



# TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

---

## PENGEMBANGAN SOAL BERBASIS HOTS MODUL PEDAGOGIK PENDIDIKAN PROFESI GURU DALAM JABATAN LPTK IAIN GORONTALO

**Lian Gafar Otaya**

IAIN Sultan Amai Gorontalo

*Email: lianotaya82@iaingorontalo.ac.id*

### ABSTRAK

Penelitian ini bertujuan untuk mengembangkan soal berbasis HOTS. Soal yang dikembangkan berupa soal-soal formatif dari kegiatan belajar pada modul pedagogik pendalaman materi yang disusun oleh dosen/instruktur/pengajar PPG Dalam Jabatan Kementerian Agama RI di LPTK IAIN Gorontalo. Jenis penelitian ini termasuk penelitian dan pengembangan *Research and Development* dengan model pengembangan tipe *formative research* yaitu: 1) tahap *analyse*, 2) tahap *design*, 3) tahap *develop*, 4) tahap *formatively evaluate*, dan 5) tahap *summatively evaluate*. Teknik analisis data secara kuantitatif untuk menganalisis kecocokan model atau model fit dan penilaian terhadap karakteristik butir soal yang dikembangkan butir menggunakan Program BILOG. Untuk validitas isi ditelaah pakar dengan menggunakan formula Gregory. Hasil temuan diperoleh bahwa item soal tes berbasis HOTS untuk modul pedagogik dikembangkan dari dimensi proses menganalisis (C4), mengevaluasi (C5), dan mencipta (C6) memiliki kriteria valid, memiliki indeks tingkat kesukaran sedang dan indeks daya beda yang baik. Sehingga telah memenuhi semua kriteria pengembangan instrumen soal yang baik. Artinya soal ini dapat digunakan untuk mengukur pemahaman mahasiswa PPG pada pendalaman materi modul pedagogik.

**Kata kunci** : Tes, HOTS, Modul Pedagogik, PPG

### ABSTRACT

*This research aims to develop HOTS-based questions. The questions developed are in the form of formative questions from learning activities in the pedagogical module for deepening the material prepared by PPG lecturers/instructors/teachers in the position of the Indonesian Ministry of Religion at LPTK IAIN Gorontalo. This type of research includes Research and Development research and development with a formative research type development model, namely: 1) analysis stage, 2) design stage, 3) develop stage, 4) formative evaluate stage, and 5) summative evaluate stage. Quantitative data analysis techniques to analyze model suitability or model fit and assess the characteristics of the items developed using the BILOG*



# TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

---

*Program. For content validity, it was reviewed by experts using the Gregory formula. The findings showed that the HOTS-based test items for the pedagogical module developed from the process dimensions of analyzing (C4), evaluating (C5), and creating (C6) had valid criteria, had a moderate level of difficulty index and a good distinguishing power index. So it has fulfilled all the criteria for developing a good question instrument. This means that this question can be used to measure PPG students' understanding of deepening the pedagogical module material.*

**Keywords:** *Test, HOTS, Pedagogical Module, PPG*

## PENDAHULUAN

Pendidikan Profesi Guru dalam Jabatan atau disingkat PPG Daljab adalah salah satu program yang dilakukan oleh pemerintah termasuk Kementerian Agama RI sebagai upaya untuk meningkatkan kompetensi dan kualitas para guru yang sudah mengajar di satuan pendidikan di bawah naungan Kementerian Agama RI agar dapat memberikan pendidikan yang lebih baik kepada peserta didiknya.<sup>1</sup> Diharapkan dengan meningkatnya profesional guru dalam mengajar maka akan berdampak pada meningkatnya kualitas pendidikan di Indonesia, sehingga dengan begitu akan dapat menciptakan lulusan yang baik dan berkualitas.<sup>2</sup> Peserta program ini biasanya telah memiliki pengalaman mengajar dan telah memenuhi syarat-syarat tertentu yang ditetapkan oleh Kementerian Agama RI dan sistem penyelenggaraannya dilakukan dengan memperhatikan prinsip-prinsip yang mendasarinya berdasarkan lingkup capaian kompetensi (capaian pembelajaran) peserta/mahasiswa Program PPG selama mengikuti program pembelajaran.

Pelaksanaan kegiatan PPG Daljab oleh Kementerian Agama RI harus dievaluasi efektifitasnya dalam meningkatkan kualitas guru, yang salah satunya adalah ditandai dengan penguasaan para mahasiswa PPG terhadap semua materi dan pengalaman belajar yang mereka alami selama mengikuti kegiatan PPG secara daring.<sup>3</sup> Penilaian menjadi isu yang sangat penting dalam proses capaian pembelajaran PPG daljab Kementerian Agama RI bagi guru Madrasah dan Pendidikan Agama Islam (PAI) yang pembelajarannya dilakukan secara online atau

---

<sup>1</sup> Maulana, Iqbal, Nia Atikah Rahma, Namira Fitri Mahfirah, Wahyu Alfarizi, and Ahmad Darlis. "Meningkatkan profesional guru dengan program pendidikan profesi guru (PPG)." *Journal on Education* 5, no. 2 (2023), h. 2158-2167.

<sup>2</sup> Hanun, Farida. "Implementasi penyelenggaraan program pendidikan profesi guru (PPG) pendidikan agama islam di LPTK UIN Serang Banten." *EDUKASI: Jurnal Penelitian Pendidikan Agama Dan Keagamaan* 19, no. 3 (2021), h. 268-285.

<sup>3</sup> Fahrurrozi. "Pengaruh Kualitas LMS dan Prilaku Belajar terhadap Kualitas Belajar Mahasiswa Ppg." *Jurnal Muara Pendidikan* 7, no. 1 (2022), h. 131-136.



## TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

---

dalam jaringan (daring) dengan memanfaatkan teknologi informasi dan komunikasi melalui LMS (*Learning Management System*) menggunakan aplikasi Space (<https://ppg.siagapendis.com/>).

Salah satu komponen yang dinilai dalam proses penyelenggaraan PPG adalah modul pendalaman materi yang terdiri dari 6 modul professional/bidang studi dan 4 modul pedagogik yang bertujuan untuk memperkuat pemahaman mahasiswa PPG tentang bidang studi, pedagogi dan keprofesian guru. Pendalaman materi modul professional/bidang studi termasuk materi esensial yang mencakup *advance materials* bidang studi yang dapat menjelaskan aspek “apa” (*isi/content*), aspek “mengapa” (filosofi), dan “bagaimana” (aplikasi/penerapan) dalam kehidupan keseharian. Selanjutnya pendalaman materi pedagogi terkait perencanaan pembelajaran, pelaksanaan hingga evaluasi pembelajaran secara terstruktur, terencana, dan mengikuti prosedur baku mulai dari penyusunan kisi-kisi sampai proses penentuan hasil penilaian. Tahapan ini merupakan tahapan yang penting dalam mengembangkan kemampuan guru agar menjadi seorang guru yang professional.<sup>4</sup> Oleh karena itu, perlu didukung dengan pemahaman yang mendalam dan penguasaan yang luas dari mahasiswa PPG tentang modul pendalaman materi tersebut.

Guna mengukur dan menilai pemahaman dan penguasaan mahasiswa PPG tentang modul pendalaman materi dilihat dari Nilai Akhir Modul (NAM) yang diperoleh dari: 1) Resume dan Refleksi Modul (30%) yang berasal dari rata-rata nilai tugas resume Modul dihitung dari nilai total resume Modul mulai dari kegiatan belajar (KB) 1 sampai KB 4 kemudian dibagi 4; 2) Analisa bahan ajar (15%) berasal dari rata-rata nilai tugas Analisa bahan ajar dari KB 1 sampai KB 4 kemudian dibagi 4; 3) Tes Formatif (10%) yang diperoleh dari rata-rata nilai Tes Formatif dihitung dari nilai total tes formatif Modul mulai dari KB 1 sampai KB 4 kemudian dibagi 4; 4) Analisa materi (35%) dengan memilih pendekatan *Problem Based Learning* (PBL) untuk modul pedagogik dan pendekatan *Project Based Learning* (PjBL) untuk modul profesional; dan 5) Tes Akhir Modul (20%) dengan nilai minimal capaian pembelajaran modul 70. Jika nilai yang diperoleh peserta/mahasiswa Program PPG kurang dari batas minimal nilai kelulusan per modul, maka dilakukan perbaikan nilai capaian pembelajaran modul dengan mengerjakan tugas tambahan/pengganti. Berat ringan tugas disesuaikan dengan selisih dari batas minimal nilai kelulusan setiap modul.

---

<sup>4</sup> Maryani, Elly. "Pendidikan profesi guru (PPG) sebagai cara untuk menjadikan guru menjadi profesional." *GUAU: Jurnal Pendidikan Profesi Guru Agama Islam* 2, no. 4 (2022), h. 171-178.



## TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

---

Tes formatif sebagai salah satu alat ukur yang digunakan untuk mengetahui capaian kemampuan mahasiswa PPG dalam memahami dan menguasai modul pada pendalaman materi yang diberikan oleh dosen/instruktur/pengajar modul dan penentu kelulusan per modul tentu saja membutuhkan instrumen tes yang baik untuk menilainya. Salah satunya berisi soal-soal berbasis *Higher Order Thinking Skills* (HOTS). Soal berbasis HOTS dapat menjamin penilaian yang dilakukan mencerminkan prinsip objektif, karena mampu menggambarkan kemampuan dari peserta tes sesuai dengan keadaan yang sebenarnya dan jika penilaian yang dilakukan objektif tentunya dapat menjamin akuntabilitas penilaian.<sup>5</sup> Oleh karena itu, para dosen/instruktur/pengajar modul diarahkan dalam penyusunan soal formatif dapat mengembangkan keterampilan berpikir tingkat tinggi (HOTS). Artinya tidak hanya mampu membuat soal-soal pada level C1 (mengingat), C2 (memahami), C3 (mengaplikasikan), tetapi juga pada level C4 (menganalisis), C5 (mengevaluasi), dan C6 (mencipta) karena dengan mengerjakan soal berbasis HOTS diharapkan dapat menstimulasi mahasiswa PPG berpikir secara luas dan mendalam untuk menjawab soal-soal yang lebih sulit dan mampu memecahkan suatu kasus masalah yang lebih kompleks tentang modul pendalaman materi yang dibelajarkan.

Berbagai penelitian menunjukkan bahwa pengembangan soal berbasis HOTS sangat penting dilakukan untuk mengarahkan pada proses berpikir kritis, logis, reflektif, metakognitif dan kreatif karena melibatkan proses bernalar dalam pengerjaannya. Diantaranya hasil penelitian Hakim, Ardiyanto & Wijayanto (2021)<sup>6</sup> bahwa jika secara bertahap pendidik mengembangkan instrumen soal pada level tingkat tinggi (tingkatan kognitif C4-C6) dan mengaplikasikannya dalam proses pembelajaran, maka kemampuan peserta didik dalam menjawab soal-soal tingkat tinggi akan meningkat. Menurut Suhady, Roza & Maimunah (2020)<sup>7</sup> bahwa supaya kemampuan HOTS bisa berkembang dengan baik maka perlu pembiasaan dengan kegiatan yang melatih HOTS itu sendiri. Soal untuk mengukur HOTS dikembangkan dengan memperhatikan karakteristik HOTS yaitu sesuai dengan indikator kemampuan menganalisis, berpikir kritis dan kemampuan berpikir

---

<sup>5</sup> Supriadi, Gito. *Pengembangan instrumen penilaian berbasis higher order thinking skill (HOTS)*. (Yogyakarta: Aswaja Pressindo, 2020), h. 4.

<sup>6</sup> Hakim, Bahtiar Rahman, Asep Ardiyanto, and Arfilia Wijayanti. "Analisis hots pada instrumen penilaian siswa kelas iv sekolah dasar." *Wawasan Pendidikan* 1, no. 2 (2021), 246-254.

<sup>7</sup> Suhady, Wandy, Yenita Roza, and Maimunah Maimunah. "Pengembangan soal untuk mengukur higher order thinking skill (hots) siswa." *Jurnal Gantang* 5, no. 2 (2020), h. 143-150.



## TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

---

kreatif.<sup>8</sup> Hal tersebut perlu diperhatikan agar soal HOTS yang dikembangkan benar-benar layak untuk digunakan dan memenuhi kriteria. Permasalahan yang terjadi di sekolah, soal-soal cenderung lebih banyak menguji aspek ingatan<sup>9</sup> yang kurang melatih keterampilan berpikir tingkat tinggi siswa.<sup>10</sup> Abosalem (2016)<sup>11</sup> menyatakan kebutuhan untuk meningkatkan keterampilan berpikir tingkat tinggi siswa sangat diperlukan untuk mendiagnosis tingkat berpikir siswa, memberikan umpan balik kepada siswa tentang tingkat pemikiran mereka dan mendorong mereka untuk berpikir dengan cara yang lebih baik.

Permasalahan yang terjadi pada pemberian tes formatif untuk soal-soal yang diupload dosen/instruktur/pengajar PPG pada modul pendalaman materi melalui akun LMS Space pada pelaksanaan PPG Dalam Jabatan Kementerian Agama RI Tahun 2021 baik untuk guru madrasah maupun PAI di Lembaga Pendidikan Tenaga Kependidikan (LPTK) IAIN Sultan Amai Gorontalo cenderung masih pada level berpikir tingkat rendah atau bersifat *Lower Order Thinking Skills* (LOTS) dengan persentase 66.83%, untuk level *Middle Order Thinking Skills* (MOTS) dengan persentase 21.01% dan untuk soal HOTS dengan persentase rendah yaitu hanya 12.16%, karena sebagian besar masih belum memahami bagaimana menyusun soal-soal yang berbasis HOTS. Meskipun dosen/instruktur/pengajar PPG di LPTK IAIN Sultan Amai Gorontalo sudah dibekali dengan pemberian bimbingan teknis (Bimtek) cara penyusunan soal berbasis HOTS mulai dari mengenali karakteristik soal yang berbasis HOTS, bagaimana pengembangannya hingga hasil analisisnya, namun dianggap belum efektif karena sebagian besar mengakui mereka masih mengalami kesulitan dalam memahaminya. Hal ini dibuktikan dengan hasil pengisian evaluasi pemahaman dosen setelah mengikuti bimteks cara penyusunan soal berbasis HOTS melalui *google form* ditemukan dari 34 peserta Bimteks terdapat 23 peserta atau 67% mengalami kesulitan menerapkan kriteria-kriteria HOTS dalam menyusun instrumen soal pilihan ganda terutama

---

<sup>8</sup> Kurniawati, Rissa Prima, & Fida Rahmantika Hadi. "Pelatihan pengembangan instrumen evaluasi berbasis HOTS untuk guru sekolah dasar." *Jurnal Altifani Penelitian dan pengabdian kepada masyarakat* 1, no. 4 (2021): 267-276.

<sup>9</sup> Fajarini, Indah, Coryna Oktaviani, and Dini Fitria. "Meningkatkan kemampuan paedagogik guru melalui pelatihan pembuatan soal ujian berbasis pembelajaran multiliterasi terintegrasi higher order thinking skills di SD Negeri 2 Seulalah Langsa." *Jurnal Ilmiah Wahana Pendidikan* 8, no. 12 (2022), h. 138-146.

<sup>10</sup> Maulina, Winda. "Pengembangan Instrumen Asesmen Berbasis Hots pada Mata Pelajaran Pendidikan Agama Islam dan Budi Pekerti di SMAN 6 Padang." *Jurnal Pendidikan dan Konseling (JPDK)* 4, no. 6 (2022), h. 7317-7326.

<sup>11</sup> Abosalem, Yousef. "Assessment techniques and students' higher-order thinking skills." *International Journal of Secondary Education* 4, no. 1 (2016), h. 1-11.



# TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

menyusun stimulus soal HOTS dan mengakui kurang mengetahui karakteristik soal HOTS.<sup>12</sup>

Mencermati permasalahan di atas, diperlukan adanya solusi dalam mengatasi permasalahan yang dihadapi dosen/instruktur/pengajar PPG di LPTK IAIN Sultan Amai Gorontalo melalui pengembangan bank soal berbasis HOTS untuk tes formatif dari modul yang dibelajarkannya difokuskan pada model bidang pedagogik yaitu modul evaluasi pembelajaran. Oleh karena itu, penelitian ini bertujuan mengembangkan soal berbasis HOTS untuk tes formatif dari modul pendalaman materi Evaluasi Pembelajaran.

## METODE PENELITIAN

Jenis penelitian ini termasuk penelitian dan pengembangan *Research and Development* yang disingkat R & D. Model R & D yang digunakan dalam penelitian ini model pengembangan tipe *formative research* dengan 5 (lima) tahap yaitu: 1) tahap *analyse*, 2) tahap *design*, 3) tahap *develop*, 4) tahap *formatively evaluate*, dan 5) tahap *summatively evaluate*. Kelima tahapan tersebut disajikan seperti pada gambar berikut:



Gambar 1. Prosedur model pengembangan tipe formative research Tessmer

<sup>12</sup> Data Hasil Evaluasi Penilaian Bimtek Penyusunan Soal Berbasis HOTS Pada PPG Dalam Jabatan Kementerian Agama RI Tahun 2021 LPTK IAIN Sultan Amai Gorontalo melalui Google Form: Selasa, 1 Juni 2021



# TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

---

## 1. Tahap analisis (*Analyse*)

Tahap ini diawali mengidentifikasi masalah yang mendasar dihadapi dosen/instruktur/pengajar modul pedagogik pada pendalaman materi di LPTK IAIN Sultan Amai Gorontalo melalui *focus discussion group* (FGD) untuk menentukan tujuan spesifikasi penyusunan instrument tes. Selanjutnya dilakukan telaah untuk menganalisis tema yang relevan atau cakupan materi yang dikembangkan dalam membuat soal HOTS berdasarkan modul pendalaman materi yang diajarkan dosen/instruktur/pengajar untuk penentuan indikator soal dan penyusunan kisi-kisi soal.

## 2. Tahap perancangan (*Design*)

Tahap ini diawali dengan penentuan indikator soal yang dilanjutkan dengan penentuan jenis soal yang dibuat dalam bentuk *multiple choice* dengan 4 alternatif pilihan jawaban yang terdiri dari 1 kunci jawaban dan 3 distraktor (pengecoh). Langkah selanjutnya kisi-kisi soal HOTS sesuai dengan indikator soal yang dibuat.

## 3. Tahap pengembangan (*Develop*)

Tahap ini dilakukan dengan menulis item pertanyaan soal sesuai dengan kisi-kisi yang telah dibuat yang terdiri dari pokok soal alternatif jawaban dengan jumlah soal yang dibuat per modul sebanyak 10 soal untuk total soal 4 modul yaitu 40 soal terdiri dari 10 soal modul Evaluasi Pembelajaran, 10 soal Modul Pengembangan Profesi Guru, 10 soal untuk modul Teori Belajar dan Pembelajaran dan 10 soal untuk modul Perkembangan Peserta Didik. Dilanjutkan dengan pembuatan petunjuk pengerjaan atau pemberian respons, instrumen penilaian dan pedoman penskoran atau kunci jawaban.

## 4. Tahap evaluasi formatif (*Formatively evaluate*)

Tahap ini dilakukan melalui tahap *One to One* dan tahap *expert review*. Pada tahap *One to One* melibatkan 10 orang mahasiswa/peserta PPG yang berkemampuan rendah, sedang dan tinggi. Soal yang telah disusun dosen/instruktur/pengajar PPG dibagikan pada mahasiswa/peserta PPG dengan waktu pengerjaan selama 60 menit. Setelah pengerjaan soal peneliti melakukan wawancara kepada masing-masing mahasiswa/peserta PPG untuk mengidentifikasi kendala yang dihadapinya dalam pengerjaan soal dan untuk menganalisis keterbacaan soal dari aspek bahasa, materi yang berhubungan dengan soal dan waktu yang dibutuhkan dalam menjawab soal. Selanjutnya untuk tahap *expert review* yaitu draft soal yang telah dikembangkan direviu *expert judgment* yang relevan dengan bidang soal yang direview. Setiap validator melakukan penilaian terhadap soal HOTS



## TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

---

yang dikembangkan dosen/instruktur/pengajar PPG berdasarkan aspek yang ditelaah dari materi, konstruksi dan Bahasa dengan mengacu pada indikator yang disajikan melalui lembar validasi. Validator juga menuliskan komentar dan saran terhadap soal yang direview. Apabila dari seluruh aspek materi, konstruksi, dan bahasa telah dipenuhi oleh butir soal, maka butir soal tersebut telah memenuhi kriteria telaah. Butir soal yang belum memenuhi kriteria penelaahan soal pilihan ganda, harus direvisi sesuai kesalahannya baik dalam aspek materi, konstruksi, maupun bahasa. Dengan catatan, soal yang direvisi dapat diperbaiki dan disempurnakan dengan benar sehingga ada kemungkinan soal menjadi baik. Selanjutnya dilakukan analisis validitas isi dengan formula Gregory.

$$\text{Validitas Isi} = \frac{D}{A+B+C+D}$$

Keterangan:

- A = Sel yang menunjukkan ketidaksesuaian antara kedua penilai
- B dan C = Sel yang menunjukkan perbedaan pandangan antara penilai
- D = Sel yang menunjukkan persetujuan yang valid antara kedua penilai

### 5. Tahap evaluasi sumatif (*Summatively evaluate*)

Pada tahap ini hasil draft instrument soal yang dibuat sudah didapat dilakukan uji coba terbatas untuk menghasilkan indikator yang kuat untuk mengukur masing-masing indikator sebagai pembentuk soal. Data hasil uji coba terbatas dianalisis untuk mengetahui validitas konstruk dan reliabilitasnya kemudian hasilnya direvisi berdasarkan hasil analisis. Selanjutnya dilanjutkan ujicoba diperluas dan dilakukan analisis yang sama untuk merakit instrument soal final dan menafsirkan hasil pengukurannya. Produk akhir yang dihasilkan adalah soal-soal formatif HOTS yang diarsipkan yang tentunya akan sangat membantu dosen/instruktur/pengajar modul pedagogik pendalaman materi PPG untuk mengukur kemampuan berpikir tingkat tinggi peserta didik secara riil karena pengembangan bank soal ini sudah mengacu pada standar pengembangan dan kriteria soal berbasis HOTS yang ditetapkan.

Teknik pengumpulan data dilakukan dengan menggunakan teknik dokumentasi. Data yang diambil adalah lembar jawaban dan soal tes formatif 4 modul pedagogik pada pendalaman materi pada modul peserta PPG Daljab Kementerian Agama tahun 2022. Menganalisis kualitas soal secara teoritis, digunakan lembar telaah butir soal guna memperoleh data tentang butir soal yang



## TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

---

memenuhi kriteria dilihat dari aspek materi, konstruksi, dan bahasa. Instrumen yang digunakan berupa lembar telaah butir soal yang ditetapkan oleh Pusat Penilaian Pendidikan (Puspendik) Kementerian Pendidikan Nasional. Teknik pengumpulan data terkait kualitas dan keefektifan produk soal yang dikembangkan peneliti menggunakan lembar observasi, angket atau kuesioner. Teknik dan instrumen disesuaikan dengan tahap uji yang dilakukan. Telaah pakar menggunakan metode FGD dan Delphi yang membutuhkan instrumen berupa angket untuk melihat kualitas dari produk yang dihasilkan, demikian juga untuk uji coba terbatas dan uji diperluas. Instrumen disesuaikan dengan tahap uji yang dilakukan.

Teknik analisis data dilakukan dengan berbagai metode yaitu langkah pertama adalah dengan melakukan analisis secara kualitatif untuk mengetahui validitas isi yang dianalisis oleh pakar. Telaah pakar (validitas isi) dengan menggunakan formula Gregory. Selanjutnya teknik deskriptif kuantitatif digunakan untuk mendeskripsikan masukan dari responden yang hadir pada acara FGD, yang bersifat kuantitatif serta penilaian respons terhadap produk yang dibuat yaitu: 1) Melakukan tabulasi semua data yang diperoleh dari validator untuk setiap aspek dan indikator dari butir penilaian yang tersedia dalam instrumen penilaian, 2) Menghitung skor total rata-rata dari setiap aspek dan indikator. Selanjutnya menganalisis karakteristik butir menggunakan Program BILOG. Setelah melakukan analisis kecocokan model atau model fit maka tahap selanjutnya yaitu penilaian terhadap karakteristik butir dalam instrumen yang dikembangkan.



# TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

## HASIL DAN PEMBAHASAN

Rancangan dasar dalam penelitian ini yaitu soal dikembangkan dari modul pedagogik PPG. Soal tes dikembangkan dari dimensi pengetahuan faktual, konseptual dan prosedural. Soal dikembangkan dari dimensi proses menganalisis (C4), mengevaluasi (C5), dan mencipta (C6). Selanjutnya soal dikembangkan dengan bentuk pilihan ganda berdasarkan kisi-kisi soal tes berbasis HOTS Taksonomi Bloom Revisi adalah sebagai berikut

Pada tahap selanjutnya yaitu melakukan perbaikan-perbaikan dalam pembuatan soal agar mendapatkan soal tes berbasis HOTS dapat layak digunakan. Selanjutnya dilakukan pengujian kecocokan model dengan menggunakan program BILOG. Setelah analisis data dengan program BILOG dilakukan maka pada output phase 2 dihasilkan statistik kecocokan suatu butir dengan model atau *goodness of fit statistic*. Kriteria yang digunakan dalam analisis kecocokan disini adalah Jika nilai  $\text{sig} < \alpha$ , maka butir dikatakan tidak cocok dengan model atau fit model. Artinya data analisis dengan model 1PL, 2PL, dan 3PL kecocokan semua butir dengan model berdasarkan kriteria  $p < \alpha$  (*item tidak cocok dengan model*) dan  $p \geq \alpha$  (*item cocok dengan model*), maka dibuat perbandingan hasil analisis kecocokan butir dengan menggunakan model 1PL, 2PL dan 3PL pada data soal tes formatif sebagaimana disajikan pada tabel 1 berikut ini.

Tabel 1

Hasil analisis kecocokan butir dengan menggunakan model 1PL, 2PL dan 3PL

Kriteria	1PL		2PL		3PL	
	No. Item	Jumlah	No. Item	Jumlah	No. Item	Jumlah
Fit Model	2,3,5,6,9,10,11	22	1,2,3,4,5,6,9,10 11,12,13,16,17	34	1,2,3,4,5,6,10,11 12,13,15,16,17,19	31
	12,13,17,21,22 24,26,29,30,32 34,35,37,38,40		19,20,21,22,23 24,25,26,27,28 29,30,31,32,33 34,35,37,38,40		20,22,23,24,25 26,28,29,30,31,32 34,35,37,38,39,40	
Tidak Fit Model	1,4,7,8,14,15,16	18	7,14,15,18 36,39	6	8,9,10,14,18,21 27,33,36	9
	18,19,20,23,25 27,28,31,33 36,39					

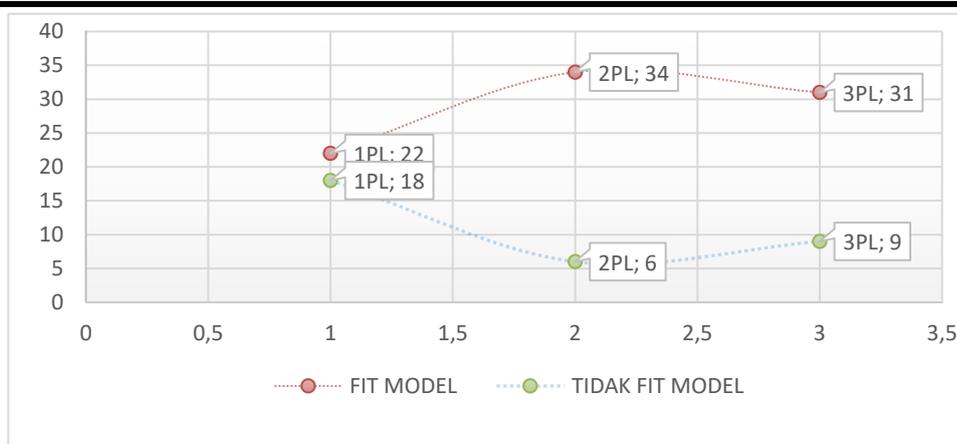
Tabel 1 menunjukkan hasil analisis kecocokan butir dengan menggunakan model 1PL, 2PL dan 3PL dan ternyata model yang menghasilkan butir yang cocok dengan model fit yang lebih banyak adalah model 2PL. Ini berarti model 2PL merupakan model yang dapat dipilih untuk analisis karakteristik butir. Untuk lebih



## TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174



jelasan butir yang cocok dengan model dibuat sebaran kecocokan butir dengan model sebagaimana disajikan pada gambar 1 berikut:

Mencermati gambar 1 diperoleh informasi bahwa model 2PL merupakan model yang paling baik dibandingkan dengan model 1PL dan 3PL. Hal ini dapat dilihat dari jumlah butir soal yang informasi estimasi tingkat kemampuan peserta (*measure of difficulty*), kesalahan pengukuran (*standard error of measurement*), kecocokan data dengan model (*infit* dan *outfit*), serta korelasi daya beda butir soal (*point biserial*) dari model 2PL paling banyak dibanding dengan model 1PL dan 3PL.

Berdasarkan hasil analisis kecocokan model tersebut, dapat disimpulkan bahwa di antara model 1 PL, 2 PL dan 3 PL yang digunakan untuk menganalisis hasil tes formatif pedagogik adalah model logistik yang paling sesuai untuk mengukur kesesuaian butir dengan model adalah 2PL, maka estimasi parameter butir dinyatakan dalam bentuk daya beda (a) dan tingkat kesulitan (b). Selanjutnya hasil estimasi parameter item dengan model 2PL yang dinyatakan dalam bentuk indeks daya beda (a) dan tingkat kesulitan (b) dijelaskan sebagai berikut.

### a. Indeks daya beda ( $a_i$ )

Parameter  $a_i$  merupakan indeks daya pembeda yang dimiliki butir ke- $i$ . Pada kurva karakteristik,  $a_i$  merupakan kemiringan (*slope*) dari kurva dititik  $b_i$  pada skala kemampuan tertentu, karena merupakan kemiringan, diperoleh semakin besar kemiringannya, maka semakin besar daya pembeda butir tersebut. Secara teoritis, nilai  $a_i$  terletak antara  $-\infty$  hingga  $+\infty$ . Pada butir yang baik nilai ini mempunyai hubungan positif dengan *performance* pada butir dengan kemampuan yang diukur, dan  $a_i$  terletak antara 0 dan 2.<sup>13</sup> Indeks daya beda dari output BILOG-MG dapat dilihat output phase2. Klasifikasi indeks daya beda item disajikan sebagai berikut:

<sup>13</sup> Hambleton, R.K & Swaminathan, H. *Item Response Theory*. (Boston, MA: Kluwer Inc, 1985), h. 37.



# TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

Tabel 1. Indeks Daya Beda Item

Tingkat Kesulitan ( $a_i$ )	Nomor Item	Jumlah		Kategori
		F	%	
$-2 \leq a_i \leq 2.0$	1,2,3,4,5,6,7,8,9,10,11	7	82.5	Baik
	12,13,14,15,16,17,18,19			
	20,21,22,23,24,26,29,30			
	32,34,35,37,38,40			
$a_i > 2.0$	25,27,28,31,33,36,39	33	17.50	Kurang Baik
Total		40	100	

Berdasarkan tabel 2 diperoleh informasi bahwa 82,5% item soal memiliki daya pembeda item yang baik. Hal ini membuktikan bahwa 82,5% dari keseluruhan item-item tes soal memiliki kemampuan untuk mempertegas perbedaan diantara peserta yang dapat menjawab dengan benar dan menjawab dengan salah.

## b. Tingkat kesulitan ( $b_i$ )

Parameter  $b_i$  merupakan suatu titik pada skala kemampuan agar peluang menjawab benar sebesar 50%. Semakin besar nilai  $b_i$ , maka semakin besar kemampuan yang diperlukan untuk menjawab benar dengan peluang 50 % atau dengan kata lain semakin besar nilai parameter  $b_i$ , maka semakin sulit butir soal tersebut. Nilai  $b_i$  (tingkat kesukaran butir) bervariasi dari -2 sampai +2. Nilai mendekati -2 menunjukkan bahwa butir tersebut sangat mudah sedangkan nilai mendekati +2 menunjukkan bahwa butir tersebut sangat sulit.<sup>14</sup>

Tingkat kesukaran item dari output BILOG-MG dapat dilihat pada nilai Threshold di output phase2. Berdasarkan hasil analisis data diperoleh informasi tingkat kesukaran seluruh item bergerak dari nilai -1.138 sampai dengan 2.799. Klasifikasi tingkat kesukaran item disajikan sebagai berikut:

<sup>14</sup> Hambleton, Ronald K., Hariharan Swaminathan, and H. Jane Rogers. *Fundamentals of item response theory*. Vol. 2. Sage, (1991), h.13



## TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

Tabel 2. Tingkat Kesukaran Item

Tingkat Kesulitan ( $a_i$ )	Nomor Item	Jumlah		Kategori
		F	%	
$-2 \leq b_i \leq 2.0$	1,2,3,4,5,6,7,8,9,10,11	38	95	Baik
	12,13,14,15,16,17,18,19			
	20,21,22,23,24,25,26,27			
	29,30,31,32,33,34,35,37			
	38,39,40			
$b_i > 2.0$	28,36	2	5	Kurang Baik
Total		40	100	

Berdasarkan tabel 2 diperoleh informasi bahwa 95% dari keseluruhan item memiliki tingkat kesulitan yang baik. Informasi ini membuktikan bahwa 95% dari keseluruhan butir tes formatif yang dikembangkan telah mampu menggambarkan fungsi dari kemampuan seseorang. Dimana mahasiswa PPG yang memiliki kemampuan tinggi akan merasa mudah mengerjakan butir soal, sebaliknya mereka yang memiliki kemampuan rendah akan merasa sulit menjawab butir soal. Sedangkan 5% lagi termasuk butir yang sulit karena memiliki indeks kesukaran lebih dari 2.0. Jika soal cenderung sulit artinya sebagian tingkat kemampuan peserta didik cenderung rendah. Pada umumnya makin mudah butir atau makin kecil  $b$ , maka makin besar probabilitas responden untuk menjawab butir itu dengan benar sehingga nilai  $P(\theta)$  menjadi besar. Sebaliknya makin sukar butir atau makin besar  $b$ , maka makin kecil probabilitas responden menjawab butir itu dengan benar sehingga nilai menjadi kecil. Dengan demikian mudah sukarnya suatu butir sering dikaitkan dengan kemampuan responden dengan taraf sukar butir yakni dengan ada butir yang sukar, ada butir yang sedang, dan ada butir yang mudah.<sup>15</sup>

Berdasarkan analisis fit model dan estimasi parameter di atas, maka dapat ditentukan berapa banyak item yang baik yang memenuhi kriteria. Untuk mengetahui kualitas item secara empiris dapat digunakan ketentuan dari Hulin, Drasgow dan Parsons (1983)<sup>16</sup> dengan kriteria baik, kurang baik, dan tidak baik. Kriteria masing-masing parameter sebagai berikut: 1) Item tergolong baik apabila item cocok dengan model, memiliki indeks kesukaran  $-2.0$  sampai  $2.0$ , indeks daya pembeda butir  $0.0$  sampai  $2.0$ , dan peluang tebakantidak lebih dari  $0.2$ ; 2) Item

<sup>15</sup> Naga, Dali S. "Karakteristik butir pada alat ukur model dikotomi." *Arkhe: Jurnal Ilmiah Psikologi* 3, no. 4 (1998), h.34

<sup>16</sup> Hullin, Ch L., F. Drasgow, and Ch K. Parsons. "Item response theory: Application to psychological measurement." *Homewood, IL: Dow Jones-Irwin* (1983), h.16-25.



## TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

---

tergolong kurang baik apabila item cocok dengan model, memiliki indeks kesukaran kurang dari -2.0 atau lebih dari 2.0, indeks daya pembeda butir lebih dari 2.0, dan peluang tebakan lebih dari 0.2; 3) Item tergolong tidak baik, jika tidak cocok dengan model. Hal ini menunjukkan dari 40 total butir soal yang dikembangkan, maka terdapat 6 item atau sekitar 15% item yang memiliki kategori kurang baik yaitu item soal nomor 7, 14, 15, 18, 36, dan nomor 39. Keenam item soal tersebut tidak memenuhi persyaratan analisis, maka keenam item tersebut tergolong item yang tidak baik.

Berdasarkan hasil analisis di atas, maka dapat disimpulkan bahwa item soal tes formatif berbasis HOTS untuk modul evaluasi pembelajaran yang dikembangkan memiliki kriteria valid, indeks daya beda yang baik dan indeks tingkat kesukaran sedang. Sehingga telah memenuhi semua kriteria uji instrumen tes baik, yaitu valid, reliabel, daya beda yang baik dan tingkat kesukaran sedang. Artinya soal ini dapat digunakan untuk mengukur pemahaman mahasiswa PPG pada pendalaman materi modul Evaluasi Pembelajaran.

### **KESIMPULAN**

Berdasarkan hasil analisis data diperoleh informasi bahwa 95% dari keseluruhan item memiliki tingkat kesulitan yang baik. Informasi ini membuktikan bahwa 95% dari keseluruhan butir tes telah mampu menggambarkan fungsi dari kemampuan seseorang. Dimana peserta PPG yang memiliki kemampuan tinggi akan merasa mudah mengerjakan butir soal, sebaliknya mereka yang memiliki kemampuan rendah akan merasa sulit menjawab butir soal. Sedangkan 5% lagi termasuk butir yang sulit karena memiliki indeks kesukaran lebih dari 2.0. Jika soal cenderung sulit artinya sebagian tingkat kemampuan peserta didik cenderung rendah. Selanjutnya 82,5% item soal memiliki daya pembeda item yang baik. Hal ini membuktikan bahwa 82,5% dari keseluruhan item-item tes soal memiliki kemampuan untuk mempertegas perbedaan di antara peserta yang dapat menjawab dengan benar dan menjawab dengan salah. Dari keseluruhan item butir soal terdapat 85% item yang memiliki karakteristik/kualitas yang baik, dimana item tersebut fit dengan model dan memenuhi estimasi parameter daya pembeda, tingkat kesukaran dan tebakan semua yang baik. Selebihnya 15% item tergolong kurang baik karena ada salah satu atau beberapa kriteria estimasi parameter item yang tidak memenuhi kriteria item yang baik, baik itu parameter daya pembeda maupun tingkat kesukaran. Implikasi dari penelitian ini dapat digunakan sebagai rekomendasi penyusunan soal tes formatif bagi modul pedagogik khususnya dan modul profesional lainnya.



## TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

---

### DAFTAR PUSTAKA

- Abosalem, Y. (2016). Assessment techniques and students' higher-order thinking skills. *International Journal of Secondary Education*, 4(1), 1-11.
- Fahrurrozi, F. (2022). Pengaruh Kualitas LMS dan Prilaku Belajar terhadap Kualitas Belajar Mahasiswa PPG. *Jurnal Muara Pendidikan*, 7(1), 131-136.
- Fajarini, I., Oktaviani, C., & Fitria, D. (2022). Meningkatkan kemampuan paedagogik guru melalui pelatihan pembuatan soal ujian berbasis pembelajaran multiliterasi terintegrasi higher order thinking skills di SD Negeri 2 Seulalah Langsa. *Jurnal Ilmiah Wahana Pendidikan*, 8(12), 138-146.
- Hakim, B. R., Ardiyanto, A., & Wijayanti, A. (2021). Analisis hots pada instrumen penilaian siswa kelas iv sekolah dasar. *Wawasan Pendidikan*, 1(2), 246-254.
- Hambleton, R.K & Swaminathan, H. (1985). *Item Response Theory*. Boston, MA: Kluwer Inc.
- Hambleton, R.K., Swaminathan H. & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication Inc.
- Hanun, F. (2021). Implementasi penyelenggaraan program pendidikan profesi guru (PPG) pendidikan agama islam di LPTK UIN Serang Banten. *EDUKASI: Jurnal Penelitian Pendidikan Agama Dan Keagamaan*, 19(3), 268-285. <https://doi.org/10.32729/edukasi.v19i3.1158>
- Hullin, C. L., Drasgow, F., & Parsons, C. K. (1983). Item response theory: Application to psychological measurement. *Homewood, IL: Dow Jones-Irwin*.
- Kurniawati, R. P., & Hadi, F. R. (2021). Pelatihan Pengembangan Instrumen Evaluasi Berbasis HOTS untuk Guru Sekolah Dasar. *Jurnal Altifani Penelitian Dan Pengabdian Kepada Masyarakat*, 1(4), 267-276. <https://doi.org/10.25008/altifani.v1i4.182>
- Maryani, E. (2022). Pendidikan profesi guru (PPG) sebagai cara untuk menjadikan guru menjadi profesional. *GUAU: Jurnal Pendidikan Profesi Guru Agama Islam*, 2(4), 171-178.



## TADBIR: Jurnal Manajemen Pendidikan Islam

P-ISSN: 2338:6673; E-ISSN 2442:8280

Vol. 12. No. 01. Februari, 2024, Hal: 159-174

---

- Maulana, I., Rahma, N., Mahfirah, N., Alfarizi, W., & Darlis, A. (2023). Meningkatkan Profesional Guru dengan Program Pendidikan Profesi Guru (PPG). *Journal on Education*, 5(2), 2158-2167. <https://doi.org/10.31004/joe.v5i2.867>.
- Maulina, W. (2022). Pengembangan Instrumen Asesmen Berbasis Hots pada Mata Pelajaran Pendidikan Agama Islam dan Budi Pekerti di SMAN 6 Padang. *Jurnal Pendidikan dan Konseling (JPDK)*, 4(6), 7317-7326.
- Naga, D. S. (1998). Karakteristik butir pada alat ukur model dikotomi. *Arkhe: Jurnal Ilmiah Psikologi*, 3(4).
- Suhady, W., Roza, Y., & Maimunah, M. (2020). Pengembangan soal untuk mengukur higher order thinking skill (hots) siswa. *Jurnal Gantang*, 5(2), 143-150.
- Supriadi, G. (2020). Pengembangan instrumen penilaian berbasis higher order thinking skill (HOTS). Yogyakarta: Aswaja Pressindo.