive, psychomotor, and affective domains. *At-Tarbawi: Jurnal Kajian Kependidikan Islam*, 10(2), 174–190.
- Moynihan, A. B., Guinote, A., & Igou, E. R. (2023). Relational dynamics and meaning in life: Dominance predicts perceived social support, belongingness, and meaning in life. *Personality and Individual Differences*, 211, Article 112249. <https://doi.org/10.1016/j.paid.2023.112249>
- Nancy, E. A. (2015). Bloom's taxonomy of cognitive learning objectives. *Journal of the Medical Library Association*, 103(3), 152–153. <https://doi.org/10.3163/1536-5050.103.3.010>
- Uzunpolat, Y. (2025). Developing an attitude scale towards religious culture and ethical knowledge course and examining it in terms of various variables. *van Ilahiyat Dergisi*, 12(21), 54–71.
- Palem, N. H., Darmawan, F., & Hasibuan, F. D. (2025). Analisis penerapan teknik non-tes dalam menilai karakter religius siswa di SDN 060834 Kecamatan Medan Petisah. *Jurnal Pendidikan Dasar*, 5(3), 456–468.
- Penthin, M., Pirner, M. L., Scheunpflug, A., & Kröner, S. (2022). Measuring student teachers' beliefs regarding

- religion in schools: Piloting and preliminary validation of two newly developed scales. *British Journal of Religious Education*, 44(3), 281–292. <https://doi.org/10.1080/01416200.2021.2007047>
- Rahman, A., et al. (2026). Effectiveness and challenges of digital and conventional Islamic religious education learning media: Quantitative analysis based on SLR. *Nidhomiyyah: Jurnal Manajemen Pendidikan Islam*, 7(1), 105–117. <https://doi.org/10.38073/nidhomiyyah.v7i1.3818>
- Rukiyati, et al. (2025). Assessing religious character among Muslim university students in Yogyakarta Indonesia. *Jurnal Pendidikan Agama Islam*, 22(1), 157–174. <https://doi.org/10.14421/jp.ai.v22i1.10927>
- Rohmah, A. A., Abdurrochim, J. A., & Habibullah, M. R. (2026). Evaluasi berbasis IT: E-assessment dan e-portofolio pada pembelajaran Pendidikan Agama Islam (PAI). *THORIQOTUNA: Jurnal Pendidikan Islam*, 9(1).
- Sanjaya Putra, I. P., et al. (2022). Instrumen penilaian sikap spiritual dan sikap sosial siswa kelas IV Sekolah Dasar tema indahnnya keberagaman di negeriku. *Jurnal Pendidikan Dasar*, 10 (1), 45–60.
- Sulani, P. (2020). Instrument for assessing spiritual attitudes of Buddhist education and manners at junior high school level. *Jurnal Pencerahan*, 14(1), 30–43. <https://paperity.org/p/341479016/instrument-for-assessing-spiritual-attitudes-of-buddhist-education-and-manners-at-junior>
- Susmiyati, S., Zurqoni, Z., Abdillah, M. H., & Saugi, W. (2023). Challenges of affective assessment of Islamic Religious Education learning in Merdeka Curriculum. *Al-Hayat: Journal of Islamic Education*, 7(2), 710. <https://doi.org/10.35723/ajie.v7i2.675>
- Triwulandari. (2023). *Pengembangan instrumen penilaian sikap terhadap agama (Attitude Toward Religious) berbasis teknik peer and self assessment* [Tesis Magister, UIN Sunan Kalijaga].
- Wijaya, A., et al. (2023). The implementation of affective and psychomotor assessment in Islamic religious education learning in junior high schools. *TARBAWY: Indonesian Journal of Islamic Education*, 10(2), 195–206. <https://doi.org/10.17509/t.v10i2.65467>
- Zubaidah, Z., Suraiya, S., Manan, A., Kamarullah, K., Embong, R., & Nurrahmi, N. (2025). Affective

assessment within madrasa
teachers: An analysis of iterative
thematic inquiry. *Jurnal Ilmiah*
Peuradeun, 14(1). [https://doi.org/
10.26811/peuradeun.v14i1.14
59](https://doi.org/10.26811/peuradeun.v14i1.1459)