

BAB III

METODE PENELITIAN

A. Rancangan Penelitian

Bab ini menguraikan seluruh perangkat metodologis yang dipakai untuk menjawab rumusan masalah penelitian, mulai dari paradigma yang mendasari, rancangan eksperimen yang dipilih, populasi dan sampel, prosedur perlakuan, instrumen pengukuran, hingga teknik analisis data. Uraian disusun tidak semata sebagai daftar langkah teknis, melainkan sebagai argumentasi metodologis: setiap keputusan yang diambil disertai alasan mengapa keputusan itu diambil dan apa konsekuensinya bagi keabsahan simpulan. Susunan bab ini sengaja dirancang sejajar dengan alur penyajian hasil pada Bab IV, sehingga setiap prosedur yang ditetapkan di sini memiliki pasangan sajian datanya di sana, dan sebaliknya setiap angka yang dilaporkan pada Bab IV dapat dilacak kembali pada prosedur yang telah ditetapkan di bab ini.¹

1. Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan landasan paradigma post-positivisme. Paradigma ini berpijak pada realisme kritis, yakni keyakinan bahwa realitas objektif memang ada, tetapi hanya dapat dikenali secara tidak sempurna dan bersifat probabilistik karena keterbatasan alat pengetahuan manusia dan kompleksitas fenomena sosial-keagamaan yang diteliti.² Dalam konteks pengukuran literasi keagamaan mahasiswa, paradigma post-positivisme mengarahkan peneliti untuk menerjemahkan gejala pemahaman, penafsiran, dan kecenderungan tindakan yang pada dasarnya abstrak menjadi indikator empiris yang terukur melalui instrumen terstandarisasi, dengan tetap menyadari bahwa hasil pengukuran itu merupakan pendekatan terhadap realitas, bukan cermin sempurna atas realitas itu sendiri.

¹ Uraian mengenai pentingnya keterlacakan prosedur metodologis terhadap sajian data hasil dapat dilihat pada John W. Creswell dan J. David Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, ed. ke-5 (Thousand Oaks, CA: Sage Publications, 2018), 147-158.

² D. C. Phillips dan Nicholas C. Burbules, *Postpositivism and Educational Research* (Lanham, MD: Rowman and Littlefield, 2000), 26-41.

Konsekuensi dari pilihan paradigmatik tersebut adalah cara kerja yang deduktif. Penelitian ini bergerak dari telaah teori pada Bab II untuk merumuskan hipotesis kausal, kemudian menguji hipotesis tersebut secara empiris melalui desain eksperimen semu, dan berakhir pada penarikan generalisasi terbatas dengan dukungan bukti statistik.³ Pilihan pendekatan kuantitatif dipandang paling tepat karena pertanyaan pokok penelitian ini bukan pertanyaan tentang makna subjektif yang dihayati mahasiswa, melainkan pertanyaan tentang besaran pengaruh: seberapa besar dua strategi media pembelajaran yang berbeda mampu meningkatkan literasi keagamaan mahasiswa, dan apakah selisih pengaruh di antara keduanya bermakna secara statistik. Pertanyaan berjenis demikian menuntut perbandingan skor antarkelompok pada sejumlah subjek yang memadai, dan hanya dapat dijawab melalui prosedur pengukuran dan pengujian inferensial.

2. Kedudukan Model ADDIE sebagai Prosedur Pra-Eksperimen

Penelitian ini melibatkan dua produk yang dikembangkan sendiri oleh peneliti, yaitu Modul Digital PAI *Islam Rahmatan Lil Alamin* dan Panduan *Project Video Based Learning* (PjBL). Karena itu, sebelum eksperimen dilaksanakan, penelitian ini mengintegrasikan prosedur pengembangan instruksional model ADDIE, yang mencakup tahap *Analysis*, *Design*, *Development*, *Implementation*, dan *Evaluation*, untuk menghasilkan perangkat stimulus pembelajaran yang layak digunakan.⁴ Perlu ditegaskan sejak awal bahwa model ADDIE dalam penelitian ini diposisikan sebagai prosedur teknis pra-eksperimen untuk menyiapkan dan menstandarkan mutu stimulus instruksional, bukan sebagai paradigma penelitian dan pengembangan yang berorientasi pada siklus penyempurnaan produk jangka panjang. Penegasan ini penting agar tidak terjadi salah baca terhadap kedudukan bab hasil, sebab yang diuji secara statistik pada Bab IV adalah efektivitas perlakuan, bukan kematangan produk.

³ Fred N. Kerlinger dan Howard B. Lee, *Foundations of Behavioral Research*, ed. ke-4 (Fort Worth, TX: Harcourt College Publishers, 2000), 17-25.

⁴ Robert Maribe Branch, *Instructional Design: The ADDIE Approach* (New York: Springer, 2009), 1-16.

Kerangka kelayakan produk yang dipakai mengacu pada tiga kriteria yang dirumuskan Plomp dan Nieveen, yaitu validitas, kepraktisan, dan keefektifan.⁵ Pembagian tugas antarbagian penelitian ini mengikuti pembagian ketiga kriteria tersebut. Validasi ahli pada fase pra-eksperimen ditujukan untuk menegakkan kriteria pertama, yaitu validitas produk. Kriteria kepraktisan ditegakkan melalui pengamatan keterlaksanaan pembelajaran oleh pengamat independen selama perlakuan berjalan. Adapun kriteria keefektifan tidak ditegakkan pada fase pengembangan, melainkan diuji secara terpisah melalui perbandingan skor *pretest* dan *posttest* dengan statistik inferensial. Dengan pembagian yang tegas ini, pernyataan sangat layak yang diperoleh dari validasi ahli tidak boleh dan tidak akan dibaca sebagai bukti efektivitas.

Guna menjaga ketelitian metodologis, peneliti menetapkan batas operasional yang tegas antara fase pengembangan dan fase pengujian. Model ADDIE diisolasi sepenuhnya pada fase persiapan, yang mencakup analisis kebutuhan materi Pendidikan Agama Islam, perancangan arsitektur digital, serta penyusunan draf Modul Digital dan Panduan Video PjBL, yang selanjutnya dinilai secara formatif oleh panel ahli melalui prosedur *expert judgment*. Setelah seluruh produk dinyatakan layak berdasarkan penilaian ahli, seluruh fitur media dikunci secara total. Selama empat pertemuan perlakuan berlangsung tidak diperkenankan adanya revisi teks, perubahan tata letak, penambahan animasi, maupun modifikasi alur interaktivitas. Penguncian ini bukan formalitas administratif, melainkan syarat validitas internal: hanya dengan produk yang dikunci, perubahan skor literasi keagamaan pada *posttest* dapat diatribusikan pada karakteristik masing-masing strategi media, bukan pada penyempurnaan produk yang berlangsung di tengah perlakuan.⁶ Batas operasional fungsional antara kedua kerangka kerja tersebut dijabarkan pada tabel dibawah ini :

Tabel 3. 1 Matriks Batas Fungsional Model ADDIE dan Desain Quasi-Eksperimen

⁵ Tjeerd Plomp dan Nienke Nieveen, ed., *An Introduction to Educational Design Research* (Enschede: SLO Netherlands Institute for Curriculum Development, 2013), 26-30.

⁶ William R. Shadish, Thomas D. Cook, dan Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Boston: Houghton Mifflin, 2002), 54-63.

Parameter Batas	Fase R&D: Model ADDIE (Persiapan Stimulus)	Fase Eksperimen Semu: Pretest-Posttest Balanced Design (Pengujian Efektivitas)
Fungsi Operasional	Merancang, membuat, dan memvalidasi kelayakan teknis Modul Digital dan Panduan Video PjBL.	Menguji dampak kausalitas (<i>causal-effect</i>) penggunaan media terhadap peningkatan literasi keagamaan mahasiswa.
Subjek yang Terlibat	Panel Ahli / <i>Expert Judgment</i> (Validator Materi, Media, Bahasa) dan sampel kecil untuk <i>try-out</i> instrumen.	Sampel riil penelitian (N = 300, dibagi rata n = 100 per kelompok dalam 6 kelas paralel).
Status Media Pembelajaran	Bersifat dinamis, mengalami iterasi, revisi, dan penyempurnaan berdasarkan catatan komparatif validator.	Bersifat statis dan dikunci (<i>locked</i>). Tidak boleh ada revisi media selama 4 pertemuan intervensi berjalan.
Kriteria kelayakan yang ditegakkan	Validitas produk menurut kerangka Plomp dan Nieveen.	Keefektifan produk menurut kerangka Plomp dan Nieveen.
Bentuk luaran analisis	Rerata skor penilaian ahli pada skala empat dan kategori kelayakan.	Nilai signifikansi uji non-parametrik, skor N-Gain, dan besaran pengaruh r.
Letak sajian pada Bab IV	Subbab A tentang Validitas Instrumen dan Produk.	Subbab B dan C tentang Deskripsi Hasil serta Analisis Uji Statistik.

3. Desain Penelitian Quasi Eksperimen (*Pretest-Posttest Control Group Design*)

Penelitian instruksional pada ranah pendidikan tinggi sering kali dihadapkan pada kendala etis, administratif, dan praktis yang membuat implementasi desain eksperimen murni (*True Experimental*) tidak mungkin dilaksanakan. Desain eksperimen murni menuntut adanya penunjukan acak individu (*Random Assignment*) untuk masuk ke dalam kelompok eksperimen maupun kontrol.

Dalam konteks institusional di Universitas Nusantara PGRI (UNP) Kediri, membubarkan kelas-kelas reguler yang sudah terbentuk secara administratif demi kepentingan acak individu akan merusak ekosistem perkuliahan, melanggar regulasi birokrasi akademik, dan memicu penolakan psikologis dari subjek

penelitian yang merasa diisolasi secara paksa dari kelompok belajarnya. Oleh karena itu, penelitian ini secara metodologis menggunakan Desain *Quasi-Experimental* (Eksperimen Semu), secara spesifik menerapkan rancangan *Pretest-Posttest Design* yang direkonstruksi menjadi struktur seimbang (*Balanced Design*).

Kelebihan utama dari desain kuasi-eksperimen ini adalah peneliti dapat menguji efektivitas intervensi dalam situasi dunia nyata (*real-world setting*) tanpa kehilangan kontrol ilmiah. Peneliti tidak memanipulasi susunan individu di dalam kelas, melainkan memanipulasi *stimulus metode pembelajaran* pada kelas-kelas utuh (*intact classes*) yang sudah tersedia secara natural di universitas, namun dengan menerapkan kontrol ketat pada karakteristik awal sampel melalui prosedur seleksi purposif. Struktur formal rancangan eksperimen ini secara matematis diwujudkan melalui skema berikut:

Kelompok Eksperimen 1	→	O_1	→	X_1	→	O_2
Kelompok Eksperimen 2	→	O_3	→	X_2	→	O_4
Kelompok Kontrol	→	O_5	→	—	→	O_6

Keterangan skema tersebut adalah sebagai berikut. Notasi O_1 , O_3 , dan O_5 menunjukkan pengukuran awal literasi keagamaan menggunakan *Parallel Form A*, yang dilaksanakan serentak pada akhir pertemuan keempat. Notasi O_2 , O_4 , dan O_6 menunjukkan pengukuran akhir menggunakan *Parallel Form B*, yang dilaksanakan serentak pada akhir pertemuan kedelapan. Notasi X_1 menunjukkan stimulus perlakuan berupa pemanfaatan Modul Digital interaktif secara mandiri terbimbing, sedangkan X_2 menunjukkan stimulus perlakuan berupa penerapan *Video Project Based Learning* secara berkelompok. Adapun tanda hubung pada kelompok kontrol menunjukkan pembelajaran konvensional berbasis ceramah bervariasi dan diskusi kelas tanpa penambahan stimulus media baru.

Satu penegasan mengenai unit analisis perlu dikemukakan pada bagian ini karena berimplikasi langsung pada penafsiran hasil. Kelompok Eksperimen 1 belajar secara individual melalui modul, sedangkan Kelompok Eksperimen 2 belajar secara kolaboratif melalui proyek video. Perbedaan moda interaksi tersebut bukanlah cacat metodologis, melainkan konsekuensi teoretis yang melekat pada kedua strategi yang justru sedang dibandingkan, yaitu belajar mandiri terarah di

satu sisi dan konstruktivisme sosial melalui bantuan antarsejawat di sisi lain.⁷ Agar perbedaan moda interaksi itu tidak berubah menjadi bias pengukuran, prosedur pengambilan data evaluasi dikunci pada level individu untuk seluruh kelompok. Setiap mahasiswa pada ketiga kelompok mengerjakan dua puluh butir soal secara mandiri tanpa diskusi maupun kerja sama, dengan durasi yang sama, yaitu empat puluh lima menit. Dengan pemisahan tegas antara format perlakuan yang boleh berbeda dan format evaluasi yang wajib seragam, varians skor yang dilaporkan pada Bab IV mencerminkan kapasitas literasi keagamaan tiap subjek, bukan efek dari format kerja kelompok.

4. Hipotesis Penelitian

Berdasarkan kerangka teoretis pada Bab II, penelitian ini menguji tiga hipotesis. Ketiga hipotesis tersebut dirumuskan atas skor komposit literasi keagamaan, yakni satu skor tunggal yang dihasilkan tes literasi keagamaan dan mencakup ketiga dimensi secara sekaligus, bukan atas skor terpisah per dimensi. Perumusan demikian dipilih secara sadar agar cakupan klaim penelitian tidak melampaui data yang benar-benar dikumpulkan.

1. Hipotesis pertama menyatakan bahwa pembelajaran menggunakan Modul Digital PAI *Islam Rahmatan Lil Alamin* berpengaruh signifikan terhadap peningkatan skor komposit literasi keagamaan mahasiswa. Hipotesis nol menyatakan tidak terdapat perbedaan yang signifikan antara skor *pretest* dan *posttest* pada kelompok Modul Digital.
2. Hipotesis kedua menyatakan bahwa pembelajaran menggunakan Video *Project Based Learning* berpengaruh signifikan terhadap peningkatan skor komposit literasi keagamaan mahasiswa. Hipotesis nol menyatakan tidak terdapat perbedaan yang signifikan antara skor *pretest* dan *posttest* pada kelompok Video PjBL.
3. Hipotesis ketiga menyatakan bahwa terdapat perbedaan efektivitas yang signifikan antara pembelajaran menggunakan Modul Digital, pembelajaran

⁷ Barry J. Zimmerman, "Becoming a Self-Regulated Learner: An Overview," *Theory into Practice* 41, no. 2 (2002): 64-70, Lev S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes* (Cambridge, MA: Harvard University Press, 1978), 84-91.

menggunakan Video PjBL, dan pembelajaran konvensional dalam meningkatkan skor komposit literasi keagamaan mahasiswa. Hipotesis nol menyatakan tidak terdapat perbedaan efektivitas yang signifikan di antara ketiga kelompok tersebut.

Ketiga hipotesis diuji pada taraf signifikansi lima persen. Hipotesis nol ditolak apabila nilai signifikansi asimtotik yang dihasilkan uji statistik lebih kecil dari 0,05. Pemetaan lengkap antara masing-masing hipotesis, jenis data, uji statistik yang dipakai, dan letak sajian hasilnya pada Bab IV disajikan pada bagian akhir bab ini.

5. Lokasi dan Waktu Penelitian

Penelitian ini dilaksanakan di lingkungan kampus Universitas Nusantara PGRI Kediri dengan subjek mahasiswa yang menempuh Mata Kuliah Wajib Kurikulum Pendidikan Agama Islam. Pemilihan lokasi didasarkan pada tiga pertimbangan. Pertama, kesiapan dan pemerataan infrastruktur teknologi informasi kampus, yang merupakan prasyarat mutlak bagi penerapan modul digital dan produksi video. Kedua, keseragaman karakteristik administratif dan kurikuler mahasiswa peserta mata kuliah, yang memungkinkan diperolehnya kelas-kelas yang sebanding. Ketiga, adanya komitmen kelembagaan terhadap pengembangan inovasi pembelajaran digital, yang memberi ruang bagi pelaksanaan eksperimen tanpa mengganggu tertib administrasi akademik.

Pelaksanaan penelitian berlangsung selama satu semester, yaitu dari 5 September 2025 sampai dengan 20 Januari 2026. Rentang waktu tersebut terbagi ke dalam dua fase. Fase pra-eksperimen berlangsung sekitar dua bulan dan meliputi analisis kebutuhan, pengembangan produk, validasi ahli, serta uji coba instrumen pada subjek di luar sampel. Fase eksperimen inti berlangsung selama delapan pertemuan terjadwal yang berjalan paralel pada kalender akademik, dengan durasi sembilan puluh menit setiap pertemuan.

B. Populasi dan Sampel

1. Populasi Penelitian

Populasi adalah keseluruhan subjek yang memiliki karakteristik tertentu yang ditetapkan peneliti sebagai dasar penarikan sampel dan sasaran generalisasi

simpulan.⁸ Populasi dalam penelitian ini adalah seluruh mahasiswa program sarjana di Universitas Nusantara PGRI Kediri yang terdaftar secara resmi menempuh Mata Kuliah Wajib Kurikulum Pendidikan Agama Islam pada Tahun Akademik 2025/2026. Berdasarkan data administrasi akademik, populasi terjangkau tercatat sebanyak 47 kelas paralel dengan sebaran 37 sampai 40 mahasiswa per kelas, sehingga jumlah keseluruhan mencapai sekitar 1.900 mahasiswa.⁹ Karakteristik populasi ini heterogen dari sisi program studi asal, namun seragam dari sisi rentang usia perkembangan dewasa awal dan keterikatan pada kurikulum universitas yang sama.

2. Teknik Sampling

Dalam pelaksanaan penelitian pendidikan di perguruan tinggi, pembagian kelas umumnya telah ditetapkan oleh institusi sehingga peneliti tidak dapat melakukan pengacakan subjek secara bebas. Kondisi ini menyebabkan penggunaan teknik sampling probabilitas menjadi kurang memungkinkan. Kondisi tersebut menjadikan teknik *probability sampling* tidak dapat diterapkan. Oleh karena itu penelitian ini menggunakan *non-probability sampling*, secara khusus *purposive sampling*, yang diterapkan pada level kelas utuh dan bukan pada level individu.¹⁰ Konsekuensi dari pilihan ini disadari sepenuhnya: karena penunjukan acak tidak dilakukan, kesetaraan kelompok tidak dapat diasumsikan begitu saja, melainkan harus dibuktikan secara empiris melalui skor *pretest*.

Penetapan ukuran sampel tidak ditentukan secara subjektif, melainkan dihitung melalui analisis kekuatan statistik atau *a priori power analysis* menggunakan perangkat lunak *GPower* dengan parameter konseptual Cohen.¹¹ Parameter yang dimasukkan adalah sebagai berikut.

1. Besaran efek yang diharapkan ditetapkan sebesar 0,25, yang merepresentasikan efek sedang menurut baku Cohen untuk penelitian instruksional. Penetapan ini

⁸ Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan R&D* (Bandung: Alfabeta, 2019), 126.

⁹ Data administrasi akademik Universitas Nusantara PGRI Kediri, Tahun Akademik 2025/2026.

¹⁰ Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*, 133-134.

¹¹ Franz Faul dkk., "GPower 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences," *Behavior Research Methods** 39, no. 2 (2007): 175-191.

berpijak pada asumsi realistis bahwa intervensi media digital pada perguruan tinggi memberikan dampak yang moderat namun bermakna secara praktis.¹²

2. Taraf signifikansi ditetapkan sebesar 0,05, yang berarti peneliti menoleransi risiko kesalahan Tipe I paling besar lima persen.

3. Kekuatan uji ditetapkan secara restriktif pada 0,95, sehingga peluang terjadinya kesalahan Tipe II, yakni gagal mendeteksi efektivitas media yang sesungguhnya nyata, ditekan hingga lima persen.

4. Jumlah kelompok ditetapkan sebanyak tiga, yaitu Eksperimen 1, Eksperimen 2, dan Kontrol.

Berdasarkan parameter tersebut, *A Priori Power Analysis* untuk pengujian komparatif tiga kelompok independen tersebut, perangkat lunak *GPower* menetapkan bahwa batas minimum ukuran sampel kumulatif yang wajib dipenuhi untuk mencapai power 95% adalah 252 responden, atau setara dengan ≈ 84 responden per kelompok. Berdasarkan ketetapan rumus *GPower* tersebut, peneliti membulatkan ukuran sampel secara presisi demi memenuhi prinsip desain seimbang (*Balanced Design*) / *attrition strategy* (20%) yaitu 16 orang, menjadi $n = 100$ responden per kelompok, sehingga diperoleh ukuran sampel total kumulatif riil dalam riset ini sebesar $N = 300$ mahasiswa. Angka inilah yang dipakai secara konsisten pada seluruh analisis di Bab IV.

Penarikan unit sampel dilakukan secara non-probabilistik melalui metode *Purposive Sampling* (sampling bertujuan) yang diaplikasikan pada level penentuan kelas utuh (*intact classes*). Langkah-langkah penyaringan dijalankan secara sistematis menggunakan kriteria inklusi operasional untuk mendapatkan kelas-kelas yang memiliki karakteristik homogen dan setara untuk diperbandingkan (*comparable*):

- a. Kriteria Inclusion 1 (Keseragaman Kurikulum): Kelas-kelas yang dipilih wajib menggunakan silabus, materi pokok, dan draf Rencana Pembelajaran Semester (RPS) PAI yang seragam di bawah naungan koordinator MKWK yang sama.

¹² Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences*, ed. ke-2 (Hillsdale, NJ: Lawrence Erlbaum Associates, 1988), 284-288.

- b. Kriteria Inclusion 2 (Homogenitas Bebas SKS): Mahasiswa di dalam kelas yang dipilih wajib memiliki beban studi aktif semesteran dalam rentang normal (18-24 SKS) untuk meminimalkan variabel pengganggu berupa bias kelelahan kognitif. Kelas khusus karyawan dikeluarkan dari daftar seleksi.
- c. Kriteria Inclusion 3 (Standardisasi Instruksional Dosen): Kelas-kelas yang dipilih harus diampu oleh dosen yang sama atau tim dosen (*team teaching*) yang memiliki kualifikasi pedagogis dan standar rubrik penilaian yang identik, guna mengeliminasi bias *Teacher Effect*.
- d. Kriteria Inclusion 4 (Homogenitas Kemampuan Awal): Mahasiswa wajib memiliki kesetaraan tingkat kemampuan dasar keagamaan awal, yang dibuktikan secara empiris melalui nilai *pretest* yang menghasilkan nilai $p > 0,05$ pada uji homogenitas varians.

3. Sampel Penelitian

Regulasi akademik Universitas Nusantara PGRI (UNP) Kediri mengenai pembatasan kapasitas fisik ruang kuliah maksimal 40 mahasiswa per kelas, ukuran sampel seimbang $n=100$ untuk tiap-tiap kelompok diwujudkan melalui mekanisme Penggabungan Tiga Kelas Homogen (*Triple Intact Class Pooling*). Peneliti memilih 9 kelas paralel yang masing-masing berisi 37 hingga 40 mahasiswa dan lolos kriteria inclusion, kemudian mengelompokkannya menjadi 3 entitas kelompok besar.

Peneliti mengakui bahwa dalam pelaksanaan eksperimen ekologis di lapangan, jumlah manifes mahasiswa asli bersifat asimetris dan tidak bulat. Data mahasiswa yang mengalami kendala kehadiran atau administratif dieliminasi secara ketat demi menjaga validitas internal. Peneliti menyusun kriteria *exclusion* yang komprehensif untuk menyaring subjek yang tidak memenuhi kualifikasi paparan eksperimen penuh, meliputi: (a) Mahasiswa tidak hadir karena alasan sakit atau izin resmi kedinasan/organisasi kampus selama rentang intervensi (pertemuan 5-8), (b) Mahasiswa terlambat mengisi atau mengosongkan lembar pengerjaan instrumen

evaluasi (*pretest* atau *posttest*), (c) Lembar jawaban mengalami kerusakan fisik atau cacat pengisian yang menyebabkan kesalahan pembacaan sistem skoring.

Melalui penerapan kriteria *exclusion* yang diperketat ini, sisa data kotor yang masih menyisakan sedikit kelebihan (selisih tipis 2-3 data responden di atas angka 100) kemudian disamakan secara presisi menggunakan prosedur penyeimbangan acak komputer (*computerized random equalization*) demi memperoleh angka seimbang mutlak $n=100$. Alur penyusutan data kotor dari populasi terjangkau menuju sampel bersih disajikan secara transparan pada Tabel 3.3:

Tabel 3. 2 Matriks Alur Filtrasi, Eksklusi, dan Penyusutan Sampel Penelitian

Metode Penilaian Sampel	Kelompok Eksperimen 1 (Kelas A + B + C)	Kelompok Eksperimen 2 (Kelas D + E + F)	Kelompok Kontrol (Kelas G + H + I)	Total Kumulatif Data
1. Jumlah mahasiswa terdaftar	Kelas A = 38 Kelas B = 40 Kelas C = 39 Total = 117	Kelas D = 40 Kelas E = 38 Kelas F = 40 Total = 118	Kelas G = 39 Kelas H = 37 Kelas I = 40 Total = 116	351 orang
2. Filtrasi Kriteria Eksklusi				307 orang
- Absen akibat Sakit / Izin resmi - Ketidakhadiran Pertemuan Inti (5-8) - Cacat Pengisian Lembar Evaluasi	- 5 mahasiswa sakit/izin - 4 mahasiswa absen kelas - 6 lembar tes cacat/kosong Sisa = 102	- 6 mahasiswa sakit/izin - 3 mahasiswa absen kelas - 6 lembar tes cacat/kosong Sisa = 103	- 4 mahasiswa sakit/izin - 5 mahasiswa absen kelas - 5 lembar tes cacat/kosong Sisa = 102	
3. Prosedur Equalization / Drop	Diundi secara acak untuk mengeluarkan	Diundi secara acak untuk mengeluarkan	Diundi secara acak untuk mengeluarkan	7 orang di drop secara acak

	2 data responden	3 data responden	2 data responden	
4. Final Sample Size) (Data masuk uji SPSS Bab IV)	$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right)$	$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right)$	$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right)$	$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right)$

Sumber: Data administrasi akademik dan rekapitulasi kehadiran penelitian.

Tabel 3.2 memperlihatkan bahwa penyusutan data berlangsung wajar, yaitu sebesar 14,53 persen dari jumlah terdaftar, dan tersebar cukup merata pada ketiga kelompok. Kemerataan penyusutan ini memiliki arti metodologis tersendiri, sebab penyusutan yang menumpuk pada satu kelompok akan menimbulkan ancaman *differential attrition* yang dapat merusak kesetaraan kelompok. Karena penyusutan tersebar merata dan alasannya bersifat administratif serta tidak berkaitan dengan kemampuan akademik, ancaman tersebut dapat dinyatakan terkendali.

C. Perlakuan dan Prosedur Penelitian

1. Penetapan Garis Dasar dan Pengendalian Efek Kebaruan

Salah satu ancaman serius terhadap validitas internal penelitian media pembelajaran adalah efek kebaruan, yaitu peningkatan motivasi semu yang muncul hanya karena subjek berhadapan dengan sesuatu yang baru. Untuk menekan ancaman tersebut, peneliti menerapkan prosedur penetapan garis dasar pra-intervensi. Sebelum stimulus Modul Digital dan Video PjBL diterapkan, seluruh mahasiswa pada ketiga kelompok menjalani empat pertemuan awal secara normal dengan durasi sembilan puluh menit setiap pertemuan. Pada keempat pertemuan awal ini seluruh kelompok memperoleh pembelajaran konvensional tatap muka yang identik tanpa sentuhan teknologi baru, sehingga tercapai stabilisasi sampel.

Setelah pertemuan keempat berakhir, dilaksanakan pengukuran awal dengan durasi tepat empat puluh lima menit menggunakan *Parallel Form A*. Skor *pretest* inilah yang menjadi titik acuan sah untuk memeriksa apakah kondisi literasi keagamaan awal mahasiswa berada pada titik yang sama sebelum menerima perlakuan sesungguhnya. Selanjutnya perlakuan diberikan

selama empat pertemuan, yaitu pertemuan kelima sampai kedelapan, dan pada akhir pertemuan kedelapan seluruh kelompok menjalani pengukuran akhir dengan durasi yang sama, yaitu tepat empat puluh lima menit, menggunakan *Parallel Form B*. Penyetaraan durasi evaluasi dijaga konstan agar perbedaan skor tidak bersumber pada perbedaan waktu pengerjaan. Demikian pula beban waktu paparan perlakuan dikunci setara pada ketiga kelompok, yaitu tiga ratus enam puluh menit, tanpa penugasan rumah, sehingga tidak ada kelompok yang memperoleh tambahan waktu belajar di luar kelas. Seluruh alur prosedur tersebut diringkas pada tabel berikut:

Tabel 3. 3 Protokol Operasional Garis Dasar dan Penyetaraan Perlakuan

Sesi atau Pertemuan	Durasi	Kelompok Eksperimen 1	Kelompok Eksperimen 2	Kelompok Kontrol	Sasaran Kendali Validitas Internal
Pertemuan 1 sampai 4	90 menit	Pembelajaran konvensional tatap muka tanpa modul	Pembelajaran konvensional tatap muka tanpa video proyek	Pembelajaran konvensional tatap muka	Stabilisasi sampel untuk meniadakan efek kebaruan
Akhir pertemuan 4	45 menit	Pretest Form A	Pretest Form A	Pretest Form A	Pemeriksaan empiris kesetaraan kemampuan awal
Pertemuan 5 sampai 8	90 menit	Eksplorasi Modul Digital secara mandiri terbimbing disertai studi kasus kontekstual	Produksi proyek video kelompok dengan jadwal ketat	Ceramah ekspositori dan diskusi kelas tanpa stimulus media baru	Penyetaraan beban waktu belajar sebesar 360 menit tanpa penugasan rumah
Akhir pertemuan 8	45 menit	Posttest Form B	Posttest Form B	Posttest Form B	Pengukuran capaian akhir literasi keagamaan

2. Tahap Persiapan

Tahap persiapan merupakan rangkaian langkah awal yang meliputi penyusunan perangkat pembelajaran, pengembangan instrumen tes, penyamaan persepsi materi, serta koordinasi akhir dengan program studi dan dosen pengampu. Kegiatan pada tahap ini berlangsung sekitar dua bulan dan

mencakup penyusunan Modul Digital, perancangan instrumen *pretest* dan *posttest* dalam dua bentuk paralel, penyiapan kuis formatif, serta penyusunan Panduan Video PjBL beserta rubrik penilaian karyanya. Bahan pembelajaran tidak hanya disusun berdasarkan capaian mata kuliah Pendidikan Agama Islam, tetapi juga dipetakan ke dalam tiga dimensi literasi keagamaan agar setiap aktivitas memiliki dasar teoretis dan tujuan instruksional yang jelas.

Justifikasi perbandingan dua perlakuan dalam satu desain perlu dikemukakan pada tahap ini. Penelitian ini bermaksud menguji secara empiris dua arus teoretis belajar yang berbeda, yaitu jalur individual-kognitif melalui belajar mandiri terarah dan jalur sosial-kolaboratif melalui bantuan antarsejawat, dalam pengaruhnya terhadap tiga dimensi literasi keagamaan. Membandingkan keduanya dalam satu rancangan yang terkendali memberikan sumbangan bagi pengambil kebijakan kurikulum dalam memetakan strategi pembelajaran digital yang paling sesuai untuk capaian kompetensi tertentu, dan bukan sekadar menyatakan bahwa media digital lebih baik daripada pembelajaran konvensional.

3. Pelaksanaan Pretest

Pretest dilaksanakan serentak pada ketiga kelompok di akhir pertemuan keempat untuk memperoleh gambaran kemampuan awal literasi keagamaan mahasiswa. Tes terdiri atas dua puluh butir soal yang dikerjakan secara mandiri melalui platform Wayground, dengan pengawasan langsung untuk menjaga kemandirian pengerjaan dan mencegah kerja sama. Skor *pretest* memiliki dua fungsi metodologis sekaligus. Pertama, ia menjadi data garis dasar untuk memeriksa kesetaraan kemampuan awal antarkelompok, sehingga bias seleksi yang melekat pada desain tanpa pengacakan dapat dikendalikan. Kedua, ia menjadi pembanding utama bagi skor *posttest* dalam penghitungan selisih dan N-Gain.

4. Perlakuan pada Kelompok Eksperimen 1 dengan Modul Digital

Pada Kelompok Eksperimen 1, mahasiswa memperoleh perlakuan berupa Modul Digital PAI *Islam Rahmatan Lil Alamin* yang dirancang dengan pendekatan belajar mandiri terarah. Modul memuat penyajian materi yang

sistematis pada tiga dimensi literasi keagamaan, video pembelajaran, kuis formatif dengan umpan balik langsung, serta ruang konsultasi akademik dengan dosen. Dimensi *Religious Knowledge* dikembangkan melalui materi fungsi agama, trilogi Islam yang mencakup akidah, syariah, dan akhlak, relasi wahyu dan nalar, *maqashid syariah*, serta *rukhsah*. Dimensi *Interpretative Understanding* dikembangkan melalui prinsip *tasamuh*, *tawazun*, *tawassut*, dan *tabayyun* digital. Adapun dimensi *Social Engagement* dikembangkan melalui tema integritas akademik, kepedulian ekologis, antikorupsi, kesehatan jiwa, resolusi konflik lintas keyakinan, dan pergaulan sehat.

Perlu ditegaskan bahwa modul ini tidak dirancang untuk mengulang materi keagamaan yang pernah dipelajari mahasiswa di jenjang sekolah, melainkan untuk menata dan mengaktifkan kembali pengetahuan yang telah dimiliki namun tidak lagi tersusun. Karena itu modul dilengkapi aktivitas penjembutan berupa refleksi awal, pemetaan konsep, dan tugas yang mengaitkan pengalaman belajar Pendidikan Agama Islam di sekolah dengan persoalan keagamaan yang dihadapi di kampus. Seluruh eksplorasi modul dilaksanakan di dalam kelas selama sembilan puluh menit setiap pertemuan di bawah pendampingan dosen, tanpa penugasan rumah, agar beban waktu belajar tetap setara dengan kelompok lain.

5. Perlakuan pada Kelompok Eksperimen 2 dengan Video PjBL

Pada Kelompok Eksperimen 2, pembelajaran dilaksanakan melalui proyek video berbasis *Project Based Learning* yang menempatkan mahasiswa sebagai subjek aktif penghasil karya.¹³ Alur proyek berlangsung dalam lima tahap. Tahap pertama adalah penentuan dan pemahaman kasus, ketika kelompok memilih isu keagamaan kontemporer di lingkungan kampus, merumuskan pesan utama, dan mengaitkannya dengan nilai-nilai Islam serta ketiga dimensi literasi keagamaan. Tahap kedua adalah penyusunan naskah yang memuat alur, dialog, pembagian adegan, dan dasar argumentasi keagamaan. Tahap ketiga adalah persiapan produksi, meliputi penelusuran

¹³ John W. Thomas, *A Review of Research on Project-Based Learning* (San Rafael, CA: The Autodesk Foundation, 2000), 1-9.

dalil, survei lokasi, penyiapan properti, dan pembagian peran sebagai penulis naskah, pemeran, juru kamera, penyunting, serta penanggung jawab konten. Tahap keempat adalah pengambilan gambar dan penyuntingan. Tahap kelima adalah peninjauan, konsultasi kepada dosen, revisi, publikasi, dan refleksi.

Seluruh proses produksi dilaksanakan di dalam kelas dalam durasi sembilan puluh menit setiap pertemuan tanpa penugasan rumah, guna menghindari bias tambahan waktu belajar yang dapat mengganggu kesetaraan perlakuan. Durasi tersebut dibagi menjadi tiga fase, yaitu dua puluh menit untuk koordinasi tim dan penelusuran dalil atas kasus yang dipilih, empat puluh menit untuk pengambilan gambar, dan tiga puluh menit untuk penyuntingan dengan aplikasi berbasis gawai. Produk akhir berupa video berdurasi tiga sampai lima menit yang dipublikasikan secara terbuka melalui kanal YouTube dengan tagar makulagamaburohmah, sebagai bentuk dokumentasi sekaligus representasi pemahaman mahasiswa atas materi yang dipelajari.

Kedudukan karya video dalam kerangka pengukuran perlu dinyatakan secara tegas sejak bab metode ini agar tidak menimbulkan salah baca pada bab hasil. Karya video merupakan bagian dari perlakuan, yaitu wahana tempat mahasiswa mengonstruksi dan mengomunikasikan pemahaman keagamaannya, dan bukan instrumen pengukur variabel terikat. Variabel terikat literasi keagamaan diukur sepenuhnya melalui tes *pretest* dan *posttest* yang menghasilkan satu skor komposit. Adapun rubrik enam aspek berkedudukan sebagai perangkat penilaian proses dan produk proyek yang menuntun mahasiswa dan dosen selama proyek berlangsung, bukan sebagai penghasil skor yang diuji secara statistik.

6. Pembelajaran pada Kelompok Kontrol

Pembelajaran pada Kelompok Kontrol dilaksanakan sebagaimana perkuliahan pada umumnya, yaitu melalui ceramah bervariasi, tanya jawab, dan diskusi kelas yang berpusat pada dosen, tanpa penambahan stimulus media baru. Materi pokok, capaian pembelajaran, cakupan dan kedalaman bahasan, serta alokasi waktu dijaga setara dengan kedua kelompok eksperimen. Dengan demikian, satu-satunya perbedaan yang sengaja diciptakan di antara ketiga

kelompok adalah jenis media dan strategi pembelajarannya, sehingga kelompok ini berfungsi sebagai pembanding yang sah bagi kedua perlakuan.

7. Pelaksanaan Posttest

Posttest dilaksanakan serentak pada ketiga kelompok di akhir pertemuan kedelapan menggunakan dua puluh butir soal *Parallel Form B* melalui platform Wayground dengan durasi tepat empat puluh lima menit. Penggunaan platform tersebut dimaksudkan agar pelaksanaan tes lebih terstruktur dan terdokumentasi, sekaligus meminimalkan potensi kecurangan karena sistem memungkinkan pengendalian waktu, akses, dan rekaman respons peserta. Seluruh pelaksanaan tes diawasi oleh pengamat independen untuk memastikan prosedur berjalan sesuai ketentuan dan setiap mahasiswa mengerjakan tes secara mandiri. Skor *posttest* selanjutnya dibandingkan dengan skor *pretest* untuk menghitung selisih dan N-Gain sebagai dasar penilaian efektivitas perlakuan.

8. Peran Ganda Peneliti dan Pengamat Independen

Dalam rancangan ini peneliti berperan ganda sebagai pengembang produk sekaligus dosen pengampu. Peran ganda tersebut menimbulkan potensi bias peneliti dan ancaman terhadap validitas internal, khususnya pada aspek kesetaraan perlakuan. Peneliti memandang perlu menyatakan potensi bias ini secara terbuka, dan kemudian menempuh langkah pengendalian yang dapat diperiksa. Langkah pengendalian yang ditempuh adalah melibatkan dua pengamat independen, yaitu Mafatih Fauzul Adzima, M.Pd. dan Nurul Mamnun, M.Pd.I., keduanya dosen Pendidikan Agama Islam yang tidak terlibat sebagai peneliti maupun pengembang produk. Kedua pengamat bertugas mengamati jalannya pembelajaran pada ketiga kelas tanpa melakukan intervensi, menggunakan Lembar Observasi Kesetaraan Perlakuan yang memeriksa kesetaraan alokasi waktu, kompetensi dasar, tujuan pembelajaran, cakupan dan kedalaman materi, pola interaksi dosen dan mahasiswa, serta identik atau paralelnya instrumen evaluasi. Hasil pengamatan tersebut menjadi dasar bagi pernyataan pada Bab IV bahwa selisih capaian yang muncul pada *posttest* lebih dapat diatribusikan pada perlakuan dan bukan pada

ketidaksetaraan pelaksanaan pembelajaran. Dengan demikian, laporan pengamat bukan pelengkap administratif, melainkan syarat yang memungkinkan klaim kausal penelitian ini dipertahankan.

D. Instrumen Penelitian

1. Prinsip Pengembangan Instrumen

Instrumen penelitian ini terdiri atas instrumen utama dan instrumen pendukung. Instrumen utama ialah Tes Literasi Keagamaan Mahasiswa yang diberikan kepada seluruh subjek penelitian, baik pada kelas eksperimen 1 (modul digital), kelas eksperimen 2 (Video Project Based Learning), maupun kelas kontrol (pembelajaran biasa). Instrumen pendukung terdiri atas lembar keterlaksanaan perlakuan, lembar observasi proses pembelajaran, dan rubrik kualitas proyek video.

Instrumen penelitian ini terdiri atas instrumen utama dan instrumen pendukung. Instrumen utama adalah Tes Literasi Keagamaan Mahasiswa, yang diberikan kepada seluruh subjek pada ketiga kelompok dan menjadi satu-satunya alat ukur variabel terikat. Instrumen pendukung terdiri atas lembar validasi Modul Digital, lembar validasi Panduan Video PjBL, rubrik penilaian karya video, lembar observasi keterlaksanaan pembelajaran, dan Lembar Observasi Kesetaraan Perlakuan. Perbedaan tegas antara instrumen utama dan pendukung ini menentukan batas klaim penelitian: hanya data dari instrumen utama yang diuji secara inferensial, sedangkan data dari instrumen pendukung dipakai untuk menegaskan kelayakan produk dan keterlaksanaan perlakuan.

Dalam desain tiga kelompok dengan pengukuran berulang, pengukuran capaian harus menggunakan instrumen yang ekuivalen agar perbedaan skor dapat ditafsirkan sebagai akibat perlakuan dan bukan akibat perbedaan alat ukur. Karena itu penelitian ini menggunakan dua bentuk paralel, yaitu Form A untuk *pretest* dan Form B untuk *posttest*. Keduanya tidak identik secara redaksi, tetapi wajib sama dalam lima hal, yaitu konstruk yang diukur, distribusi indikator per dimensi, jumlah butir, tingkat kesukaran yang ditargetkan, dan

jenjang kognitif serta sistem penskoran.¹⁴ Pemetaan variabel penelitian ke instrumen yang mengukurnya disajikan pada tabel dibawah ini.

Tabel 3. 4 Matriks Variabel dan Instrumen Penelitian

Variabel	Kedudukan	Indikator	Instrumen	Perlakuan Data
Modul Digital	Variabel perlakuan pertama	Keterbacaan materi, sistematika penyajian, interaktivitas media	Lembar validasi modul dan lembar keterlaksanaan modul	Deskriptif, tidak diuji inferensial
Video PjBL	Variabel perlakuan kedua	Eksplorasi konsep, kolaborasi proyek, produksi karya	Lembar validasi panduan, lembar keterlaksanaan, rubrik karya video	Deskriptif, tidak diuji inferensial
Pembelajaran konvensional	Kondisi pembanding	Penyampaian materi reguler, tanya jawab, diskusi	Lembar observasi kelas kontrol	Deskriptif, tidak diuji inferensial
Literasi keagamaan	Variabel terikat	Religious Knowledge, Interpretative Understanding, Social Engagement	Tes Literasi Keagamaan Form A dan Form B	Skor komposit, diuji inferensial pada Bab IV

2. Spesifikasi Tes Literasi Keagamaan

Konstruksi instrumen evaluasi literasi keagamaan dalam penelitian ini dikembangkan secara deduktif dengan menurunkan dimensi teoretis makro menjadi indikator perilaku yang terukur (*measurable behavior*). Spesifikasi tes dikunci secara ketat dengan mengadopsi taksonomi dimensi Literasi Keagamaan menurut Diane L. Moore yang diselaraskan dengan capaian pembelajaran kurikulum Mata Kuliah Wajib Kurikulum (MKWK) Pendidikan Agama Islam (PAI) di Perguruan Tinggi Umum. Guna menjamin rigoritas pengukuran pada tingkat studi doctoral, instrumen dirancang tidak untuk menguji aspek hafalan kognitif tingkat rendah (*low-order thinking skills*), melainkan menggunakan stimulus berbasis studi kasus fenomena sosio-religius kontemporer di Indonesia untuk menguji kemampuan

¹⁴ Linda Crocker dan James Algina, *Introduction to Classical and Modern Test Theory* (New York: Holt, Rinehart and Winston, 1986), 131-137.

berpikir tingkat tinggi (*Higher Order Thinking Skills / HOTS*). Distribusi dan bobot butir soal dikonstruksikan secara berimbang ke dalam total 20 butir soal objektif pilihan ganda (dengan opsi jawaban A, B, C, D, E) untuk *Parallel Form A* (Pretest) dan *Parallel Form B* (Posttest), sebagaimana disajikan dalam Tabel 3.5:

Tabel 3. 5 Kisi-kisi Tes Literasi Keagamaan (Form A dan Form B Paralel)

No	Dimensi Teoretis	Indikator Kompetensi Operasional	Karakteristik Stimulus Soal	Jenjang Kognitif	Nomor Butir Form A	Nomor Butir Form B	Jumlah dan Bobot
1	Religious Knowledge	Menganalisis teks keagamaan secara tekstual dan kontekstual serta mengidentifikasi perkembangan pemikiran hukum Islam secara objektif	Teks kasus perbedaan penafsiran ayat hukum; mahasiswa menentukan metodologi istinbath hukum yang tepat	Memahami sampai menganalisis	1 sampai 7	1 sampai 7	7 butir, 35 persen
2	Interpretative Understanding	Menguraikan keragaman internal dalam tradisi umat Islam dan memvalidasi sikap toleransi terhadap ekspresi sosial-keagamaan	Narasi dilematis benturan tradisi lokal dengan ritus keagamaan; mahasiswa menganalisis batas toleransi inklusif	Menganalisis sampai mengevaluasi	8 sampai 13	8 sampai 13	6 butir, 30 persen
3	Social Engagement	Merumuskan solusi etis-religius atas konflik sosio-religius dan menunjukkan kesiapan tindakan nyata	Studi kasus penyebaran hoaks bernuansa keagamaan di media sosial; mahasiswa memilih keputusan tindakan yang paling maslahat	Menganalisis sampai mengevaluasi	14 sampai 20	14 sampai 20	7 butir, 35 persen
	Total	Mencakup tiga dimensi literasi keagamaan	Objektif pilihan ganda lima opsi berbasis kasus	Berpikir jenjang tinggi	20 butir	20 butir	20 butir, 100 persen

3. Bentuk Paralel Form A dan Form B

Penggunaan dua bentuk paralel ditujukan untuk mencegah efek pengujian, yaitu meningkatnya skor semata karena mahasiswa telah mengenal butir yang sama pada *pretest*. Paralelitas kedua bentuk diuji secara empiris pada tahap uji coba melalui tiga prosedur. Pertama, korelasi skor Form A dan Form B pada subjek uji coba, dengan ambang keterandalan bentuk paralel minimum 0,80. Kedua, uji perbedaan rerata kedua bentuk pada subjek uji coba, yang disyaratkan tidak signifikan pada taraf lima persen. Ketiga, pembandingan distribusi tingkat kesukaran antarbentuk. Apabila salah satu prosedur menunjukkan ketidaksetaraan, butir yang menjadi sumber ketidaksetaraan direvisi sebelum instrumen digunakan pada sampel penelitian.

Peneliti menegaskan argumentasi metodologis yang kokoh untuk menepis potensi bias pengganggu (*confounding effect*) yang diakibatkan oleh perbedaan moda interaksi, di mana Kelompok Eksperimen 1 melakukan aktivitas belajar secara individu mandiri (Modul Digital), sedangkan Kelompok Eksperimen 2 melakukan aktivitas belajar secara kelompok kolaboratif (*Video Project Based Learning*). Secara epistemologis, perbedaan format interaksi fisik tersebut bukanlah sebuah bias eksperimental (*methodological flaw*), melainkan representasi dari pengaplikasian teoretis inheren yang membedakan karakteristik kedua media yang sedang diperbandingkan efektivitasnya. Pembelajaran berbasis Modul Digital dirancang untuk menguji daya akselerasi kemandirian kognitif (*self-regulated learning*), sedangkan *Video Project Based Learning* dirancang untuk menguji daya akselerasi konstruktivisme sosial melalui efek kerja sama sejawat (*peer-scaffolding effect*).

Untuk mengisolasi dampak kausalitas agar murni merepresentasikan tingkat efektivitas substansi media dan terbebas dari bias perlakuan makro tersebut, peneliti menetapkan kontrol ketat pada Unit Analisis Pengukuran Data. Walaupun proses eksekusi proyek Video PjBL dilakukan secara berkelompok oleh mahasiswa, prosedur pengambilan data evaluasi tingkat literasi keagamaan pada sesi *Pretest* (Form A) maupun *Posttest* (Form B) dikunci secara mutlak pada level individu mandiri untuk seluruh kelompok perlakuan (Eksperimen 1, Eksperimen 2, dan Kontrol). Seluruh mahasiswa pengerjaannya diwajibkan menyelesaikan 20 butir

soal evaluasi tersebut secara individual tanpa diperkenankan melakukan diskusi atau kerja sama.

Melalui pemisahan yang tegas antara format intervensi ekologis (kelompok/individu) dengan format evaluasi psikometris (individu mandiri) yang seragam berdurasi tepat 45 menit, maka varians skor yang dihasilkan pada Bab IV murni mencerminkan kapasitas literasi keagamaan dari tiap-tiap subjek responden, sehingga penarikan kesimpulan melalui uji statistik *Kruskal-Wallis* dan uji lanjut *Mann-Whitney U* memiliki derajat validitas internal yang valid dan terbebas dari bias metode pengganggu.

4. Penskoran dan Kedudukan Skor

Tes literasi keagamaan terdiri atas dua puluh butir dengan penskoran dikotomus, yaitu jawaban benar diberi skor lima dan jawaban salah diberi skor nol, sehingga skor maksimum yang dapat dicapai adalah seratus dan skor minimum adalah nol. Waktu pengerjaan ditetapkan empat puluh lima menit, sama untuk *pretest* maupun *posttest*. Skor total dari dua puluh butir tersebut diperlakukan sebagai satu skor komposit literasi keagamaan yang mencakup ketiga dimensi secara sekaligus.

Keputusan memperlakukan skor sebagai komposit tunggal memiliki konsekuensi yang harus dinyatakan secara terbuka. Karena jumlah butir per dimensi relatif kecil, yaitu enam sampai tujuh butir, penskoran terpisah per dimensi tidak memiliki kecukupan psikometris untuk diuji secara inferensial. Pembatasan ini sengaja ditetapkan agar klaim penelitian tetap berada dalam jangkauan data yang benar-benar dikumpulkan, dan analisis per dimensi direkomendasikan sebagai agenda penelitian lanjutan.

5. Lembar Validasi Modul Digital

Lembar validasi Modul Digital memuat dua puluh empat butir penilaian yang tersebar pada lima aspek. Penilaian menggunakan skala Likert empat titik, yaitu satu untuk tidak sesuai, dua untuk kurang sesuai, tiga untuk sesuai, dan empat untuk sangat sesuai. Skala genap dipilih secara sengaja untuk menghindari kecenderungan penilai memilih posisi tengah yang aman, yang lazim terjadi pada

skala ganjil. Indikator butir disintesis dari standar kelayakan bahan ajar Badan Standar Nasional Pendidikan, kerangka literasi keagamaan Moore, dan prinsip desain pembelajaran multimedia Mayer.¹⁵ Sebaran butir per aspek disajikan pada tabel berikut :

Tabel 3. 6 Sebaran Butir Lembar Validasi Modul Digital

No	Aspek Penilaian	Cakupan Indikator	Jumlah Butir
1	Kesesuaian isi dan materi	Keselarasan materi dengan tujuan pembelajaran, kerangka Islam rahmatan lil alamin, muatan ketiga dimensi literasi keagamaan, dan relevansi dengan konteks kampus	6
2	Kesesuaian instrumen dan aktivitas pembelajaran	Keselarasan instrumen pretest dan posttest dengan tujuan, keterwakilan ketiga dimensi, kejelasan redaksi butir, serta dukungan aktivitas, kuis, video, dan proyek	5
3	Desain instruksional	Kelogisan struktur modul, kejelasan alur penggunaan, kesesuaian dengan pendekatan blended learning mandiri, dan dukungan panduan dosen	5
4	Bahasa dan keterbacaan	Kesesuaian bahasa dengan karakteristik mahasiswa, kejelasan kalimat, ketepatan istilah keagamaan, dan kejelasan petunjuk penggunaan	4
5	Tampilan media	Daya tarik visual, tata letak, warna dan tipografi, kemudahan navigasi fitur, serta kelayakan desain untuk perkuliahan	4
	Jumlah	Lima aspek penilaian	24

6. Lembar Validasi Panduan Video PjBL

Lembar validasi Panduan Video PjBL disusun terpisah dari lembar validasi modul karena keduanya menilai objek yang berbeda watak. Lembar ini memuat dua puluh empat butir penilaian yang tersebar pada enam aspek dengan skala Likert empat titik yang sama. Perbedaan pokok dengan lembar validasi modul terletak pada masuknya aspek kesesuaian dengan prinsip *Project Based Learning* dan aspek rubrik penilaian video, dua hal yang tidak relevan bagi modul namun menentukan bagi panduan proyek. Sebaran butir per aspek disajikan pada tabel berikut :

¹⁵ Richard E. Mayer, *Multimedia Learning*, ed. ke-3 (Cambridge: Cambridge University Press, 2021), 57-72.

Tabel 3. 7 Sebaran Butir Lembar Validasi Panduan Video PjBL

No	Aspek Penilaian	Cakupan Indikator	Jumlah Butir
1	Kesesuaian konten dan landasan normatif	Ketepatan rujukan dalil, keselarasan dengan kerangka Islam rahmatan lil alamin, dan keterwakilan ketiga dimensi literasi keagamaan	5
2	Kesesuaian dengan prinsip Project Based Learning	Keberadaan pertanyaan pemandu, kemandirian kelompok, produk nyata, dan refleksi	4
3	Rubrik penilaian video	Kejelasan kriteria, ketepatan deskriptor tiap jenjang, dan keterpetaan aspek rubrik pada dimensi literasi keagamaan	5
4	Desain aktivitas dan ketentuan proyek	Kelogisan lima tahap proyek, kejelasan ketentuan durasi dan publikasi, serta kelayakan alokasi waktu di kelas	4
5	Bahasa dan keterbacaan	Kesesuaian bahasa, kejelasan instruksi, dan ketepatan istilah teknis maupun keagamaan	3
6	Tampilan dan kemudahan penggunaan panduan	Tata letak, kejelasan ilustrasi visual, dan kemudahan panduan diikuti mahasiswa secara mandiri	3
	Jumlah	Enam aspek penilaian	24

7. Rubrik Penilaian Karya Video

Rubrik penilaian karya video merupakan bagian dari Panduan Video PjBL dan telah dinilai kelayakannya oleh ahli sebagai salah satu aspek validasi panduan. Rubrik terdiri atas enam aspek, tiga di antaranya dipetakan langsung pada ketiga dimensi literasi keagamaan, sedangkan tiga lainnya berfungsi sebagai aspek pendukung yang menopang kreativitas, komunikasi multimedia, dan internalisasi nilai. Pemetaan konseptual antara aspek rubrik dan dimensi literasi keagamaan disajikan pada tabel berikut :

Tabel 3. 8 Pemetaan Aspek Rubrik Karya Video dengan Dimensi Literasi Keagamaan

Aspek Rubrik	Dimensi Literasi Keagamaan	Fokus Pengamatan pada Karya Video
Akurasi konten	Religious Knowledge	Ketepatan rujukan ayat dan hadis serta akurasi konsep akidah, syariah, akhlak, dan maqashid syariah
Kedalaman interpretasi	Interpretative Understanding	Penggunaan prinsip tasamuh, tawazun, tawassut, dan tabayyun serta kemampuan membaca konteks

Aspek Rubrik	Dimensi Literasi Keagamaan	Fokus Pengamatan pada Karya Video
Dampak sosial	Social Engagement	Relevansi isu dengan kehidupan kampus, ajakan tindakan, dan kontribusi pada moderasi serta integritas
Kolaborasi tim	Aspek pendukung kolaborasi	Pembagian peran, pengakuan kontribusi anggota, dan keutuhan narasi kolektif
Kualitas produksi	Aspek pendukung komunikasi multimedia	Kejernihan gambar dan suara, mutu penyuntingan, dan kepatuhan pada ketentuan durasi
Refleksi nilai	Aspek pendukung internalisasi nilai	Penegasan pelajaran, ajakan moral, dan kontekstualisasi nilai Islam rahmatan lil alamin

Sesuai penegasan pada bagian perlakuan, rubrik ini tidak menghasilkan data yang diuji secara statistik. Penilaian karya video dalam penelitian ini berhenti pada pertimbangan kelayakan secara kualitatif dan tidak diskorkan per aspek untuk setiap karya, sehingga tidak tersedia data kuantitatif per dimensi yang dapat dianalisis secara inferensial. Penskoran rubrik secara penuh atas seluruh karya beserta analisis isi video merupakan agenda lanjutan yang direkomendasikan bagi penelitian berikutnya.

8. Lembar Observasi Kesetaraan Perlakuan

Lembar Observasi Kesetaraan Perlakuan merupakan instrumen pendukung yang secara khusus dirancang untuk menjawab ancaman bias yang timbul dari peran ganda peneliti. Lembar ini diisi oleh dua pengamat independen dan memeriksa enam butir kesetaraan, yaitu alokasi waktu setiap pertemuan, kompetensi dasar yang dibelajarkan, tujuan pembelajaran, cakupan dan kedalaman materi, pola interaksi dosen dan mahasiswa, serta identik atau paralelnya instrumen evaluasi. Setiap butir diisi dengan pilihan setara atau tidak setara disertai kolom catatan kualitatif. Instrumen ini juga memeriksa secara eksplisit bahwa perbedaan yang muncul antarkelas hanya terletak pada jenis media dan model pembelajaran, sesuai rancangan penelitian. Hasil pengisian lembar observasi inilah yang dilaporkan pada Subbab A Bab IV sebagai landasan bagi penafsiran kausal atas selisih capaian antarkelompok.

E. Validitas dan Reliabilitas

1. Validitas Isi Produk melalui Penilaian Ahli

Validasi ahli menempati posisi penting dalam penelitian yang mengembangkan produk lalu mengujinya secara kuasi-eksperimen. Pada tahap ini produk tidak lagi dinilai pada tataran gagasan, melainkan dihadapkan pada penilaian profesional para ahli yang kompeten di bidang substansi keagamaan, desain instruksional, dan media pembelajaran digital. Dalam kerangka ADDIE, validasi berada pada fase pengembangan dan berfungsi sebagai mekanisme penjaminan mutu sebelum produk diimplementasikan di lapangan.

Penilaian ahli dilakukan atas dua produk, yaitu Modul Digital dan Panduan Video PjBL, menggunakan lembar terstruktur sebagaimana diuraikan pada Tabel 3.6 dan Tabel 3.7. Rerata skor penilaian dikonversi ke kategori kelayakan mengikuti rentang yang dirumuskan Akbar, yaitu 3,26 sampai 4,00 untuk kategori sangat layak tanpa revisi, 2,51 sampai 3,25 untuk kategori layak dengan revisi kecil, 1,76 sampai 2,50 untuk kategori kurang layak dengan revisi besar, dan 1,00 sampai 1,75 untuk kategori tidak layak.¹⁶ Produk dinyatakan dapat digunakan sebagai stimulus eksperimen apabila mencapai kategori sekurang-kurangnya layak dan seluruh catatan perbaikan validator telah ditindaklanjuti.

Panel ahli dalam penelitian ini terdiri atas lima validator dengan latar keahlian yang berbeda dan saling melengkapi, sebagaimana disajikan pada Tabel 3.9. Penetapan lima validator dari sudut keahlian yang berlainan bukan sekadar upaya memperbanyak penilai, melainkan bentuk triangulasi ahli yang memungkinkan validitas isi diperiksa dari beberapa sisi sekaligus, yaitu sisi keilmuan Pendidikan Agama Islam pada perguruan tinggi, sisi praktik dakwah digital dan moderasi beragama, sisi kontinuitas kurikulum dari jenjang sekolah, sisi keahlian multimedia dengan rekam jejak internasional, dan sisi keterlaksanaan pada kapasitas teknologi institusi tempat penelitian berlangsung.

¹⁶ Sadun Akbar, *Instrumen Perangkat Pembelajaran* (Bandung: Remaja Rosdakarya, 2013), 157-158.

Tabel 3. 9 Panel Validator Ahli dan Fokus Penilaian

No	Validator	Latar Keahlian	Objek dan Fokus Penilaian
1	Ridwan, S.Ag., M.Pd.	Dosen senior bidang pendidikan agama, Universitas Nusantara PGRI Kediri	Modul Digital, kesesuaian substansi keagamaan, ketepatan konsep, teori, dan pendekatan
2	Ning Dhomirotul Firdaus, M.Pd.	Praktisi dan pendidik yang aktif pada ruang publik dan media dakwah digital	Modul Digital, kesesuaian dengan prinsip moderasi beragama, sensitivitas gender, dan otoritas dalil
3	Melly Mahmudah, M.Pd.	Pembelajaran Pendidikan Agama Islam jenjang dasar dan menengah	Modul Digital, kontinuitas kurikulum, penghindaran repetisi verbalistik, dan ketersediaan penata awal
4	Ir. Samodra, Ph.D.	Dosen Universitas Muhammadiyah Gresik, doktor bidang teknologi dan magister computer and information science	Panduan Video PjBL, kerangka pedagogis Project Based Learning, kerangka konten, dan kerangka teknis multimedia
5	Dr. Erna Daniati, M.Kom.	Dosen senior bidang pengembangan media digital dan pembelajaran elektronik, Universitas Nusantara PGRI Kediri	Panduan Video PjBL, keterlaksanaan panduan pada kultur, infrastruktur, dan kapasitas teknologi institusi

Satu batas penafsiran perlu dinyatakan sejak bab metode ini. Kelayakan yang dihasilkan prosedur validasi adalah kelayakan produk, bukan bukti efektivitas, dan keduanya berbeda kedudukan. Validasi memastikan bahwa stimulus yang akan diuji memang layak diuji, tetapi belum menyatakan apa pun tentang besar pengaruhnya. Karena itu legitimasi yang diperoleh dari validasi berhenti pada pemberian dasar yang sah untuk melangkah ke tahap implementasi, sedangkan seluruh klaim mengenai daya pengaruh baru ditegakkan setelah data *pretest* dan *posttest* diuji secara inferensial.

2. Validitas Isi Instrumen Tes

Validitas isi instrumen tes ditegakkan melalui penilaian ahli atas setiap butir soal pada empat aspek, yaitu kesesuaian butir dengan dimensi yang diukur, kesesuaian butir dengan indikator operasional, ketepatan bahasa dan keterbacaan, serta kesesuaian jenjang kognitif dengan sasaran. Penilaian menggunakan skala empat titik yang sama dengan penilaian produk. Butir dinyatakan memenuhi validitas isi apabila memperoleh penilaian sekurang-kurangnya sesuai dari seluruh validator, dan butir yang memperoleh penilaian di bawah ambang tersebut direvisi

atau digantikan sebelum instrumen dipakai. Prosedur ini penting karena keterwakilan ketiga dimensi pada dua puluh butir soal merupakan dasar yang menopang seluruh penghitungan N-Gain dan pengujian efektivitas.

3. Validitas Konstruk

Validitas konstruk diperiksa melalui analisis butir atas data uji coba yang dikumpulkan dari subjek di luar sampel penelitian namun memiliki karakteristik setara. Karena instrumen berbentuk dikotomis, analisis menggunakan korelasi *point-biserial* antara skor butir dan skor total, dengan ambang butir dinyatakan valid apabila koefisiennya sekurang-kurangnya 0,30.¹⁷ Selain koefisien korelasi butir, dilaporkan pula dua karakteristik psikometris tambahan, yaitu tingkat kesukaran butir yang diharapkan berada pada rentang 0,30 sampai 0,80 agar sebaran butir seimbang, dan daya pembeda butir yang diharapkan sekurang-kurangnya 0,30 agar butir mampu membedakan mahasiswa berkemampuan tinggi dan rendah. Butir yang tidak memenuhi ketiga kriteria tersebut direvisi atau dikeluarkan, dan penggantinya diuji ulang sebelum instrumen dinyatakan siap dipakai.

Rumus korelasi *point-biserial* yang digunakan adalah sebagai berikut.

$$r(pbi) = ((Mp - Mt) / St) \times \text{akar}(p / q)$$

Keterangan: $r(pbi)$ adalah koefisien korelasi point-biserial, Mp adalah rerata skor total pada kelompok yang menjawab butir dengan benar, Mt adalah rerata skor total seluruh peserta, St adalah simpangan baku skor total, p adalah proporsi peserta yang menjawab benar, dan q adalah proporsi peserta yang menjawab salah, yaitu satu dikurangi p .

4. Reliabilitas Instrumen

Reliabilitas berkaitan dengan keajegan instrumen, yaitu kemampuan alat ukur menghasilkan data yang konsisten apabila digunakan berulang pada kondisi yang relatif sama. Konsep ini berpijak pada teori pengukuran klasik yang menyatakan bahwa skor hasil pengukuran merupakan gabungan antara skor sejati dan galat pengukuran, sehingga instrumen dengan reliabilitas tinggi menunjukkan galat pengukuran yang rendah.¹⁸

¹⁷ Crocker dan Algina, *Introduction to Classical and Modern Test Theory*, 313-320.

¹⁸ Crocker dan Algina, *Introduction to Classical and Modern Test Theory*, 105-112.

Karena instrumen tes dalam penelitian ini berbentuk dikotomus, koefisien reliabilitas yang digunakan adalah Kuder-Richardson 20, yaitu bentuk khusus dari konsistensi internal yang tepat bagi data penskoran benar dan salah.¹⁹ Penetapan ini merupakan penyesuaian yang disengaja, sebab koefisien alfa Cronbach lebih tepat diperuntukkan bagi instrumen berskala politomus seperti angket, sehingga penggunaannya pada tes dikotomus akan menimbulkan ketidakserasian antara jenis data dan teknik analisis. Rumus yang digunakan adalah sebagai berikut.

$$KR-20 = (k / (k - 1)) \times (1 - (jumlah\ p.q / varians\ total))$$

Keterangan: k adalah jumlah butir, p adalah proporsi peserta yang menjawab butir dengan benar, q adalah proporsi peserta yang menjawab butir dengan salah, dan varians total adalah varians skor total peserta. Instrumen dinyatakan reliabel apabila koefisien KR-20 sekurang-kurangnya 0,70, yang merupakan ambang lazim bagi tes hasil belajar.

Selain konsistensi internal, penelitian ini juga memeriksa reliabilitas bentuk paralel, yaitu korelasi antara skor Form A dan Form B pada subjek uji coba, dengan ambang minimum 0,80. Pemeriksaan ganda ini diperlukan karena instrumen dipakai dalam dua bentuk yang berbeda pada dua waktu pengukuran, sehingga keajegan tidak cukup dibuktikan pada satu bentuk saja, melainkan harus dibuktikan pula pada kesetaraan antarbentuk. Adapun kesepakatan antarpemilai pada lembar validasi dan lembar observasi diperiksa melalui perbandingan konsistensi penilaian antarvalidator dan antarpengamat.

F. Teknik Analisis Data

Analisis data dalam *Quasi-eksperimen* merupakan proses mengubah data empiris menjadi dasar simpulan yang dapat dipertanggungjawabkan secara metodologis. Karena desain ini tidak menggunakan pengacakan individu, prosedur analisis dituntut lebih ketat agar simpulan kausal yang dihasilkan tetap memiliki legitimasi ilmiah. Oleh sebab itu tahapan analisis disusun berjenjang, dimulai dari pembersihan data, eksplorasi karakteristik data, pengujian asumsi statistik, pemeriksaan kesetaraan kemampuan awal, pengujian hipotesis di dalam kelompok,

¹⁹ G. Frederic Kuder dan M. W. Richardson, "The Theory of the Estimation of Test Reliability," *Psychometrika* 2, no. 3 (1937): 151-160.

pengukuran efektivitas, hingga pembandingan efektivitas antarkelompok. Seluruh pengolahan dilakukan dengan bantuan perangkat lunak IBM SPSS Statistics. Berikut merupakan tahapan-tahapan analisis dalam penelitian:

1. Prosedur Pembersihan Data

Sebelum analisis dilakukan, data diperiksa kelengkapan dan kewajaran nilainya. Pemeriksaan mencakup tiga hal, yaitu penelusuran data yang hilang, penelusuran skor yang berada di luar rentang instrumen, dan pemeriksaan kewajaran sebaran melalui nilai kemencengan serta keruncingan. Skor yang ditemukan berada di luar rentang nol sampai seratus dikoreksi ke batas rentang instrumen, karena nilai di luar rentang tidak mungkin dihasilkan oleh alat ukur yang digunakan. Setiap koreksi yang dilakukan dilaporkan secara terbuka pada Bab IV sebagai bagian dari prosedur pembersihan data, sehingga pembaca dapat menilai sendiri pengaruh koreksi tersebut terhadap hasil. Prinsip yang dipegang adalah keterbukaan yakni perbaikan data dilaporkan, bukan disembunyikan.

2. Analisis Deskriptif

Analisis deskriptif dilakukan sebagai langkah awal untuk memahami struktur data secara menyeluruh dan menyediakan konteks bagi penafsiran hasil analisis inferensial. Tanpa pemahaman atas karakteristik data, hasil pengujian statistik berpotensi kehilangan makna substantif. Parameter yang dilaporkan untuk setiap kelompok pada *pretest* maupun *posttest* meliputi jumlah subjek, rerata, galat baku rerata, selang kepercayaan sembilan puluh lima persen untuk rerata, rerata terpangkas lima persen, median, modus, varians, simpangan baku, skor minimum dan maksimum, rentang, rentang antarkuartil, serta kemencengan dan keruncingan beserta galat bakunya.

Pemilihan parameter tersebut tidak bersifat sekadar melengkapi. Rerata dan median dipakai untuk membaca kecenderungan pusat, simpangan baku dan rentang antarkuartil dipakai untuk membaca sebaran, sedangkan kemencengan dan keruncingan dipakai untuk membaca bentuk sebaran yang kelak menjelaskan hasil uji normalitas. Rerata terpangkas lima persen dan selang kepercayaan dilaporkan

untuk memeriksa ketahanan rerata terhadap nilai ekstrem. Parameter yang digunakan meliputi:

a) Nilai Mean (Rata-rata)

Nilai mean menggambarkan posisi pusat data yang merepresentasikan capaian rata-rata mahasiswa dalam suatu kelompok. Mean diperoleh dari penjumlahan nilai seluruh kelompok dan dibagi total kelompok yang ada. Rumus perhitungan mean sebagai berikut ini:

$$\bar{X} = \frac{\sum X_i}{n}$$

Keterangan:

$\bar{X}$ = Mean atau rata-rata

$\sum$ = Jumlah

X_n = Variabel ke n

n = banyaknya data atau sampel

b) Nilai Median (Nilai Tengah)

Nilai Median menggambarkan nilai tengah dari sekumpulan data yang telah diurutkan dari yang terkecil hingga terbesar ataupun sebaliknya dari yang terbesar ke terkecil. Rumus perhitungan median sebagai berikut:

$$X_{\text{Median}} = \frac{X_1 + X_2}{2}$$

Keterangan:

X_1 = Nilai tengah pertama dimana median terletak

X_2 = nilai tengah kedua dimana median terletak

c) Nilai Modus (Nilai yang sering muncul)

Nilai modus menggambarkan nilai sekumpulan data yang paling sering muncul atau memiliki frekuensi tertinggi dalam suatu kelompok. Rumus perhitungan modus untuk data berkelompok sebagai berikut:

$$SD = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N - 1}}$$

Keterangan:

Tb : Tepi bawah kelas modus (batas bawah kelas dikurangi 0,5).

$d1$: Selisih frekuensi kelas modus dengan kelas sebelumnya.

$d2$: Selisih frekuensi kelas modus dengan kelas setelahnya.

p : Panjang interval kelas

d) Nilai Maksimum dan Minimum

Nilai maksimum menggambarkan nilai tertinggi dari suatu kelompok, sedangkan nilai minimum menggambarkan nilai terkecil dari suatu kelompok. Rumus perhitungan nilai maksimum dan minimum sebagai berikut:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

e) Standar Deviasi

Standar deviasi menggambarkan tingkat penyebaran skor individu terhadap nilai rata-rata. Nilai ini penting untuk mengetahui tingkat heterogenitas kemampuan mahasiswa. Variasi data yang terlalu besar dapat memengaruhi stabilitas hasil analisis inferensial. Rumus perhitungan standar deviasi sebagai berikut:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Keterangan:

S = Standar deviasi

N = Jumlah data

X_i = Nilai X ke I sampai ke-n

$\bar{x}$ = Nilai rata-rata x

3. Uji Prasyarat Analisis (Normalitas & Homogenitas)

Uji prasyarat analisis dilakukan untuk memastikan bahwa data memenuhi asumsi penggunaan teknik statistik tertentu. Tahap ini merupakan bentuk kontrol metodologis agar pemilihan uji statistik sesuai dengan karakteristik data empiris. Dalam penelitian ini, uji prasyarat analisis yang digunakan adalah uji normalitas dan homogenitas.

a) Uji Normalitas

Uji normalitas bertujuan untuk mengidentifikasi data penelitian bersifat normal atau tidak. Distribusi normal menjadi asumsi penting dalam statistik parametrik karena berkaitan dengan akurasi estimasi parameter populasi. Rumus perhitungan uji normalitas sebagai berikut:

1) Uji Chi-Kuadrat (Chi-Square Goodness of Fit)

Uji Chi-Kuadrat digunakan untuk menguji kesesuaian distribusi data empiris dengan distribusi normal teoritis. Teknik ini umumnya diterapkan pada data yang telah dikelompokkan dalam bentuk tabel distribusi frekuensi. Secara teoritik, uji ini membandingkan frekuensi hasil pengamatan dengan frekuensi yang diharapkan apabila data benar-benar mengikuti distribusi normal. Semakin kecil selisih antara keduanya, semakin besar kemungkinan data berdistribusi normal. Rumus perhitungan uji normalitas menggunakan chi-kuadrat sebagai berikut:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Keterangan:

O_i = frekuensi observasi (observed frequency)

E_i = frekuensi harapan (expected frequency)

i = kelas interval data

2) Uji Kolmogorov-Smirnov (K-S)

Uji Kolmogorov-Smirnov merupakan metode nonparametrik yang digunakan untuk membandingkan distribusi kumulatif empiris data dengan distribusi kumulatif teoritis normal. Uji ini banyak digunakan pada data tunggal dengan ukuran sampel relatif besar. Rumus perhitungan uji normalitas menggunakan Kolmogorov-smirnov sebagai berikut:

$$D = \max |F_o(x) - F_t(x)|$$

Keterangan:

D = deviasi maksimum

$F_o(x)$ = distribusi kumulatif empiris

$F_t(x)$ = distribusi kumulatif teoritis normal

3) Uji Shapiro-Wilk

Uji Shapiro Wilk merupakan metode yang sangat direkomendasikan terutama untuk ukuran sampel kecil hingga menengah. Secara teoritis, uji ini memiliki kekuatan statistik yang lebih tinggi dibandingkan metode normalitas lainnya. Rumus perhitungan uji normalitas menggunakan Shapiro-Wilk sebagai berikut:

$$W = \frac{(\sum_{i=1}^n a_i X_{(i)})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Keterangan:

W = Nilai statistik saphiro-wilk

X_i = Data sampel ke-i yang telah diurutkan dari terkecil hingga terbesar

a_i = Koefisien Shapiro-Wilk dari tabel koefisien berdasarkan ukuran sampel

$\bar{X}$ = Nilai rata-rata sampel

Berikut merupakan uraian hipotesis yang digunakan dalam uji normalitas:

H_0 : Data sampel berasal dari populasi yang berdistribusi normal.

H_1 : Data sampel berasal dari populasi yang tidak berdistribusi normal.

b) Uji Homogenitas

Uji homogenitas bertujuan mengetahui kesamaan varians antar kelompok. Kesamaan varians menunjukkan bahwa kelompok eksperimen dan kontrol memiliki tingkat variasi kemampuan yang relatif sebanding, sehingga perbandingan hasil belajar menjadi lebih valid. Tahap prasyarat ini memastikan bahwa analisis selanjutnya dilakukan secara tepat dan tidak menghasilkan kesimpulan yang bias. Rumus perhitungan uji homogenitas sebagai berikut.

$$F = \frac{S_1^2}{S_2^2}$$

Keterangan:

F = Nilai F hitung

S_1^2 = Varians kelompok besar

S_2^2 = Varians kelompok kecil

Berikut merupakan uraian hipotesis yang digunakan dalam uji normalitas:

H_0 : Distribusi variabel kategorikal homogen untuk setiap subkelompok atau populasi.

H₁: Distribusi variabel kategorikal heterogen untuk setiap subkelompok atau populasi.

4. Uji Kesetaraan Kemampuan Awal (Pretest)

Uji kesetaraan kemampuan awal (pretest) diperlukan pengujian kesetaraan kemampuan awal untuk memastikan bahwa kedua kelompok memiliki kondisi awal yang relatif sama. Pengujian dilakukan menggunakan skor pretest untuk mengontrol kemungkinan bias seleksi. Untuk menguji kemampuan awal ada kriteria inclusion yaitu menggunakan kriteria inklusi operasional untuk mendapatkan kelas-kelas yang memiliki karakteristik homogen dan setara untuk diperbandingkan (*comparable*):

- a. Kriteria Inclusion 1 (Keseragaman Kurikulum): Kelas-kelas yang dipilih wajib menggunakan silabus, materi pokok, dan draf Rencana Pembelajaran Semester (RPS) PAI yang seragam di bawah naungan koordinator MKWK yang sama.
- b. Kriteria Inclusion 2 (Homogenitas Bebas SKS): Mahasiswa di dalam kelas yang dipilih wajib memiliki beban studi aktif semesteran dalam rentang normal (18-24 SKS) untuk meminimalkan variabel pengganggu berupa bias kelelahan kognitif. Kelas khusus karyawan dikeluarkan dari daftar seleksi.
- c. Kriteria Inclusion 3 (Standardisasi Instruksional Dosen): Kelas-kelas yang dipilih harus diampu oleh dosen yang sama atau tim dosen (*team teaching*) yang memiliki kualifikasi pedagogis dan standar rubrik penilaian yang identik, guna mengeliminasi bias *Teacher Effect*.
- d. Kriteria Inclusion 4 (Homogenitas Kemampuan Awal): Mahasiswa wajib memiliki kesetaraan tingkat kemampuan dasar keagamaan awal, yang dibuktikan secara empiris melalui nilai *pretest* yang menghasilkan nilai $p > 0,05$ pada uji homogenitas varians.

5. Uji Efektivitas (N-Gain/Effect Size)

Uji N-Gain (*Normalized Gain*) adalah metode analisis untuk mengukur efektivitas perlakuan melalui metode/media pembelajaran dengan membandingkan selisih nilai *posttest* dan *pretest* terhadap skor maksimum, yang menunjukkan

peningkatan pemahaman mahasiswa. Uji besarnya peningkatan menggunakan skor gain ternormalisasi (N-Gain) menurut Hake (1998), dengan rumus:

$$N\text{-Gain } (g) = (Skor\ posttest - Skor\ pretest) / (Skor\ maksimum - Skor\ pretest)$$

Kategori N-Gain mengikuti ketentuan pada table dibawah ini.

Tabel 3. 10 Kategori Skor N-Gain

Nilai N-Gain (g)	Kategori
$g \geq 0,70$	Tinggi
$0,30 \leq g < 0,70$	Sedang
$g < 0,30$	Rendah

Sumber: Hake (1998).

Tabel 3. 11 Tafsiran Tingkat Efektivitas Berdasarkan Persentase N-Gain

Persentase N-Gain	Tafsiran Efektivitas
Lebih dari 76 persen	Efektif
56 sampai 75 persen	Cukup efektif
40 sampai 55 persen	Kurang efektif
Kurang dari 40 persen	Tidak efektif

Kedua kerangka tafsiran tersebut dilaporkan bersama-sama pada Bab IV agar pembaca dapat menilai peningkatan dari dua sudut sekaligus, yaitu besaran gain sebagai proporsi dan tingkat keefektifan sebagai kategori pedagogis. Perlu dicatat bahwa N-Gain merupakan ukuran deskriptif atas besar peningkatan dan bukan uji signifikansi, sehingga simpulan mengenai perbedaan efektivitas antarkelompok tetap harus ditegakkan melalui pengujian inferensial pada tahap berikutnya.

6. Uji Hipotesis (Perbedaan/Peningkatan hasil belajar)

Pengujian hipotesis pertama dan hipotesis kedua dilakukan melalui uji Wilcoxon *Signed Ranks*, yaitu uji non-parametrik untuk dua sampel berpasangan yang menguji signifikansi perubahan skor dari *pretest* ke *posttest* pada masing-masing kelompok.²⁰ Uji ini dipilih menggantikan uji-t sampel berpasangan karena asumsi normalitas tidak dipenuhi. Luaran yang dilaporkan meliputi distribusi arah perubahan berupa peringkat negatif, peringkat positif, dan angka kembar, nilai Z,

²⁰ Frank Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin* 1, no. 6 (1945): 80-83.

nilai signifikansi asimtotik dua sisi, serta besaran pengaruh r . Hipotesis nol ditolak apabila nilai signifikansi lebih kecil dari 0,05, yang berarti terdapat peningkatan skor literasi keagamaan yang signifikan pada kelompok yang diuji.

Satu batas penafsiran perlu ditetapkan pada bagian ini. Uji Wilcoxon menjawab pertanyaan apakah terjadi peningkatan, tetapi tidak menjawab pertanyaan apakah suatu perlakuan lebih unggul daripada perlakuan lain. Kelompok kontrol pun berpeluang mengalami peningkatan yang signifikan, sebab pembelajaran konvensional tetap merupakan pembelajaran yang menyampaikan materi yang sama. Karena itu signifikansi pada uji Wilcoxon tidak boleh dijadikan dasar untuk menyatakan keunggulan, dan analisis wajib dilanjutkan pada pengukuran besar peningkatan serta perbandingan antarkelompok.

7. Uji Perbedaan Efektivitas Antarkelompok

Pengujian hipotesis ketiga dilakukan melalui uji Kruskal-Wallis sebagai uji simultan atas ketiga kelompok, yang dilanjutkan uji Mann-Whitney sebagai uji lanjut berpasangan.²¹ Uji Kruskal-Wallis dipilih menggantikan analisis varians satu jalur karena asumsi normalitas dan homogenitas varians tidak dipenuhi, sedangkan uji Mann-Whitney dipilih menggantikan uji-t sampel bebas atas dasar pertimbangan yang sama.

Satu keputusan analitis perlu dijelaskan alasannya karena berpengaruh langsung pada penafsiran. Pengujian antarkelompok dilakukan terhadap skor selisih, yaitu selisih antara skor *posttest* dan skor *pretest* setiap mahasiswa, dan bukan terhadap skor *posttest* semata. Alasannya adalah bahwa skor selisih memisahkan besar perubahan dari perbedaan titik awal, sehingga perbandingan yang dihasilkan merupakan perbandingan efektivitas perlakuan dan bukan perbandingan capaian akhir yang masih mengandung pengaruh kondisi awal. Hipotesis nol ditolak apabila nilai signifikansi asimtotik lebih kecil dari 0,05, yang berarti terdapat perbedaan efektivitas yang signifikan pada sekurang-kurangnya satu pasang kelompok. Setiap hasil uji dilengkapi besaran pengaruh. Besaran

²¹ William H. Kruskal dan W. Allen Wallis, "Use of Ranks in One-Criterion Variance Analysis," *Journal of the American Statistical Association* 47, no. 260 (1952): 583-621. H. B. Mann dan D. R. Whitney, "On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other," *Annals of Mathematical Statistics* 18, no. 1 (1947): 50-60.

pengaruh untuk uji non-parametrik berbasis peringkat dihitung dengan rumus berikut.²²

$$r = \text{nilai mutlak } Z \text{ dibagi akar } N$$

Keterangan: r adalah besaran pengaruh Z adalah nilai statistik uji yang dihasilkan uji Wilcoxon atau Mann-Whitney dan N adalah jumlah observasi yang terlibat dalam perbandingan. Nilai r ditafsirkan mengikuti baku Cohen, yaitu 0,10 untuk pengaruh kecil, 0,30 untuk pengaruh sedang, dan 0,50 untuk pengaruh besar.²³ Pelaporan besaran pengaruh di samping nilai signifikansi merupakan praktik pelaporan statistik yang dianjurkan, sebab pada ukuran sampel besar nilai signifikansi mudah tercapai sehingga tanpa besaran pengaruh pembaca tidak memperoleh gambaran mengenai kebermaknaan praktis temuan.²⁴

Sebagai pelengkap deskriptif, selisih rerata *posttest* antarkelompok juga dilaporkan. Namun besaran pengaruh parametrik atas selisih rerata tersebut tidak dijadikan tumpuan simpulan mengingat asumsi normalitas tidak dipenuhi, sehingga seluruh penarikan simpulan dalam penelitian ini bersandar pada uji non-parametrik beserta besaran pengaruh r .

G. Sinkronisasi Rancangan Metode dengan Penyajian Hasil

Sebagai penutup bab ini, disajikan pemetaan yang menautkan setiap hipotesis dengan jenis data, uji statistik, kriteria keputusan, dan letak sajian hasilnya pada Bab IV. Pemetaan ini berfungsi sebagai alat kendali agar tidak ada prosedur yang ditetapkan pada bab metode namun tidak dilaporkan hasilnya, dan sebaliknya tidak ada hasil yang dilaporkan tanpa dasar prosedur yang telah ditetapkan.

²² Robert Rosenthal, "Parametric Measures of Effect Size," dalam *The Handbook of Research Synthesis*, ed. Harris Cooper dan Larry V. Hedges (New York: Russell Sage Foundation, 1994), 231-244.

²³ Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 79-81.

²⁴ Andy Field, *Discovering Statistics Using IBM SPSS Statistics*, ed. ke-5 (London: Sage Publications, 2018), 90-96.

Tabel 3. 12 Pemetaan Hipotesis, Uji Statistik, dan Letak Sajian Hasil pada Bab IV

Hipotesis	Data yang Dianalisis	Uji Statistik	Kriteria Keputusan	Letak Sajian pada Bab IV
Kelayakan produk sebagai prasyarat pengujian	Skor penilaian lima validator pada skala empat titik	Rerata skor dan konversi kategori kelayakan Akbar	Kategori sekurang-kurangnya layak dan catatan validator ditindaklanjuti	Subbab A tentang Validitas Instrumen dan Produk
Kesetaraan kemampuan awal	Skor pretest ketiga kelompok	Statistik deskriptif dan uji beda non-parametrik	Rerata berdekatan, selang kepercayaan bertumpang tindih, signifikansi lebih besar dari 0,05	Subbab B tentang Deskripsi Hasil Pretest dan Posttest
Prasyarat pemilihan uji	Skor pretest dan posttest keenam sebaran	Kolmogorov-Smirnov dengan koreksi Lilliefors, Shapiro-Wilk, dan Levene	Signifikansi lebih kecil dari 0,05 mengarahkan pada statistik non-parametrik	Subbab C angka 1 tentang Uji Prasyarat Analisis
Hipotesis 1: pengaruh Modul Digital	Skor pretest dan posttest kelompok Modul Digital	Wilcoxon Signed Ranks dengan besaran pengaruh r	Hipotesis nol ditolak jika signifikansi lebih kecil dari 0,05	Subbab C angka 2 tentang Uji Hipotesis dalam Kelompok
Hipotesis 2: pengaruh Video PjBL	Skor pretest dan posttest kelompok Video PjBL	Wilcoxon Signed Ranks dengan besaran pengaruh r	Hipotesis nol ditolak jika signifikansi lebih kecil dari 0,05	Subbab C angka 2 tentang Uji Hipotesis dalam Kelompok
Besar peningkatan setiap kelompok	Skor N-Gain individual ketiga kelompok	Rerata N-Gain, kategori Hake, dan tafsiran persentase efektivitas	Kategori efektivitas menurut Tabel 3.10 dan Tabel 3.11	Subbab C angka 3 tentang Uji Efektivitas melalui N-Gain
Hipotesis 3: perbedaan efektivitas antarkelompok	Skor selisih posttest dikurangi pretest ketiga kelompok	Kruskal-Wallis dilanjutkan Mann-Whitney dengan besaran pengaruh r	Hipotesis nol ditolak jika signifikansi lebih kecil dari 0,05	Subbab C angka 4 tentang Uji Perbedaan Efektivitas Antarkelompok
Keterlaksanaan dan kesetaraan perlakuan	Lembar Observasi Kesetaraan Perlakuan dari dua pengamat independen	Telaah deskriptif kualitatif	Seluruh butir kesetaraan tercatat setara kecuali jenis media dan model pembelajaran	Subbab A angka 5 tentang Kesetaraan Perlakuan dan Peran Observer Independen