

Penerapan Model Kirkpatrick dalam Mengukur Efektivitas Pembelajaran Evaluasi Program dan Kebijakan Pendidikan Islam

Budihartono⁽¹⁾, Sitti Mania⁽²⁾, Muhammad Nur Akbar Rasyid⁽³⁾

^{1,2,3} Universitas Islam Negeri Alauddin Makassar, Indonesia

Email: ¹198208152010011032@kemenag.go.id, ²sitti.mania@uin-alauddin.ac.id,
³akbar.rasyid@uin-alauddin.ac.id.

Abstrak: Penelitian ini menganalisis efektivitas perkuliahan Evaluasi Program dan Kebijakan Pendidikan Islam di UIN Alauddin Makassar menggunakan kerangka evaluasi empat tingkat Kirkpatrick. Metode kuantitatif deskriptif diterapkan dan diperkaya data kualitatif dari pertanyaan terbuka. Partisipan adalah 18 mahasiswa doktoral dengan teknik total sampling. Data dikumpulkan melalui kuesioner skala Likert lima tingkat yang mencakup empat dimensi: reaksi, pembelajaran, perilaku, dan hasil. Temuan menunjukkan perkuliahan efektif pada seluruh dimensi: reaksi (M=4,63; 92,50%), pembelajaran (M=4,31; 86,30%), dan hasil (M=4,38; 87,56%) berkategori sangat baik, sedangkan perilaku (M=4,07; 81,48%) berkategori baik. Transfer gap antara penguasaan konseptual dan implementasi praktis pada dimensi perilaku merupakan temuan kritis yang memerlukan respons pedagogis. Penelitian ini berkontribusi pada validasi dan ekspansi kerangka Kirkpatrick ke konteks pendidikan tinggi keagamaan Islam (PTKI) doktoral yang sebelumnya belum tereksplorasi, sekaligus menyediakan basis empiris bagi penguatan penjaminan mutu akademik di lingkungan PTKI.

Tersedia Online di

http://journal.unublitar.ac.id/pendidikan/index.php/Riset_Konseptual

Sejarah Artikel

Diterima pada : 01-02-2026

Disetujui pada : 20-03-2026

Dipublikasikan pada : 28-04-2026

Kata Kunci:

Model Kirkpatrick, efektivitas pembelajaran, evaluasi pembelajaran, evaluasi program, kebijakan pendidikan Islam

DOI:

http://doi.org/10.28926/riset_konseptual.v10i2.1480

PENDAHULUAN

Penilaian merupakan komponen fundamental dalam pengelolaan pendidikan tinggi untuk memastikan pembelajaran tidak hanya berjalan secara prosedural, tetapi juga mencapai tujuan yang telah ditetapkan (Anshari, 2014; Fitriyyah, 2025). Dalam konteks perguruan tinggi keagamaan Islam (PTKI), penilaian memiliki makna yang lebih strategis karena harus mengintegrasikan capaian akademik, kompetensi profesional, nilai-nilai Islam, serta kebijakan pendidikan yang kontekstual (Kurniawan et al., 2025; Sudradjat & Alim, 2025). Integrasi berbagai dimensi ini menjadikan penilaian di PTKI lebih kompleks dibandingkan dengan perguruan tinggi umum, sekaligus menegaskan urgensi pengembangan sistem evaluasi pembelajaran yang komprehensif dan berbasis bukti.

Salah satu mata kuliah yang berperan penting dalam membangun kompetensi evaluatif mahasiswa adalah Evaluasi Program dan Kebijakan Pendidikan di UIN Alauddin Makassar, khususnya pada jenjang doktoral. Mata kuliah ini dirancang untuk membekali mahasiswa dengan kemampuan teoretis dan praktis dalam mengevaluasi program dan kebijakan pendidikan secara sistematis dan ilmiah (Taali et al., 2024). Mahasiswa diharapkan mampu memahami berbagai model evaluasi, mengidentifikasi kekuatan dan kelemahan program, serta menyusun rekomendasi berbasis data yang relevan dengan konteks kebijakan pendidikan Islam. Namun, keberadaan mata kuliah tidak otomatis menjamin efektivitas pembelajaran tanpa adanya evaluasi komprehensif terhadap proses dan hasil perkuliahan itu sendiri. Selama ini, penilaian mata kuliah cenderung berfokus pada capaian kognitif atau kepuasan mahasiswa semata, tanpa

mengkaji secara menyeluruh hubungan antara pengalaman belajar, perubahan perilaku akademik, dan dampaknya terhadap praktik evaluatif di lapangan.

Salah satu kerangka evaluasi yang banyak digunakan dan terbukti efektif adalah model Kirkpatrick, yang mencakup empat dimensi: reaksi, pembelajaran, perilaku, dan hasil (D. Kirkpatrick & Kirkpatrick, 2006). Model ini tidak hanya mengukur kepuasan peserta, tetapi juga peningkatan pengetahuan, perubahan perilaku, serta dampak konkret program. Awalnya dikembangkan untuk pelatihan organisasi pada akhir 1950-an, model Kirkpatrick kemudian diadaptasi secara luas dalam konteks pendidikan formal, termasuk pendidikan tinggi (D. Kirkpatrick & Kirkpatrick, 2006). Kirkpatrick menekankan bahwa efektivitas pembelajaran harus dinilai secara berjenjang, mulai dari pengalaman awal peserta hingga dampak akhir yang dihasilkan, sehingga memberikan gambaran yang utuh dan multidimensi tentang kualitas suatu program pembelajaran.

Sejumlah penelitian telah membuktikan efektivitas model ini dalam konteks Indonesia. Ramadhon (2016) menunjukkan bahwa penerapan model Kirkpatrick di Pusdiklat Migas mampu mengoptimalkan fungsi evaluasi diklat aparatur secara sistematis, sehingga kelebihan dan kekurangan program dapat diidentifikasi dan ditindaklanjuti secara tepat. Sopacua dan Budijanto (2007) membuktikan efektivitasnya sebagai alat evaluasi pascapelatihan yang valid dan reliabel dalam mengukur kepuasan peserta, capaian pembelajaran, dan dampak terhadap kinerja. Utomo dan Tehupeior (2014) menemukan bahwa evaluasi berbasis Kirkpatrick mampu memberikan umpan balik bermakna untuk peningkatan kualitas pelatihan melalui perbaikan kompetensi trainer, penyesuaian materi, dan pengukuran dampak terhadap performa kerja peserta.

Namun, terdapat kesenjangan penelitian yang signifikan: seluruh studi tersebut diterapkan pada konteks pelatihan aparatur dan organisasi, bukan pada evaluasi mata kuliah di lingkungan PTKI. Penerapan kerangka Kirkpatrick pada konteks PTKI memiliki kompleksitas tersendiri karena harus mengakomodasi dimensi keilmuan Islam, tuntutan kompetensi penelitian doktoral, dan konteks kebijakan pendidikan nasional yang tidak dijumpai dalam evaluasi pelatihan konvensional. Dengan demikian, tidak dapat diasumsikan bahwa temuan dari konteks pelatihan aparatur secara langsung berlaku untuk mata kuliah doktoral di lingkungan UIN, sehingga penelitian khusus pada konteks ini menjadi suatu keharusan ilmiah, bukan sekadar pelengkap literatur.

Kesenjangan penelitian ini dapat diidentifikasi dalam tiga dimensi yang saling berkaitan. Pertama, belum adanya studi yang secara spesifik menerapkan kerangka Kirkpatrick untuk mengevaluasi mata kuliah—bukan program pelatihan—di lingkungan PTKI. Kedua, absennya bukti empiris mengenai bagaimana mahasiswa doktoral pendidikan Islam mentransformasikan pemahaman evaluatif ke dalam praktik akademik dan penelitian, khususnya dalam konteks penyusunan disertasi dan perencanaan evaluasi kebijakan. Ketiga, minimnya instrumen evaluasi pembelajaran yang dikontekstualisasikan dengan nilai, tujuan, dan karakteristik pendidikan Islam di jenjang doktoral. Tanpa pengisian gap ini, pengelola program studi tidak memiliki basis empiris yang memadai untuk memastikan bahwa lulusan benar-benar memiliki kemampuan evaluatif yang aplikatif dan kontekstual sesuai tuntutan kebijakan pendidikan nasional.

Berdasarkan identifikasi kesenjangan tersebut, penelitian ini merumuskan empat pertanyaan penelitian sebagai berikut: (1) Bagaimana reaksi mahasiswa doktoral terhadap pelaksanaan perkuliahan Evaluasi Program dan Kebijakan Pendidikan Islam? (2) Sejauh mana peningkatan pemahaman dan penguasaan konsep evaluasi mahasiswa setelah mengikuti perkuliahan? (3) Bagaimana mahasiswa mengimplementasikan konsep evaluasi dalam aktivitas akademik mereka? (4) Apa dampak akhir perkuliahan terhadap pengembangan kompetensi evaluatif mahasiswa doktoral?

Secara konseptual, penilaian program merupakan proses sistematis untuk mengumpulkan, menganalisis, dan menafsirkan informasi guna menilai nilai, efektivitas, dan dampak suatu program dalam mencapai tujuannya (Fitzpatrick et al., 2011; Stufflebeam & Coryn, 2014). Dalam pendidikan tinggi, penilaian program berfungsi sebagai instrumen pengendalian mutu sekaligus dasar pengambilan keputusan untuk perbaikan berkelanjutan. Sukardi (2022) menegaskan bahwa penilaian harus mencakup seluruh komponen program—mulai dari perencanaan, pelaksanaan, hingga hasil—sehingga evaluasi tidak hanya berorientasi pada output, tetapi juga pada proses yang membentuknya. Dalam konteks PTKI, penilaian memiliki dimensi tambahan karena harus menjamin ketercapaian tujuan pendidikan Islam yang mencakup aspek kognitif, afektif, spiritual, dan profesional secara terintegrasi (Ramadhon, 2016). Keempat dimensi Kirkpatrick—reaksi, pembelajaran, perilaku, dan hasil—saling berkaitan secara logis dan fungsional, serta bersifat fleksibel untuk berbagai konteks pembelajaran termasuk pendidikan tinggi berbasis kompetensi (D. Kirkpatrick & Kirkpatrick, 2006), menjadikannya kerangka yang tepat untuk menilai mata kuliah sebagai program pendidikan mikro pada jenjang doktoral.

METODE

Penelitian ini didesain untuk menilai pelaksanaan perkuliahan Evaluasi Program dan Kebijakan Pendidikan pada Program Doktor UIN Alauddin Makassar Tahun 2025 dengan memanfaatkan kerangka evaluasi Kirkpatrick sebagai landasan utama. Pendekatan metodologis yang diterapkan adalah pendekatan kuantitatif dengan jenis penelitian deskriptif. Pendekatan ini dipilih karena tujuan utama kajian bukan untuk menguji hipotesis atau menjelaskan hubungan kausal antarvariabel secara inferensial, melainkan untuk memperoleh gambaran empiris yang sistematis mengenai tingkat efektivitas pembelajaran berdasarkan empat dimensi evaluasi Kirkpatrick (Creswell, 2012). Pendekatan deskriptif memungkinkan peneliti memetakan kondisi aktual pelaksanaan mata kuliah secara objektif berdasarkan persepsi dan pengalaman mahasiswa.

Kerangka evaluasi Kirkpatrick digunakan secara utuh tanpa modifikasi atau penggabungan dengan model penilaian lain. Penggunaan model ini dimaksudkan untuk memetakan pengalaman belajar mahasiswa secara berjenjang, mulai dari respons awal terhadap proses pembelajaran, pemahaman dan penguasaan materi, penerapan hasil pembelajaran dalam aktivitas akademik, hingga dampak akhir terhadap kompetensi evaluatif mahasiswa doktoral. Keempat dimensi tersebut dipahami sebagai aspek evaluatif yang berdiri sendiri, sehingga hasil analisis difokuskan pada deskripsi capaian masing-masing dimensi.

Partisipan kajian adalah seluruh mahasiswa Program Doktor UIN Alauddin Makassar Tahun 2025 yang mengikuti perkuliahan Evaluasi Program dan Kebijakan Pendidikan Islam. Mengingat jumlah mahasiswa yang relatif terbatas dan seluruh populasi dapat dijangkau, kajian ini menggunakan teknik total sampling, sehingga seluruh mahasiswa yang mengikuti mata kuliah tersebut dijadikan responden. Jumlah responden yang mengisi kuesioner secara lengkap sebanyak 18 mahasiswa dengan latar belakang pendidikan magister yang beragam, meliputi pendidikan Islam, manajemen pendidikan, dan studi kebijakan pendidikan. Sebagian besar responden berstatus sebagai dosen dan praktisi pendidikan, sehingga memiliki pengalaman akademik dan profesional yang relevan untuk memberikan penilaian yang kritis dan reflektif terhadap perkuliahan.

Instrumen kajian disusun secara sistematis mengacu pada indikator-indikator yang relevan dengan setiap dimensi evaluasi Kirkpatrick. Pada dimensi reaksi, delapan indikator mencakup kejelasan penyampaian materi, relevansi materi dengan kebutuhan akademik doktoral, efektivitas metode pembelajaran, penguasaan dosen, peran media pembelajaran, kualitas interaksi dosen-mahasiswa, kemampuan diskusi mendorong pemikiran kritis, dan kepuasan umum. Pada dimensi pembelajaran, enam indikator mencakup pemahaman konsep dasar evaluasi program, pemahaman

evaluasi kebijakan pendidikan Islam, kemampuan membedakan evaluasi formatif dan sumatif, pemahaman langkah penyusunan desain evaluasi, kemampuan menganalisis studi kasus, dan persepsi peningkatan pengetahuan secara signifikan. Pada dimensi perilaku, enam indikator mencakup penerapan konsep dalam tugas, penerapan dalam analisis kebijakan, penerapan penyusunan desain evaluasi, penerapan dalam rencana disertai, penerapan dalam mengkritisi kebijakan secara evaluatif, dan pengaruh terhadap cara berpikir akademik. Pada dimensi hasil, lima indikator mencakup peningkatan kompetensi evaluatif, kesiapan melakukan evaluasi program, dampak terhadap kualitas penelitian, kontribusi pada pengambilan kebijakan, dan kontribusi pada pengembangan keilmuan.

Selain instrumen skala Likert lima tingkat, kajian ini mengumpulkan data kualitatif melalui pertanyaan terbuka yang meliputi kelebihan utama mata kuliah, kendala selama perkuliahan, saran peningkatan, dan pengaruh terhadap disertai atau penelitian evaluasi. Data kualitatif ini berfungsi sebagai triangulasi untuk memperkuat dan mengonfirmasi temuan kuantitatif, meningkatkan kredibilitas hasil penelitian melalui berbagai perspektif (Miles et al., 2013).

Kelayakan instrumen dikonfirmasi melalui uji validitas dan reliabilitas. Validitas konstruk diuji melalui nilai loading factor ($\geq 0,70$) dan Average Variance Extracted ($\geq 0,50$), sedangkan reliabilitas konstruk diuji melalui Composite Reliability dan Cronbach's Alpha ($\geq 0,70$). Perlu ditegaskan bahwa pengujian ini dilakukan semata-mata untuk memastikan kelayakan instrumen pengukuran—bukan untuk menguji model struktural antarkonstruk—sesuai dengan desain penelitian deskriptif yang tidak memerlukan analisis hubungan antarvariabel secara inferensial. Seluruh indikator memenuhi ambang batas yang ditetapkan, sehingga instrumen dinyatakan valid dan reliabel untuk digunakan dalam pengumpulan data.

Data yang diperoleh dari kuesioner dianalisis menggunakan statistik deskriptif sederhana dengan menghitung skor rata-rata dan persentase respons untuk setiap item dan setiap dimensi evaluasi Kirkpatrick. Hasil analisis disajikan dalam bentuk tabel dan uraian naratif yang menjelaskan makna temuan empiris tersebut. Analisis tidak melibatkan uji hipotesis, uji signifikansi statistik, maupun analisis inferensial lainnya, karena tujuan kajian secara khusus dibatasi pada pemetaan dan deskripsi hasil penilaian per dimensi.

HASIL dan PEMBAHASAN

Bagian ini menyajikan hasil evaluasi perkuliahan Evaluasi Program dan Kebijakan Pendidikan Islam pada Program Doktor UIN Alauddin Makassar Tahun 2025 berdasarkan keempat dimensi kerangka Kirkpatrick. Data yang dianalisis berasal dari kuesioner skala Likert yang dilengkapi pertanyaan terbuka, sehingga hasil kajian tidak hanya bersifat kuantitatif deskriptif tetapi juga diperkaya narasi reflektif mahasiswa..

Hasil Evaluasi Dimensi Reaksi

Dimensi reaksi dalam kerangka Kirkpatrick mengukur respons awal peserta terhadap pelaksanaan program pembelajaran, mencakup kepuasan terhadap materi, metode, fasilitator, dan kondisi belajar secara keseluruhan. Temuan pada dimensi ini memperlihatkan bahwa mahasiswa memberikan tanggapan yang sangat positif terhadap pelaksanaan perkuliahan, dengan rata-rata skor 4,63 dan persentase 92,50 persen pada kategori sangat baik.

Tabel 1. Hasil Evaluasi Dimensi Reaksi Mahasiswa.

No	Indikator Reaksi	Mean	Persentase (%)	Kategori
1	Materi mata kuliah disampaikan dengan jelas dan sistematis	4,78	95,56	Sangat Baik
2	Materi perkuliahan relevan dengan kebutuhan akademik mahasiswa doktoral	4,78	95,56	Sangat Baik
3	Metode pembelajaran yang digunakan	4,50	90,00	Sangat Baik

dosen efektif				
4	Dosen menguasai materi evaluasi program dan kebijakan pendidikan Islam	4,72	94,44	Sangat Baik
5	Media pembelajaran membantu pemahaman materi	4,56	91,11	Sangat Baik
6	Interaksi dosen dan mahasiswa berjalan dengan baik	4,72	94,44	Sangat Baik
7	Diskusi kelas mendorong pemikiran kritis dan reflektif	4,39	87,78	Sangat Baik
8	Secara umum puas mengikuti mata kuliah evaluasi	4,56	91,11	Sangat Baik
Rata-rata		4,63	92,50	Sangat Baik

Berdasarkan Tabel 1, seluruh indikator dimensi reaksi berada pada kategori sangat baik tanpa pengecualian. Aspek kejelasan dan sistematika penyampaian materi serta relevansi materi dengan kebutuhan akademik mahasiswa doktoral memperoleh skor tertinggi, masing-masing 4,78 dengan persentase 95,56 persen. Skor yang tinggi pada kedua aspek ini mengindikasikan bahwa mahasiswa menilai konten perkuliahan sangat sesuai dengan kebutuhan dan konteks akademik mereka sebagai calon doktor, bukan sekadar pengetahuan generik yang kurang relevan dengan tuntutan riset doktoral. Penguasaan materi oleh dosen dan kualitas interaksi dosen-mahasiswa masing-masing memperoleh skor 4,72 dengan persentase 94,44 persen, menunjukkan bahwa kompetensi substantif pengajar dan kemampuannya membangun komunikasi akademik yang produktif diapresiasi sangat tinggi oleh mahasiswa.

Kepuasan umum mahasiswa dalam mengikuti perkuliahan dan efektivitas media pembelajaran memperoleh skor 4,56 dengan persentase 91,11 persen, mencerminkan bahwa penggunaan berbagai media pendukung pembelajaran telah berhasil memperkuat proses transfer pengetahuan secara efektif. Efektivitas metode pembelajaran memperoleh skor 4,50 dengan persentase 90,00 persen, mengindikasikan bahwa strategi pengajaran yang diterapkan berhasil memfasilitasi keterlibatan aktif mahasiswa dalam proses belajar. Indikator dengan skor terendah pada dimensi ini adalah kemampuan diskusi kelas mendorong pemikiran kritis dan reflektif, yakni 4,39 dengan persentase 87,78 persen. Meskipun skor ini sedikit lebih rendah dibandingkan indikator lain, nilainya tetap berada pada kategori sangat baik dan mengindikasikan bahwa kegiatan diskusi telah berhasil membangun atmosfer intelektual yang kondusif bagi pengembangan pemikiran kritis mahasiswa.

Skor yang konsisten tinggi pada seluruh indikator mengonfirmasi terpenuhinya prasyarat awal pembelajaran yang efektif. Namun secara kritis, tingginya skor reaksi tidak dengan sendirinya menjamin kedalaman pembelajaran yang sesungguhnya. Kirkpatrick dan Kirkpatrick (2016) memperingatkan potensi "entertainment effect"—kondisi di mana peserta merasa puas dan nyaman tetapi belum tentu memperoleh peningkatan kompetensi yang substansial. Dalam konteks doktoral, risiko ini relatif lebih rendah mengingat mahasiswa memiliki kemampuan metakognitif yang matang dalam menilai kualitas pembelajaran secara kritis. Berbeda dari studi Ramadhon (2016) dan Sopacua & Budijanto (2007) yang melibatkan peserta pelatihan aparatur dengan kapasitas evaluatif yang beragam, mahasiswa doktoral dalam penelitian ini merupakan akademisi berpengalaman yang penilaiannya lebih terstruktur dan berbasis pengalaman akademik yang kaya. Oleh karena itu, skor reaksi tinggi dalam konteks ini memiliki bobot validitas yang lebih kuat. Data kualitatif memperkuat temuan ini: mahasiswa mengungkapkan bahwa suasana perkuliahan memberikan ruang yang nyaman dan produktif untuk dialog ilmiah, pertukaran gagasan antar akademisi, dan pendalaman konsep evaluasi dalam perspektif pendidikan Islam. Mahasiswa juga mengapresiasi pendekatan interaktif yang mendorong partisipasi aktif dan refleksi kritis dalam setiap sesi perkuliahan.

Hasil Evaluasi Dimensi Pembelajaran

Dimensi pembelajaran dalam kerangka Kirkpatrick mengukur sejauh mana peserta mengalami peningkatan pengetahuan, pemahaman, dan keterampilan sebagai hasil dari program pembelajaran yang diikuti. Temuan pada dimensi ini memperlihatkan peningkatan pemahaman yang signifikan terhadap konsep evaluasi program dan kebijakan pendidikan Islam, dengan rata-rata skor 4,31 dan persentase 86,30 persen pada kategori sangat baik.

Tabel 2. Hasil Evaluasi Dimensi Pembelajaran Mahasiswa

No	Indikator Pembelajaran	Mean	Persentase (%)	Kategori
1	Memahami konsep dasar evaluasi program pendidikan	4,39	87,78	Sangat Baik
2	Memahami konsep evaluasi kebijakan pendidikan Islam	4,39	87,78	Sangat Baik
3	Mampu membedakan evaluasi formatif dan sumatif	4,33	86,67	Sangat Baik
4	Memahami langkah-langkah menyusun desain evaluasi program	4,22	84,44	Sangat Baik
5	Mampu menganalisis studi kasus evaluasi pendidikan Islam	4,17	83,33	Baik
6	Merasa pengetahuan tentang evaluasi meningkat secara signifikan	4,39	87,78	Sangat Baik
	Rata-rata	4,31	86,30	Sangat Baik

Berdasarkan Tabel 2, tiga indikator memperoleh skor tertinggi secara bersamaan, yakni pemahaman konsep dasar evaluasi program pendidikan, pemahaman konsep evaluasi kebijakan pendidikan Islam, dan persepsi peningkatan pengetahuan secara signifikan, masing-masing mencapai skor 4,39 dengan persentase 87,78 persen. Ketiganya berada pada kategori sangat baik dan mengindikasikan bahwa perkuliahan berhasil membangun fondasi konseptual yang kokoh dalam diri mahasiswa, baik pada tataran pemahaman teoritis maupun kesadaran akan pertumbuhan pengetahuan yang mereka rasakan sendiri. Kemampuan membedakan evaluasi formatif dan sumatif memperoleh skor 4,33 dengan persentase 86,67 persen, menunjukkan bahwa mahasiswa telah menguasai perbedaan dua pendekatan evaluasi fundamental yang menjadi basis analisis dalam berbagai model evaluasi program.

Pemahaman langkah-langkah menyusun desain evaluasi program memperoleh skor 4,22 dengan persentase 84,44 persen. Meskipun skor ini sedikit lebih rendah dibandingkan indikator konseptual, hasilnya tetap berada pada kategori sangat baik dan mengindikasikan bahwa mahasiswa telah menguasai pengetahuan prosedural yang diperlukan untuk merancang evaluasi program secara sistematis. Indikator dengan skor terendah pada dimensi pembelajaran adalah kemampuan menganalisis studi kasus evaluasi pendidikan Islam, yakni 4,17 dengan persentase 83,33 persen, yang berada pada kategori baik. Meski selisihnya tidak terlalu besar, penurunan ini secara teoretis bermakna penting.

Pola skor ini mengungkap temuan yang secara teoretis penting: mahasiswa lebih mudah menguasai declarative knowledge (pemahaman konsep dan pengetahuan faktual) dibandingkan procedural knowledge (kemampuan aplikasi dan analisis berbasis kasus nyata). Kesenjangan antara penguasaan konseptual dan kemampuan analisis kasus ini sejalan dengan kritik terhadap pembelajaran evaluasi yang cenderung berhenti pada transfer konsep dan belum sepenuhnya mencapai tahap aplikasi kritis dalam situasi kompleks (Stufflebeam & Coryn, 2014). Dalam konteks doktoral—di mana kemampuan berpikir analitis, sintesis, dan evaluatif merupakan kompetensi utama—gap antara declarative dan procedural knowledge ini menjadi sinyal kurikuler yang perlu mendapat respons serius dalam perancangan pembelajaran

ke depan, bukan sekadar variasi statistik yang dapat diabaikan. Studi-studi sebelumnya yang menggunakan kerangka Kirkpatrick dalam konteks pelatihan aparatur (Ramadhon, 2016; Sopacua & Budijanto, 2007; Utomo & Tehupeiry, 2014) tidak melaporkan pola ini karena instrumen yang digunakan tidak membedakan dua level kompetensi kognitif tersebut secara eksplisit. Hal ini menjadi salah satu kontribusi metodologis penelitian ini dalam memperkaya penerapan kerangka Kirkpatrick pada konteks pendidikan tinggi. Mahasiswa juga menyatakan melalui pertanyaan terbuka bahwa perkuliahan membantu mereka mengaitkan teori evaluasi dengan praktik nyata dalam pendidikan Islam, termasuk mengalami transformasi cara pandang dalam memaknai evaluasi sebagai instrumen reflektif dan pengambilan keputusan kebijakan.

Hasil Evaluasi Dimensi Perilaku

Dimensi perilaku dalam kerangka Kirkpatrick mengukur sejauh mana peserta mengimplementasikan pengetahuan dan keterampilan yang diperoleh ke dalam perilaku nyata pascapembelajaran. Temuan pada dimensi ini memperlihatkan bahwa hasil pembelajaran telah mulai diterapkan dalam perilaku akademik mahasiswa, dengan rata-rata skor 4,07 dan persentase 81,48 persen pada kategori baik. Dimensi perilaku menjadi satu-satunya dimensi yang tidak mencapai kategori sangat baik, sebuah temuan yang memiliki implikasi penting bagi pengembangan perkuliahan.

Tabel 3. Hasil Evaluasi Dimensi Perilaku Mahasiswa

No	Indikator Perilaku	Mean	Persentase (%)	Kategori
1	Penerapan konsep evaluasi dalam tugas	4,39	87,78	Sangat Baik
2	Penerapan pendekatan evaluasi dalam analisis kebijakan pendidikan Islam	4,06	81,11	Baik
3	Penerapan menyusun desain evaluasi program pendidikan	3,67	73,33	Baik
4	Penerapan konsep evaluasi dalam penelitian atau rencana disertasi	4,11	82,22	Baik
5	Penerapan dalam mengkritisi program/kebijakan pendidikan secara evaluatif	3,83	76,67	Baik
6	Konsep evaluasi yang dipelajari memengaruhi cara berpikir akademik	4,39	87,78	Sangat Baik
Rata-rata		4,07	81,48	Baik

Berdasarkan Tabel 3, terdapat variasi yang lebih besar antara indikator pada dimensi perilaku dibandingkan dimensi sebelumnya. Dua indikator memperoleh skor tertinggi dan masuk kategori sangat baik, yakni penerapan konsep evaluasi dalam tugas akademik (4,39; 87,78%) dan pengaruh konsep evaluasi terhadap cara berpikir akademik (4,39; 87,78%). Kedua indikator ini menunjukkan bahwa perkuliahan telah berhasil membentuk pola pikir evaluatif mahasiswa dan mendorong mereka mengaplikasikan konsep yang dipelajari dalam penyelesaian tugas-tugas akademik. Khususnya, perubahan cara berpikir akademik merupakan indikator yang sangat bermakna dalam konteks doktoral, karena mencerminkan internalisasi nilai-nilai evaluatif yang melampaui sekadar penyelesaian tugas formal.

Penerapan konsep evaluasi dalam penelitian atau rencana disertasi memperoleh skor 4,11 dengan persentase 82,22 persen, mengindikasikan bahwa mahasiswa mulai mengintegrasikan perspektif evaluatif dalam perencanaan penelitian doktoral mereka—sebuah capaian yang strategis mengingat disertasi merupakan produk akademik utama jenjang doktoral. Penerapan pendekatan evaluasi dalam analisis kebijakan pendidikan Islam memperoleh skor 4,06 dengan persentase 81,11 persen, mencerminkan bahwa mahasiswa dapat menggunakan kerangka evaluatif untuk menganalisis kebijakan pendidikan secara lebih sistematis. Indikator penerapan dalam mengkritisi program atau kebijakan pendidikan secara evaluatif memperoleh skor 3,83 dengan persentase 76,67 persen, menandakan bahwa kemampuan evaluatif kritis terhadap kebijakan masih perlu penguatan lebih lanjut. Indikator dengan skor

terendah adalah penerapan menyusun desain evaluasi program pendidikan, yakni 3,67 dengan persentase 73,33 persen. Meskipun masih berada pada kategori baik, skor ini paling rendah pada seluruh dimensi perilaku dan mengindikasikan bahwa kompetensi teknis menyusun desain evaluasi—yang merupakan keterampilan praktis kompleks—belum terinternalisasi secara optimal dalam aktivitas akademik mahasiswa.

Penurunan skor pada dimensi perilaku dibandingkan dimensi reaksi dan pembelajaran bukan sekadar fenomena yang dapat dimaklumi begitu saja, melainkan perlu dicermati secara kritis sebagai temuan yang paling signifikan secara pedagogis. Kesenjangan antara skor pembelajaran (4,31) dan skor perilaku (4,07) mengonfirmasi adanya transfer gap—jarak antara penguasaan konseptual yang tinggi dengan implementasi praktis yang lebih rendah—sebuah fenomena yang dalam literatur pendidikan dikenal sebagai knowing-doing gap (Pfeffer & Salancik, 2003). Dalam konteks pendidikan doctoral yang berorientasi pada pembentukan peneliti dan pengambil kebijakan, gap ini memiliki implikasi serius: jika mahasiswa yang akan menjadi evaluator, peneliti kebijakan, dan pengambil keputusan tidak dapat secara efektif mentransfer pemahaman mereka ke dalam praktik evaluatif nyata, maka tujuan strategis mata kuliah belum sepenuhnya tercapai. Dua faktor utama dapat menjelaskan transfer gap ini: pertama, keterbatasan alokasi waktu perkuliahan untuk kegiatan praktik langsung dan simulasi evaluasi; kedua, minimnya scaffolding sistematis yang memandu mahasiswa dalam mengaplikasikan konsep ke dalam konteks penelitian dan kebijakan yang spesifik. Temuan ini sekaligus menjadi kritik konstruktif terhadap desain perkuliahan yang perlu memperbanyak komponen praktik berbasis proyek nyata dan studi kasus kontekstual.

Data kualitatif memberikan bukti yang lebih kaya dan kongkret mengenai transfer pembelajaran ini. Seorang mahasiswa menyatakan: "Setelah mengikuti perkuliahan ini, saya mulai menyusun kerangka evaluasi dalam proposal disertai saya dengan mengacu pada model Kirkpatrick, meskipun belum sempurna dalam implementasinya." Mahasiswa lain mengemukakan: "Saya lebih kritis sekarang dalam menilai program-program pendidikan Islam—saya tidak lagi menerima laporan evaluasi begitu saja tanpa mempertanyakan metodologi dan indikator yang digunakan." Pernyataan-pernyataan ini mengindikasikan bahwa internalisasi awal pemahaman evaluatif telah terjadi secara nyata, namun implementasi teknis—khususnya dalam menyusun desain evaluasi yang komprehensif—masih dalam tahap berkembang dan memerlukan dukungan pembelajaran yang lebih terstruktur.

Hasil Evaluasi Dimensi Hasil

Dimensi hasil dalam kerangka Kirkpatrick mengukur dampak akhir program pembelajaran terhadap peningkatan kapasitas, kinerja, dan kontribusi peserta pada level yang lebih luas. Temuan pada dimensi ini memperlihatkan bahwa perkuliahan memberikan dampak positif yang signifikan terhadap pengembangan kompetensi evaluatif mahasiswa doctoral, dengan rata-rata skor 4,38 dan persentase 87,56 persen pada kategori sangat baik.

Tabel 4. Hasil Evaluasi Dimensi Hasil Mahasiswa

No	Indikator Hasil	Mean	Persentase (%)	Kategori
1	Peningkatan kompetensi evaluatif	4,39	87,78	Sangat Baik
2	Kesiapan melakukan evaluasi program	4,33	86,67	Sangat Baik
3	Dampak terhadap kualitas penelitian	4,39	87,78	Sangat Baik
4	Kontribusi pada pengambilan kebijakan	4,39	87,78	Sangat Baik
5	Kontribusi pada pengembangan keilmuan	4,39	87,78	Sangat Baik
	Rata-rata	4,38	87,56	Sangat Baik

Berdasarkan Tabel 4, empat dari lima indikator dimensi hasil memperoleh skor yang sama tinggi, yakni 4,39 dengan persentase 87,78 persen, mencakup peningkatan kompetensi evaluatif, dampak terhadap kualitas penelitian, kontribusi pada pengambilan kebijakan, dan kontribusi pada pengembangan keilmuan.

Konsistensi skor yang tinggi pada keempat indikator ini mengindikasikan bahwa dampak perkuliahan bersifat menyeluruh dan tidak terbatas pada satu aspek saja, melainkan menjangkau dimensi kompetensi personal, kualitas penelitian, relevansi kebijakan, dan kontribusi keilmuan secara bersamaan. Peningkatan kompetensi evaluatif yang dirasakan mahasiswa mencerminkan keberhasilan perkuliahan dalam membangun kapasitas mendasar sebagai evaluator program dan kebijakan pendidikan Islam. Dampak positif terhadap kualitas penelitian mengindikasikan bahwa pemahaman evaluatif yang diperoleh telah memberikan pengaruh konkret terhadap cara mahasiswa merancang dan melaksanakan penelitian doctoral mereka. Kontribusi pada pengambilan kebijakan dan pengembangan keilmuan menunjukkan bahwa perkuliahan tidak hanya berdampak pada level individual, tetapi juga dirasakan mahasiswa sebagai bekal untuk berkontribusi pada level institusional dan akademik yang lebih luas.

Indikator kesiapan melakukan evaluasi program memperoleh skor sedikit lebih rendah, yakni 4,33 dengan persentase 86,67 persen, namun tetap berada pada kategori sangat baik. Skor yang sedikit lebih rendah pada indikator ini dapat dipahami sebagai cerminan kehati-hatian mahasiswa dalam menilai kesiapan praktis mereka—sebuah sikap kritis yang justru mencerminkan kematangan akademik, karena mereka menyadari bahwa kesiapan untuk melakukan evaluasi yang sesungguhnya memerlukan lebih dari sekadar pemahaman konseptual.

Temuan menarik pada dimensi ini adalah skor hasil (4,38) yang lebih tinggi dari skor perilaku (4,07), padahal secara logis hierarki Kirkpatrick mengasumsikan bahwa hasil merupakan produk kumulatif dari perubahan perilaku yang telah terkonsolidasi. Paradoks ini perlu dicermati secara kritis dan dapat dipahami melalui dua penjelasan yang saling melengkapi. Pertama, mahasiswa cenderung mempersepsikan "hasil" secara lebih aspirasional—yakni mengacu pada potensi dan kapasitas yang mereka yakini akan terwujud ke depan—dibandingkan perilaku aktual yang harus mereka demonstrasikan saat ini. Kedua, pengukuran berbasis persepsi diri memiliki keterbatasan inheren dalam membedakan antara kompetensi yang telah dimiliki secara aktual dan kompetensi yang diharapkan atau diantisipasi. Kondisi ini menggarisbawahi pentingnya triangulasi dengan penilaian pihak eksternal, observasi langsung, atau data kinerja jangka panjang dalam studi lanjutan agar gambaran mengenai dampak perkuliahan menjadi lebih akurat dan objektif.

Data kualitatif memberikan gambaran yang lebih konkret dan personal mengenai dampak perkuliahan. Seorang mahasiswa menyatakan: "Mata kuliah ini menjadi fondasi berpikir saya dalam menyusun bab metodologi disertasi—saya kini memahami bagaimana merancang evaluasi yang tidak hanya valid secara teknis, tetapi juga relevan dengan konteks pendidikan Islam." Mahasiswa lain mengungkapkan: "Saya merasa lebih siap untuk terlibat dalam penyusunan kebijakan di institusi saya karena saya kini mampu mengevaluasi program secara sistematis dan berbasis data." Pernyataan-pernyataan ini mengonfirmasi bahwa perkuliahan telah memberikan dampak yang bermakna tidak hanya pada tataran akademik, tetapi juga pada kesiapan praktis mahasiswa dalam menjalankan peran profesional mereka sebagai pendidik dan pengambil kebijakan.

Ringkasan Hasil Evaluasi Berdasarkan Kerangka Kirkpatrick

Secara keseluruhan, hasil evaluasi pada keempat dimensi Kirkpatrick dapat dirangkum dalam tabel berikut untuk memberikan gambaran komprehensif tentang efektivitas perkuliahan.

Tabel 5. Hasil Statistik Deskriptif Evaluasi Berdasarkan Kerangka Kirkpatrick

Dimensi Evaluasi	Jumlah Item	Rata-rata (Mean)	Persentase (%)	Interpretasi
Reaksi	8	4,63	92,50	Sangat Baik
Pembelajaran	6	4,31	86,30	Sangat Baik
Perilaku	6	4,07	81,48	Baik

Hasil	5	4,38	87,56	Sangat Baik
-------	---	------	-------	-------------

Tabel 5 merangkum capaian keempat dimensi evaluasi Kirkpatrick secara komprehensif. Tiga dari empat dimensi berada pada kategori sangat baik, dengan dimensi reaksi memperoleh nilai tertinggi (4,63; 92,50%), diikuti dimensi hasil (4,38; 87,56%) dan dimensi pembelajaran (4,31; 86,30%). Dimensi perilaku sebagai satu-satunya dimensi berkategori baik (4,07; 81,48%) menjadi titik kritis yang memerlukan perhatian khusus dalam pengembangan perkuliahan ke depan.

Pola temuan secara keseluruhan mengonfirmasi relevansi hierarki Kirkpatrick: respons positif pada dimensi reaksi mendukung pencapaian pembelajaran yang signifikan, yang pada gilirannya mendorong—meskipun tidak secara linear dan otomatis—transformasi perilaku akademik dan dampak akhir yang bermakna. Transfer gap yang teridentifikasi antara dimensi pembelajaran dan perilaku mempertegas bahwa pengetahuan tidak secara otomatis bertransformasi menjadi perilaku, dan bahwa investasi pedagogis tambahan diperlukan untuk menjembatani gap tersebut.

Secara kritis, penelitian ini menempatkan diri berbeda dari studi-studi sebelumnya dalam dua hal mendasar. Pertama, konteks doktoral PTKI menghadirkan peserta yang bukan sekadar penerima pelatihan melainkan akademisi kritis dengan agenda penelitian aktif—kondisi ini mengubah makna dan bobot setiap dimensi Kirkpatrick secara substantif. Mahasiswa doktoral tidak hanya mengevaluasi perkuliahan dari perspektif kepuasan belajar, tetapi juga dari relevansinya terhadap proyek penelitian dan karier akademik mereka. Kedua, dimensi keislaman dalam pembelajaran menciptakan lapisan pemaknaan tambahan yang tidak ditangkap oleh instrumen evaluasi konvensional yang digunakan dalam studi-studi pelatihan aparatur terdahulu.

Kontribusi ilmiah penelitian ini bersifat ganda. Secara teoretis, penelitian ini memvalidasi dan mengekspansi penerapan kerangka Kirkpatrick ke dalam konteks PTKI doktoral yang sebelumnya belum tereksplorasi dalam literatur. Temuan ini menunjukkan bahwa model Kirkpatrick tidak hanya relevan untuk pelatihan organisasi, tetapi juga mampu mengungkap nuansa pembelajaran akademik tingkat lanjut—termasuk pola declarative-procedural knowledge gap dan knowing-doing gap—yang tidak dapat ditangkap oleh evaluasi parsial berbasis kepuasan mahasiswa semata. Secara praktis, penelitian ini menyediakan profil empiris multidimensi yang dapat menjadi baseline bagi pengembangan kurikulum, desain pembelajaran, dan sistem penjaminan mutu internal di lingkungan PTKI.

KESIMPULAN

Penelitian ini bertujuan menilai efektivitas perkuliahan Evaluasi Program dan Kebijakan Pendidikan Islam di UIN Alauddin Makassar menggunakan kerangka Kirkpatrick. Secara keseluruhan, perkuliahan terbukti efektif pada keempat dimensi evaluasi, dengan tiga dimensi berkategori sangat baik (reaksi: $M=4,63$; pembelajaran: $M=4,31$; hasil: $M=4,38$) dan satu dimensi berkategori baik (perilaku: $M=4,07$).

Pada dimensi reaksi, mahasiswa memberikan tanggapan sangat positif terhadap kejelasan materi, relevansinya dengan kebutuhan doktoral, kompetensi dosen, dan metode pembelajaran partisipatif. Skor tinggi yang dihasilkan oleh akademisi berpengalaman ini memiliki bobot validitas yang lebih kuat dibandingkan konteks evaluasi pelatihan konvensional. Pada dimensi pembelajaran, perkuliahan berhasil meningkatkan penguasaan konsep evaluasi secara signifikan. Namun, teridentifikasi gap antara declarative knowledge (pemahaman konsep) dan procedural knowledge (kemampuan analisis kasus) yang menjadi sinyal kurikuler penting bagi pengembangan desain pembelajaran. Pada dimensi perilaku, internalisasi awal konsep evaluasi dalam aktivitas akademik telah terjadi, dibuktikan melalui data kualitatif yang kaya. Namun, transfer gap antara pemahaman konseptual dan implementasi teknis—khususnya dalam penyusunan desain evaluasi program—mengindikasikan perlunya penguatan komponen praktik berbasis proyek nyata secara sistematis. Pada dimensi

hasil, perkuliahan memberikan dampak positif yang menyeluruh terhadap kompetensi evaluatif, kualitas penelitian, dan kesiapan mahasiswa sebagai evaluator dan pengambil kebijakan pendidikan Islam.

Kontribusi ilmiah penelitian ini mencakup dua dimensi. Secara teoretis, penelitian ini memvalidasi dan mengekspansi penerapan kerangka Kirkpatrick ke konteks PTKI doktoral yang sebelumnya belum tereksplorasi, sekaligus mengidentifikasi pola knowing-doing gap dan declarative-procedural knowledge gap yang belum dilaporkan dalam studi evaluasi pelatihan sebelumnya. Secara praktis, penelitian ini menyediakan profil empiris multidimensi yang dapat menjadi landasan pengembangan kurikulum dan sistem penjaminan mutu akademik di lingkungan PTKI.

Berdasarkan temuan, disarankan agar: (1) dosen memperbanyak kegiatan praktik langsung, simulasi evaluasi, dan studi kasus kontekstual pendidikan Islam untuk menutup transfer gap pada dimensi perilaku; (2) program studi mengintegrasikan penilaian berbasis kinerja evaluatif nyata dalam sistem asesmen mata kuliah. Kajian selanjutnya disarankan menggunakan desain longitudinal, pengukuran multi-sumber (self-report, peer assessment, dan penilaian dosen), serta pendekatan inferensial untuk menguji hubungan kausal antardimensi Kirkpatrick dalam konteks PTKI secara lebih mendalam.

Keterbatasan penelitian ini mencakup: (1) desain deskriptif yang tidak dapat menjelaskan hubungan kausal antardimensi secara statistik; (2) sampel terbatas (n=18) dari satu mata kuliah pada satu institusi yang membatasi generalisasi temuan; (3) pengukuran berbasis persepsi diri yang rentan terhadap bias sosial-desirabilitas, khususnya pada dimensi perilaku dan hasil; serta (4) pelaksanaan perkuliahan yang sebagian dilakukan secara daring yang berpotensi memengaruhi pengalaman belajar dan tingkat penerapan konsep evaluatif.

DAFTAR RUJUKAN

- Anshari, H. (2014). *Pengantar Ilmu Pendidikan*. Usaha Nasional.
- Creswell, J. W. (2012). *Pendekatan Kualitatif, Kuantitatif, dan Mixed (Penerjemah :Achmad Fawaid)*. Pustaka Pelajar.
<http://kin.perpusnas.go.id/DisplayData.aspx?pld=42264&pRegionCode=JIUNM AL&pClientId=111>
- Fitriyyah, R. (2025). Manajemen Sistem Penilaian Yang Efektif Dalam Meningkatkan Mutu Pendidikan Islam. *Andragogi: Jurnal Pendidikan Islam Dan Manajemen Pendidikan Islam*, 7(1), 66–85.
<https://jurnalptiq.id/index.php/andragogi/article/view/45>
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program Evaluation: Alternative Approaches and Practical Guidelines*. Pearson Education.
- Kirkpatrick, D., & Kirkpatrick, J. (2006). *Evaluating Training Programs: The Four Levels*. Berrett-Koehler Publishers.
- Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's Four Levels of Training Evaluation*. Association for Talent Development.
- Kurniawan, W., Sriwahyuni, T., & Zen, B. Y. (2025). Revitalisasi Pendidikan Agama Islam di Perguruan Tinggi: Membangun Mahasiswa Yang Intellektual dan Spiritual. *Reflection: Islamic Education Journal*, 2(2), 227–239.
<https://doi.org/10.61132/reflection.v2i2.835>
- Miles, M. B., Huberman, A. M., & Saldana, J. (2013). *Qualitative Data Analysis: A Methods Sourcebook*. SAGE Publications.
- Pfeffer, J., & Salancik, G. R. (2003). *The External Control of Organizations: A Resource Dependence Perspective*. Stanford University Press.
- Ramadhon, S. (2016). Penerapan Model Empat Level Kirkpatrick Dalam Evaluasi Program Pendidikan Dan Pelatihan Aparatur Di Pusdiklat Migas. *Swara Patra : Majalah Ilmiah PPSDM Migas*, 6(1).
<https://ejurnal.ppsdmmigas.esdm.go.id/sp/index.php/swarapatra/article/view/101>

- Sopacua, E., & Budijanto, D. (2007). Evaluasi 4 Tahap dari Kirkpatrick sebagai Alat dalam Evaluasi Pasca Pelatihan. *Buletin Penelitian Sistem Kesehatan*, 10(4), 21049. <https://www.neliti.com/publications/21049/>
- Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation Theory, Models, and Applications*. John Wiley & Sons.
- Sudradjat, A., & Alim, A. (2025). Membangun Efektivitas Pembelajaran Pendidikan Agama Islam Pada Mahasiswa Di Perguruan Tinggi Umum. *Pendas : Jurnal Ilmiah Pendidikan Dasar*, 10(03), 217–239. <https://doi.org/10.23969/jp.v10i03.32633>
- Sukardi. (2022). *Evaluasi Program Pendidikan dan Kepelatihan*. Bumi Aksara.
- Taali, M., Darmawan, A., & Maduwinarti, A. (2024). *Teori dan Model Evaluasi Kebijakan: Kajian kebijakan kurikulum pendidikan*. PT. Sonpedia Publishing Indonesia.
- Utomo, A. P., & Tehupeiry, K. P. (2014). Evaluasi Pelatihan dengan Metode Kirkpatrick Analysis. *Jurnal Telematika*, 9(2), 37–37. <https://doi.org/10.61769/telematika.v9i2.87>