Jurnal Pendidikan MIPA
Volume 16. Nomor 5, Tahun 2026 | ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i5.5188

Analisis Pencapaian Ujian Komprehensif dalam Program Pendidikan Profesi Guru dalam Jabatan Bidang Fisika

Gazali Rachman[1),*], Erlin Eveline[1)], Izaak Hendrik Wenno[1)]

[1)]Pendidikan Fisika, Universitas Pattimura

*Coresponding Author: gazali.rachman@lecturer.unpatti.ac.id

ABSTRAK

Penelitian ini bertujuan untuk menganalisis hasil tes komprehensif mata pelajaran Fisika peserta Program Pendidikan Profesi Guru (PPG) dalam Jabatan. Tujuan utamanya adalah untuk menilai efektivitas program PPG berdasarkan nilai ujian peserta dan untuk mengetahui perbedaan penilaian antara dua penguji. Dengan menggunakan pendekatan kuantitatif, penelitian ini memfokuskan pada 27 peserta yang terdaftar dalam program pendidikan profesi guru sebagai sampel penelitian. Data dianalisis menggunakan metode statistik deskriptif dan inferensial. Temuan menunjukkan nilai rata-rata yang diberikan oleh dua penguji adalah 85,41 dan 87,26. Hasil uji-t menunjukkan bahwa nilai tes komprehensif peserta memenuhi atau melampaui standar kelulusan minimal 70. Selain itu, terdapat perbedaan antara penilaian kedua penguji, di mana satu penguji secara konsisten memberikan nilai yang lebih rendah, namun selisih rerata antar-penguji tergolong kecil. Kedua penguji juga secara praktis memberikan penilaian yang cukup konsisten. Temuan-temuan dalam penelitian ini menunjukkan bahwa peserta pada program ini telah memenuhi standar kelulusan. Perbedaan penilaian menjelaskan perlunya standar kriteria di antara para penguji. Penelitian ini berfungsi sebagai sumber informasi penting dalam mempertimbangkan pengembangan sistem penilaian standar pada program, yang pada akhirnya bertujuan untuk meningkatkan hasil pendidikan bagi para pendidik di masa depan.

Kata Kunci: Program Pendidikan Profesional Guru; Fisika; Tes

This is an open access article under the CC - BY license.

PENDAHULUAN

Berbagai bidang telah mengalami perubahan yang signifikan dalam beberapa tahun terakhir sebagai bagian dari perkembangan masyarakat yang lebih luas, termasuk bidang pendidikan. Para guru di Indonesia dituntut untuk terus meningkatkan kompetensi profesional mereka di lingkungan sekolah (Irfan dkk., 2022). Menanggapi hal tersebut, pemerintah Indonesia berkomitmen untuk memajukan kualitas guru. Upaya-upaya ini mencakup pelaksanaan Program Pendidikan Profesi Guru (PPG), memfasilitasi kolaborasi dan pengembangan guru—seperti melalui inisiatif "Guru Berbagi" dan "Guru Belajar"—serta mendorong pembelajaran berkelanjutan bagi para pendidik, termasuk melalui program "Guru Penggerak" (Irfan dkk., 2022; Manao dkk., 2022; Susanto dkk., 2024). Sejalan dengan peningkatan profesionalisme pendidik tersebut, evaluasi pelaksanaan program PPG menjadi bagian yang perlu diperhatikan sehingga dapat menjadi pedoman dalam peningkatan program.

Guru di Indonesia diharapkan memiliki kompetensi yang dibuktikan dengan sertifikat pendidik profesional (Irfan dkk., 2022). Sertifikat ini diperoleh melalui proses sertifikasi guru dalam program profesi guru, yang dikenal sebagai Program Pendidikan Profesi Guru (PPG). Program ini diselenggarakan oleh Kementerian bekerja sama dengan perguruan tinggi. PPG merupakan program yang bertujuan meningkatkan kualitas guru agar mereka menjadi guru profesional. Sertifikat tersebut menjadi bukti bahwa seorang guru memiliki kompetensi dan kualitas yang baik, atau merupakan guru profesional. Dengan demikian, PPG yang diselenggarakan oleh perguruan tinggi bertujuan untuk menghasilkan guru profesional yang memiliki sertifikat pendidik.

Beberapa penelitian terkait evaluasi pelaksanaan PPG telah dilakukan. Sadsabila dkk. (2022) menguji efektivitas guru lulusan PPG dalam proses pembelajaran bahasa Indonesia di sekolah dasar; hasil penelitian menunjukkan tingkat efektivitas yang tergolong cukup baik. Darmawang dkk. (2021) menguji efektivitas PPG melalui tes komprehensif dan memperoleh hasil dalam kategori sangat baik. Daud dkk. (2020) meneliti dampak PPG berdasarkan pendapat peserta program untuk mata pelajaran bahasa Inggris; hasilnya menunjukkan bahwa

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i5.5188

69,2% peserta menyatakan PPG memberikan dampak signifikan terhadap pengembangan profesi mengajar mereka. Meskipun program ini dinilai mampu meningkatkan kompetensi guru, efektivitasnya masih jarang dievaluasi secara kuantitatif melalui tes komprehensif. Selain itu, proses penilaian memungkinkan adanya variasi yang dapat memengaruhi validitas hasil belajar peserta didik.

Kegiatan pembelajaran PPG Dalam Jabatan dirancang agar guru memiliki kompetensi profesional, termasuk keterampilan dalam literasi teknologi informasi dan komunikasi, inovasi, dan kemahiran berbahasa. Keterampilan-keterampilan ini digunakan untuk merancang pembelajaran berbasis masalah (*problem-based learning*) dan pembelajaran berbasis proyek (*project-based learning*). Terdapat beberapa tahapan dalam pelaksanaan PPG Dalam Jabatan yang disebut sebagai alur pembelajaran. Alur pembelajaran tersebut mencakup pendalaman materi, penyusunan perangkat pembelajaran, dan praktik pengalaman lapangan, serta dilaksanakan dalam dua siklus. Namun, pada siklus pertama, dilakukan tes komprehensif. Tes komprehensif merupakan penilaian menyeluruh terhadap kemampuan peserta PPG dalam mengidentifikasi masalah, menentukan penyebab, memilih solusi atas masalah yang teridentifikasi, serta kemampuan mereka dalam merancang tindakan dan rencana evaluasi. Tes komprehensif dilaksanakan melalui presentasi rancangan pembelajaran inovatif—yang dihasilkan dari mata kuliah pengembangan perangkat pembelajaran—dalam bentuk seminar. Tes komprehensif ini diselenggarakan pada siklus 1 sebagai syarat untuk mengikuti praktik pengalaman lapangan (PPL).

Dalam pelaksanaan penilaian, terdapat kriteria yang telah ditetapkan sebagai standar penilaian yang baik. Kriteria penilaian dalam pendidikan harus memenuhi beberapa prinsip, seperti valid, objektif, adil, terpadu, transparan, menyeluruh dan berkesinambungan, sistematis, beracuan kriteria, serta akuntabel (Kemdiktisaintek, 2022). Selain itu, kriteria penilaian juga dapat ditinjau dari aspek validitas dan reliabilitas. Validitas dan reliabilitas merupakan dua indikator yang menunjukkan kualitas suatu penilaian (Bashooir & Supahar, 2018). Validitas menunjukkan bahwa suatu tes mengukur apa yang seharusnya diukur, sedangkan reliabilitas berkaitan dengan konsistensi dalam pengukuran (Alias, 2005). Jika hasil penilaian dari dua penilai berbeda terhadap objek penilaian yang sama menunjukkan konsistensi, maka penilaian tersebut dapat dikatakan reliabel. Validitas dapat ditinjau melalui validitas isi, validitas konstruk, dan validitas bahasa. Validitas ini mencakup evaluasi terhadap isi instrumen oleh para ahli atau penilaian mengenai apakah instrumen yang dikembangkan telah mencakup seluruh aspek yang dinilai (Ilahi dkk., 2024). Terdapat juga validitas empiris yang menganalisis hasil jawaban peserta ujian (Bashooir & Supahar, 2018; Ramadhan dkk., 2020). Sementara itu, reliabilitas dapat dilihat melalui reliabilitas antar-penilai melalui koefisien korelasi intrakelas *(Intraclass Correlation Coefficient-ICC)* yang digunakan untuk menilai konsistensi antar penilai dalam data kualitatif. Dengan menggunakan interval keyakinan 95% dari estimasi ICC, kriteria nilai ICC kurang dari 0,5, antara 0,5 dan 0,75, antara 0,75 dan 0,9, dan lebih besar dari 0,90 secara berturut-turut mengindikasikan reliabilitas yang buruk, sedang, baik, dan sangat baik (Koo dkk., 2016). Kriteria ini menjadi acuan dalam pelaksanaan evaluasi. Seluruh penilai wajib memperhatikan dan mematuhi ketentuan yang tercantum dalam kriteria tersebut. Dengan mengikuti kriteria ini, diharapkan hasil penilaian benar-benar mencerminkan kondisi peserta (Mariati, 2012; Salirawati, 2013; Syaifuddin dkk., 2023).

Penilaian dalam ujian komprehensif memegang peranan penting dalam mengukur kelayakan peserta. Penilaian ini perlu dilaksanakan secara objektif dan konsisten guna menghasilkan hasil yang dapat diandalkan. Ujian ini dilakukan oleh dua orang dosen yang memiliki bidang keahlian relevan. Namun, dalam praktiknya, perbedaan nilai antara kedua penguji mungkin saja terjadi. Perbedaan tersebut dapat disebabkan oleh berbagai faktor, baik yang berasal dari pihak penguji maupun dari instrumen penilaian itu sendiri. Faktor yang berasal dari penguji antara lain meliputi subjektivitas dan perbedaan interpretasi terhadap rubrik penilaian; hal ini dapat terjadi apabila penguji tidak mematuhi prinsip-prinsip penilaian. Jika terdapat perbedaan signifikan dalam penilaian yang dilakukan oleh kedua penguji, akurasi hasil ujian komprehensif akan berkurang. Kondisi ini berdampak pada peserta PPG karena hasil penilaian tidak sepenuhnya mencerminkan kompetensi peserta secara menyeluruh.

Di sisi lain, ujian komprehensif juga dapat memberikan informasi mengenai efektivitas pembelajaran yang telah dilaksanakan dalam program PPG. Pembelajaran dalam PPG dinilai berhasil jika mampu meningkatkan kompetensi guru, yang ditunjukkan dengan perolehan nilai ujian komprehensif minimal 70. Sabon (2018) menemukan bahwa meskipun terdapat peningkatan nilai rata-rata setelah mengikuti program pendidikan dan pelatihan profesi guru, kompetensi pedagogik dan profesional guru masih tergolong sangat rendah karena jauh dari standar kelulusan yang ditetapkan. Faktor-faktor seperti interaksi antarpeserta, motivasi belajar, kompetensi

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i5.5188

guru, metode pengajaran, dan pelaksanaan tes pengetahuan secara berkala memiliki dampak besar terhadap efektivitas proses pendidikan (Lu dkk., 2021; Vasilev dkk., 2024). Namun, tantangan seperti kurangnya variasi metode pengajaran, keterbatasan fasilitas, dan keterbatasan waktu menjadi hambatan dalam mencapai pembelajaran yang optimal pada pelaksanaan PPG. Oleh karena itu, perlu dilakukan analisis terhadap hasil penilaian dari dua dosen untuk mengetahui efektivitas pembelajaran serta perbedaan dalam evaluasi yang mereka berikan. Berdasarkan hal tersebut, penelitian ini bertujuan untuk menganalisis efektivitas pembelajaran dalam PPG dengan menggunakan tes komprehensif sebagai alat evaluasi, serta menganalisis perbedaan penilaian yang diberikan oleh dua dosen dalam Tes Komprehensif PPG Dalam Jabatan.

METODE

Penelitian ini bertujuan untuk menganalisis efektivitas pembelajaran serta perbedaan penilaian dari dua penguji dalam program PPG berdasarkan tes komprehensif. Subjek penelitian ini adalah 27 mahasiswa PPG dalam jabatan tahun 2023 untuk mata pelajaran Fisika. Mahasiswa PPG mengikuti pembelajaran yang terdiri dari sembilan tahapan, dengan alur yang mencakup pendalaman materi, pengembangan bahan ajar, dan praktik pembelajaran inovatif pada siklus 1. Tahapan-tahapan tersebut meliputi identifikasi awal, eksplorasi penyebab masalah, penentuan penyebab masalah, eksplorasi alternatif solusi, penentuan solusi, penyusunan rencana tindakan, penyusunan rencana evaluasi, pelaksanaan rencana tindakan dan evaluasi, refleksi, serta perencanaan tindak lanjut. Sebelum memasuki tahapan kedelapan dan kesembilan (pelaksanaan rencana tindakan dan evaluasi, refleksi, serta perencanaan tindak lanjut), tes komprehensif dilaksanakan terhadap para mahasiswa.

Pertanyaan dalam penelitian ini adalah: (1) seberapa efektif pembelajaran program PPG bagi mahasiswa dalam jabatan (in-service) berdasarkan ujian komprehensif, dan (2) apakah terdapat perbedaan penilaian berdasarkan hasil evaluasi dari dua orang dosen. Untuk menjawab pertanyaan-pertanyaan tersebut, dilakukan analisis deskriptif dan perbandingan terhadap hasil ujian komprehensif mahasiswa PPG. Ujian komprehensif dilaksanakan selama maksimal 50 menit oleh dua dosen dari bidang studi terkait. Penilaian terhadap 27 peserta dilakukan oleh dua orang dosen sebagai tim penilai (penguji) dengan menggunakan pasangan penguji yang sama untuk setiap peserta. Penggunaan data peserta telah melalui prosedur perizinan serta memperoleh persetujuan dari pihak terkait. Setelah itu, dosen melakukan pengujian secara lisan melalui daring. Materi ujian mencakup identifikasi masalah, eksplorasi penyebab masalah, penentuan penyebab masalah, eksplorasi penentuan solusi, penentuan solusi, rencana tindakan dan rencana evaluasi, materi yang mendukung keterampilan berpikir tingkat tinggi (*Higher-Order Thinking Skills*-HOTS), literasi numerasi, serta integrasi teknologi informasi dan komunikasi, substansi materi pelajaran, dan pedagogi ke dalam rancangan pembelajaran (metode). Instrumen penilaian dalam ujian komprehensif menggunakan skala Likert dengan rentang dari sangat kurang hingga sangat baik.

Instrumen penilaian dalam ujian komprehensif menggunakan skala Likert dengan rentang dari sangat kurang hingga sangat baik (mulai dari 1 poin untuk penilaian sangat kurang hingga 5 poin untuk penilaian sangat baik). Terdapat 10 aspek yang dinilai yaitu: (1) kemampuan mengidentifikasi masalah, (2) kemampuan mengeksplorasi penyebab masalah, (3) kemampuan menentukan penyebab masalah, (4) kemampuan eksplorasi penentuan solusi, (5) kemampuan menentukan solusi, (6) kemampuan membuat rencana aksi, (7) kemampuan membuat rencana evaluasi, (8) penguasaan materi yang mendukung HOTS, (9) penguasaan literasi numerasi, dan (10) kemampuan mengintegrasikan TPAC dalam pembelajaran. Nilai yang diperoleh kemudian dikonversi ke dalam skala 0-100 dengan menggunakan persamaan berikut.

$$Nilai = \frac{poin\ yang\ didapat}{poin\ maksimum} \times 100$$

Terdapat dua hipotesis yang diajukan dalam penelitian ini. Hipotesis pertama adalah: (1) H_0 = nilai rata-rata tes komprehensif mahasiswa PPG lebih kecil atau sama dengan 70 (H_0: $\mu \leq 70$) ; Ha: nilai rata-rata tes komprehensif mahasiswa PPG lebih dari 70 (H_a: $\mu > 70$). Jika nilai t-hitung lebih besar daripada nilai t-tabel, maka H_0 ditolak. Jika nilai t-hitung lebih kecil daripada nilai t-tabel, maka H_0 diterima dan Ha ditolak. Hipotesis kedua adalah: H_0 = Tidak terdapat perbedaan rata-rata penilaian antara Dosen 1 dan Dosen 2 dalam tes komprehensif PPG; Ha: Terdapat perbedaan rata-rata penilaian antara Dosen 1 dan Dosen 2 dalam tes komprehensif PPG. Jika nilai signifikansi yang diperoleh dari uji perbedaan kurang dari atau sama dengan 0,05, maka H_0 ditolak; sedangkan jika nilai signifikansi lebih besar dari 0,05, maka H_0 diterima atau Ha ditolak (Ismail, 2018; Sugiyono, 2016, 2025).

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i5.5188

HASIL DAN PEMBAHASAN

Hasil penilaian ujian komprehensif oleh dua evaluator (Dosen 1 dan Dosen 2) dianalisis menggunakan statistik deskriptif untuk memperoleh gambaran mengenai data yang diperoleh. Hasil analisis menunjukkan karakteristik data sebagaimana disajikan pada Tabel 1.

Tabel 1. Hasil Uji Statistik Deskriptif

Assessment	Statistik	
Uji Komprehensif – Dosen 1	Rata-Rata	85,4074
	Median	84,0000
	Varians	45,174
	Standar Deviasi (SD)	6,72114
	Nilai Maksimum	98,00
	Nilai Minimum	72,00
	Skewness	0,146
	Kurtosis	-0,576
Uji Komprehensif – Dosen 2	Rata-Rata	87,2593
	Median	86,0000
	Varians	28,969
	Standar Deviasi (SD)	5,38225
	Nilai Maksimum	98,00
	Nilai Minimum	78,00
	Skewness	0,325
	Kurtosis	-0,565

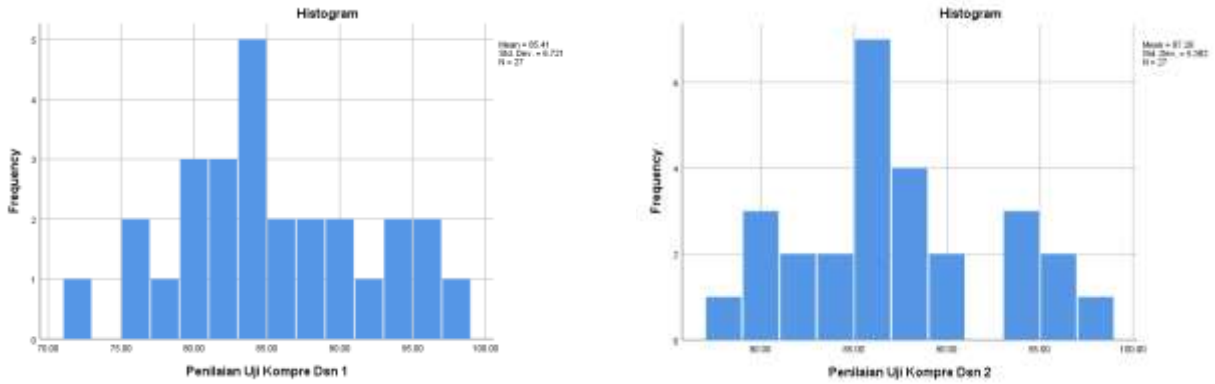
Tabel 1 menyajikan hasil statistik deskriptif dari penilaian oleh dua orang dosen. Skor rata-rata yang diberikan oleh Dosen 1 adalah 85,41, sedangkan skor rata-rata dari Dosen 2 adalah 87,26. Skor rata-rata Dosen 2 lebih tinggi dibandingkan Dosen 1. Sebaran skor dari Dosen 1 (Varians: 45,174; SD: 6,72) lebih homogen dibandingkan Dosen 2 (Varians: 28,969; SD: 5,38). Sebaran skor dari kedua dosen tersebut cenderung simetris, sebagaimana terlihat dari nilai kemencengan (*skewness*) yang mendekati nol (0,146; 0,325). Skor rata-rata dari kedua dosen termasuk dalam kategori tinggi, yang menunjukkan bahwa secara keseluruhan, peserta menunjukkan kinerja yang baik dalam tes komprehensif tersebut. Sementara itu, nilai standar deviasi yang kecil mengindikasikan bahwa skor peserta tidak tersebar terlalu jauh dan cenderung mengelompok di sekitar nilai rata-rata.

Tabel 2. Hasil Uji Tes Normalitas

No	Kolmogorov-Smirnov (Sig.)	Shapiro-Wilk (Sig.)
Uji Komprehensif – Dosen 1	0,198	0,677
Uji Komprehensif – Dosen 2	0,128	0,208

Tabel 2 menyajikan hasil uji normalitas yang dilakukan sebelum uji perbedaan. Berdasarkan hasil uji normalitas tersebut, nilai signifikansi untuk penilaian ujian komprehensif oleh Dosen 1 dan Dosen 2 masing-masing adalah 0,677 dan 0,208; karena nilai-nilai tersebut lebih besar dari 0,05, maka disimpulkan bahwa data berdistribusi normal menurut uji Shapiro-Wilk. Kesimpulan serupa juga diperoleh dari uji Kolmogorov-Smirnov, dengan nilai signifikansi untuk kedua penilai masing-masing sebesar 0,198 dan 0,128, yang juga lebih besar dari 0,05. Dengan demikian, uji perbedaan statistik seperti uji-t dapat dilanjutkan. Sebaran data dapat dilihat pada Gambar 1, yang menunjukkan bahwa data tersebar secara merata dan mengikuti kurva normal.

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i5.5188



Gambar 1. Histogram Nilai Uji Komprehensif untuk Dosen 1 dan Nilai Uji Komprehensif untuk Dosen 2

Gambar 1 menampilkan histogram distribusi nilai dari kedua dosen. Distribusi nilai untuk kedua dosen menunjukkan pola yang mendekati distribusi normal. Distribusi normal juga dapat dilihat dari nilai kemencengan (*skewness*) kedua kelompok data, yaitu 0,146 untuk Dosen 1 dan 0,325 untuk Dosen 2. Nilai-nilai tersebut mendekati nol, sehingga distribusinya tergolong normal. Selain itu, nilai kurtosis negatif menunjukkan bahwa distribusi nilai sedikit lebih datar dibandingkan dengan distribusi normal. Namun, nilai-nilai tersebut masih berada dalam rentang yang dapat diterima untuk analisis parametrik.

Tabel 3. Hasil Uji T Satu Sampel

Test value= 70			
	t	Sig. (2-tailed)	Mean Difference
Uji Komprehensif – Dosen 1	11,912	0,000	15.40741
Uji Komprehensif – Dosen 2	16,662	0,000	17.25926

Tabel 3 menyajikan hasil uji-t satu sampel bagi mahasiswa dari kedua penilai. Hasil uji-t menunjukkan nilai t-hitung sebesar 11,912 untuk Dosen 1 dan 16,662 untuk Dosen 2, sedangkan nilai t-tabel yang diperoleh adalah 1,706. Karena nilai t-tabel lebih kecil daripada nilai t-hitung, maka H_0 ditolak; hal ini berarti nilai rata-rata hasil tes komprehensif mahasiswa secara statistik bernilai secara signifikan lebih tinggi dari 70. Lebih lanjut, nilai mahasiswa mencapai nilai minimum sehingga dapat dikatakan bahwa temuan ini mengindikasikan bahwa capaian peserta memenuhi standar kelulusan pada pelaksanaan program PPG tahun 2023.

Kemampuan literasi teknologi informasi dan komunikasi berhasil dicapai karena program Pendidikan Profesi Guru (PPG) dalam jabatan tahun 2023 dilaksanakan secara daring. Peserta didorong untuk mengadopsi dan menguasai teknologi sebagai dampak dari pembelajaran daring atau jarak jauh. Pembelajaran tersebut memanfaatkan platform seperti *Learning Management System* (LMS). Bashori dkk. (2024) menemukan bahwa penggunaan LMS dalam PPG dapat berjalan secara efektif dan efisien. Kemampuan inovasi dan keterampilan berbahasa peserta juga dilatih melalui tahapan pembelajaran dalam PPG. Proses ini menuntut peserta untuk merancang pembelajaran berdasarkan langkah-langkah yang dimulai dari identifikasi masalah, analisis penyebab masalah, penentuan penyebab masalah, eksplorasi alternatif solusi, pengambilan keputusan terkait solusi, penyusunan rencana tindakan, hingga pembuatan rencana evaluasi. Setiap tahapan tersebut mengharuskan peserta untuk bersikap inovatif serta mengomunikasikan gagasan mereka, baik secara lisan maupun tertulis. Oleh karena itu, kegiatan pembelajaran dalam program PPG mendukung pengembangan kemampuan inovasi dan keterampilan berbahasa peserta. Manurung dkk. (2024) menunjukkan bahwa meskipun masih terdapat tantangan terkait akses internet dan literasi digital, program PPG daring bagi guru dalam jabatan terbukti efektif dalam meningkatkan kompetensi guru. Selain itu, keterampilan pedagogik dan profesional peserta juga dapat meningkat melalui program PPG (Mardhatillah & Surjanti, 2023). Namun, pelaksanaan PPG menghadapi berbagai tantangan, seperti kualitas lembaga penyelenggara, kualifikasi calon peserta, penyerapan lulusan, akses internet yang tidak merata, serta rendahnya literasi digital di kalangan peserta. Keterbatasan akses internet dan rendahnya literasi digital terutama terjadi di daerah terpencil dan di kalangan peserta yang berusia lebih tua (Azis dkk., 2021; Manurung dkk., 2024).

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i5.5188

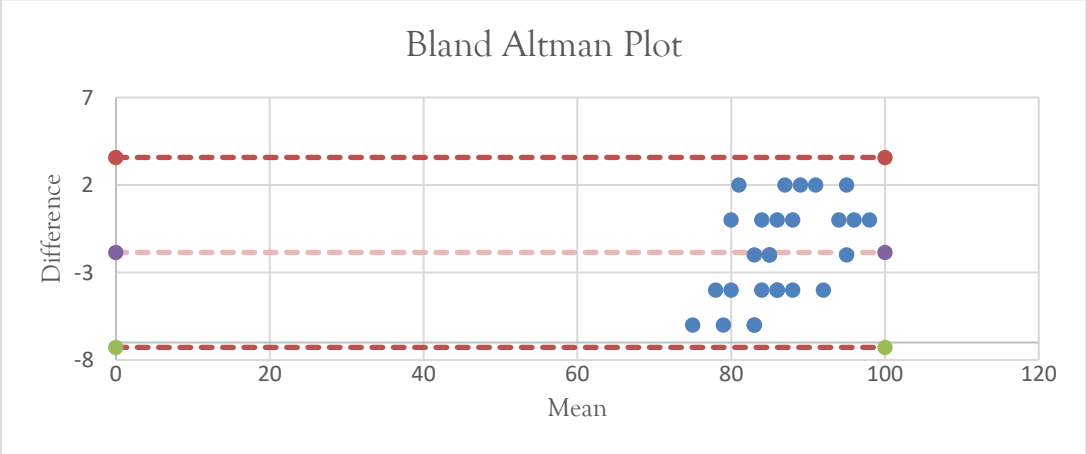
Tabel 4. Hasil Uji T Sampel Berpasangan

	Paired Differences			
	Mean	Standar Deviasi	t	Sig. (2-tailed)
Uji Komprehensif – Dosen 1 Uji Komprehensif – Dosen 2	-1.85185	2.76939	-3.475	0,002

Selanjutnya, Tabel 4 menyajikan hasil uji-t berpasangan yang bertujuan membandingkan penilaian antara kedua penilai. Berdasarkan hasil uji-t tersebut, nilai t-hitung adalah -3,475 (dengan nilai mutlak |t| = 3,475, df=26), dan nilai t-tabel adalah 2,056. Nilai signifikansinya adalah 0,002, yang berada di bawah ambang batas 0,05. Hasil ini menunjukkan adanya perbedaan rata-rata yang signifikan secara statistik dalam penilaian yang diberikan oleh Dosen 1 dan Dosen 2 pada ujian komprehensif PPG. Hal ini mengindikasikan bahwa meskipun secara umum peserta PPG mencapai nilai yang ditetapkan, terdapat perbedaan perspektif antara kedua dosen dalam melakukan penilaian. Meskipun terdapat perbedaan berdasarkan hasil uji analisis statistik, namun selisih rerata antar-penilai tergolong kecil, yaitu sebesar 1,85 poin. Selain itu, nilai *effect size* sebesar dz ≈ 0,67 (kategori sedang) menginformasikan juga bahwa secara praktis kedua dosen cukup konsisten. Hasil tersebut mengungkapkan bahwa penilaian dari kedua dosen tersebut sesungguhnya cukup konsisten.

Tabel 5. *Interclass Correlation Coefficient*

Model Pengukuran	ICC	F	Signifikansi
Single measures	0,897	18,334	0,000
Average measures	0,945	18,334	0,000
Cronbach's alpha	0,945		



Gambar 2. Grafik Bland Altman

Analisis dilanjutkan terhadap konsistensi penilaian antar penilai untuk memastikan lebih lanjut hasil penilaian antara dosen 1 dan dosen 2. Tabel 5 merupakan hasil uji reliabilitas statistik. Data yang dianalisis sebanyak 27 peserta dan tidak ada dara yang dikeluarkan untuk dianalisis. Nilai koefisien Cronbach's Alpha sebesar 0,945 yang mengindikasikan bahwa skor yang diberikan oleh kedua dosen tersebut memiliki pola yang konsisten. Kemudian, terdapat pula nilai *intraclass correlation coefficient* (ICC) dengan model *two-way mixed effects* digunakan untuk menilai tingkat kesepakan pengukuran. Berdasarkan Tabel 1, nilai ICC (*single measures*: 0,897; dan *average measures*: 0,945) menunjukkan reliabilitas skor yang diperoleh dari rata-rata kedua penilai tergolong baik. Dengan kata lain, skor yang diberikan oleh kedua penilai memiliki konsistensi yang tinggi. Analisis antar-penilai dapat juga dilakukan menggunakan metode Bland-Altman. Berdasarkan Gambar 2, seluruh nilai peserta berada di dalam batas kesesuaian (*limit of agreement-LOA)*. Data yang tersebar menunjukkan pola acak yang konsisten tanpa membentuk tren kemiringan naik atau turun yang ekstrem dan membentuk corong. Hal ini memperkuat temuan yaitu kedua penilai memiliki tingkat kesesuaian dan konsistensi pengukuran yang baik secara empiris.

Perbedaan dalam pemberian nilai ini dapat disebabkan oleh berbagai faktor, seperti perbedaan interpretasi terhadap rubrik penilaian, subjektivitas penilai, kriteria evaluasi yang digunakan, kejelasan rubrik, serta ekspektasi masing-masing penilai. Setiap dosen mungkin memiliki pemahaman yang berbeda mengenai

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i5.5188

rubrik penilaian tersebut. Menurut Nitko & Brookhart (2014), perbedaan dalam cara menginterpretasikan rubrik penilaian dapat memengaruhi validitas evaluasi yang dilakukan oleh penilai. Selain itu, penelitian oleh Masinta (2016) menyoroti pentingnya validasi rubrik penilaian. Yusuf (2023) menyatakan bahwa penggunaan rubrik penilaian serta pengembangan instrumen penilaian yang valid dan reliabel dapat meningkatkan objektivitas evaluasi. Rubrik yang tidak jelas atau tidak tepat dapat menyebabkan penilaian menjadi tidak akurat. Dalam praktik penilaian, masih terdapat permasalahan seperti kriteria evaluasi yang tidak jelas, ketidakpastian dalam pemberian nilai, dan subjektivitas dalam penilaian. Hasil ini dapat menjadi dasar untuk memastikan bahwa evaluasi memenuhi standar, seperti objektivitas dan konsistensi dalam penilaian ujian mahasiswa. Dalam sistem penilaian PPG, yang mencakup ujian komprehensif, seluruh evaluasi dilakukan oleh dosen melalui teknik observasi dengan menggunakan rubrik penilaian. Namun, tidak disebutkan adanya mekanisme untuk menyamakan persepsi di antara para dosen. Adanya mekanisme penyamaan persepsi memungkinkan penilaian yang lebih akurat; tanpa mekanisme tersebut, dapat terjadi perbedaan dalam cara dosen memberikan nilai. Secara keseluruhan, pelaksanaan program PPG berhasil memberikan dampak positif. Hal ini sejalan dengan temuan Islamiyah dkk. (2025) yang menunjukkan bahwa program PPG dapat meningkatkan pengetahuan dan keterampilan mengajar guru. Program ini berjalan secara efektif melalui penerapan pendekatan terintegrasi, pelatihan praktis, dan evaluasi berkelanjutan.

SIMPULAN

Penelitian ini menunjukkan bahwa peserta program PPG dalam jabatan untuk mata pelajaran Fisika telah mencapai standar kelulusan atau kompetensi. Hasil tersebut tercermin dalam hasil tes komprehensif, di mana seluruh peserta memperoleh nilai 70 atau lebih tinggi. Nilai rata-rata dari kedua penilai tergolong tinggi: 85,41 untuk penilai pertama (Dosen 1) dan 87,26 untuk penilai kedua (Dosen 2). Namun, berdasarkan hasil uji-t, terdapat perbedaan signifikan antara nilai yang diberikan oleh kedua dosen tersebut. Penilai kedua cenderung memberikan nilai yang lebih tinggi dibandingkan penilai pertama. Selain itu, kedua penilai memiliki konsistensi yang sangat baik. Perbedaan ini mengindikasikan adanya perbedaan interpretasi di antara para dosen. Variasi dalam penilaian ini dapat disebabkan oleh berbagai faktor, seperti perbedaan penafsiran terhadap rubrik penilaian, subjektivitas penilai, kriteria penilaian yang digunakan, tingkat kejelasan rubrik penilaian, serta ekspektasi masing-masing penilai. Studi ini hanya melibatkan dua penilai dan terbatas pada mata pelajaran Fisika, sehingga tidak dapat merepresentasikan keseluruhan sistem penilaian dalam program PPG. Penelitian lanjutan yang tidak hanya berfokus pada hasil kuantitatif tetapi juga menggali lebih dalam faktor-faktor eksternal yang memengaruhi pendekatan penilaian perlu dilakukan. Berdasarkan hasil studi ini, diperlukan pengujian efektivitas PPG melalui perbandingan kondisi sebelum dan sesudah mengikuti program; diperlukan pula standardisasi rubrik penilaian yang lebih jelas serta penyamaan persepsi dalam penilaian dengan melibatkan lebih banyak subjek penelitian (penguji dan mata pelajaran lain), serta analisis kualitatif untuk mengidentifikasi faktor-faktor penyebab perbedaan penilaian tersebut. Langkah-langkah ini bertujuan agar pelaksanaan program dapat ditingkatkan berdasarkan hasil evaluasi. Evaluasi program PPG harus dilakukan secara berkelanjutan guna menjamin konsistensi dan kualitas lulusan program.

Daftar Pustaka

Alias, M. (2005). *Assessment of learning outcomes: validity and reliability of classroom tests*. *4*(2). https://www.researchgate.net/publication/242477615

Azis, A. A., Arifin, A. N., & Daud, F. (2021). Pemanfaatan Aplikasi Evaluasi Pembelajaran Digital ISpring bagi Guru Sekolah Menengah kabupaten Gowa. *SMART: Jurnal Pengabdian Kepada Masyarakat*, *1*(1), 1. https://doi.org/10.35580/smart.v1i1.24431

Bashooir, K., & Supahar, S. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM. *Jurnal Penelitian dan Evaluasi Pendidikan*, *22*(2), 219–230. https://doi.org/10.21831/pep.v22i2.19590

Bashori, Novebri, & Prasetyo, M. A. M. (2024). Pembelajaran Daring di Lembaga Pendidikan dan Tenaga Kependidikan: Analisis Efektivitas Penggunaan Learning Management System Space. *POTENSIA: Jurnal Kependidikan Islam*, *10*(1), 1–12.

Darmawang, Bakharani, & Ahmad Zaki. (2021). Efektivitas Model Pembelajaran PPG dalam Menghasilkan

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i5.5188

Guru Profesional yang Berkarakter. *Seminar Nasional Hasil Penelitian "Penguatan Riset, Inovasi, dan Kreativitas Peneliti di Era Pandemi Covid-19."*

Daud, A., Novitri, & Hardian, M. (2020). Evaluation of the Teacher Professional Education Program (PPG): English Students' Perspective. *Al-Ishlah: Jurnal Pendidikan* , *12*(2), 185–200.

Ilahi, D. P., Festiyed, Utami, L., & Ratih, A. (2024). Pengembangan dan Validitas Instrumen Metakognitif Siswa pada Pembelajaran Fisika SMA. *Biogenerasi: Jurnal Pendidikan Biologi*, *9*(2). https://e-journal.my.id/biogenerasi

Irfan, A. M., Wahyuni, A. S. A., & Nulaela. (2022). Analisis Kemampuan Praktik Pembelajaran Inovatif pada Mahasiswa Pendidikan Profesi Guru dalam Jabatan Universitas Negeri Makassar. *Seminar Nasional Hasil Penelitian*.

Islamiyah, S., Nasution, H. A., & Aziz, A. (2025). *Efektivitas Program Pendidikan Profesi Guru (PPG) terhadap Profesionalisme Guru di UPT SMP Negeri 6 Medan*.

Ismail, F. (2018). *Statistika untuk Penelitian dan Ilmu-ilmu Sosial*. Prenadamedia Group.

Kemdiktisaintek. (2022). *Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 21 Tahun 2022 tentang Standar Penilaian Pendidikan pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah*. https://peraturan.bpk.go.id/Details/224425/permendikbudriset-no-21-tahun-2022

Koo, T. K., Li, M. Y., Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/J.JCM.2016.02.012

Lu, K., Yang, H. H., Shi, Y., & Wang, X. (2021). Examining the key influencing factors on college students' higher-order thinking skills in the smart classroom environment. *International Journal of Educational Technology in Higher Education*, *18*(1). https://doi.org/10.1186/s41239-020-00238-7

Manao, M. M., Parulian Sijabat, O., Situmorang, A. R., Hutauruk, A., & Panjaitan, S. (2022). Mengatur Kualitas Guru melalui Program Guru Penggerak. *Journal of Educational Learning and Innovation*, *2*(1), 130. https://doi.org/10.46229/elia.v2i1

Manurung, M. A. P., Chaniago, N. S., & Akhyar, S. (2024). Efektivitas Pelaksanaan Program PPG dalam Jabatan secara Online di LPTK UIN Sumatera Utara Medan. *Jurnal Research and Education Studies*, *3*(2).

Mardhatillah, O., & Surjanti, J. (2023). Peningkatan Kompetensi Pedagogik dan Profesionalitas Guru di Indonesia Melalui Pendidikan Profesi Guru (PPG). *Jurnal Pendidikan Ekonomi Undiksha*, *15*(1), 102–111. https://doi.org/10.23887/jjpe.v15

Mariati, I. (2012). Analisis Butir Soal dengan Teori Tes Klasik (Classical Test Theory) dan Teori Respons Butir (Item Response Theory) (Studi Kasus: Soal Ujian Olimpiade Sains Provinsi Bidang Informatika 2009). *PYTHAGORAS Jurnal Pendidikan Matematika*, *5*(2). https://doi.org/10.21831/pg.v5i2.536

Masinta, A. D. (2016). Validasi Rubrik sebagai Pedoman untuk Menilai Kemampuan Pemecahan Masalah Siswa SMA Berbasis Inkuiri dengan Menggunakan Tes Uraian [UIN Sumatera]. Dalam *Unnes Physics Education Journal* (Vol. 5, Nomor 2). http://journal.unnes.ac.id/sju/index.php/upej

Nitko, A. J., & Brookhart, S. M. (2014). *Educational Assessment of Students (7th ed.)*. Pearson.

Ramadhan, S., Sumiharsono, R., Mardapi, D., & Prasetyo, Z. K. (2020). The quality of test instruments constructed by teachers in bima regency, Indonesia: Document analysis. *International Journal of Instruction*, *13*(2), 507–518. https://doi.org/10.29333/iji.2020.13235a

Sabon, S. S. (2018). Efektivitas Pelatihan Guru melalui Pendidikan dan Latihan Profesi Guru. *Jurnal Penelitian Kebijakan Pendidikan*, *11*(3), 159–182.

Sadsabila, I., Puspitasari, M., & Larasati, N. (2022). Efektivitas PPG Terhadap Proses Pembelajaran Bahasa Indonesia Kelas VI di SDIT Bina Mulia Kota Depok. *Proseding Didaktis: Seminar Nasional Pendidikan Dasar*, 223–233.

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i5.5188

Salirawati, D. (2013). Pengembangan Instrumen Pendeteksi Miskonsepsi Kesetimbangan Kimia pada Peserta Didik SMA. *Jurnal Penelitian dan Evaluasi Pendidikan*, *15*(2), 232–249. https://doi.org/10.21831/pep.v15i2.1095

Sugiyono. (2016). *Metode Penelitian Kombinasi (Mixed Methods)*. Bandung: Alfabeta.

Sugiyono, S. (2025). Transforming Boring Grammar Teaching Methods into Fun Ones. *Jurnal Pendidikan dan Sastra Inggris*, *5*(2), 387–399. https://doi.org/10.55606/jupensi.v5i2.5385

Susanto, T. T. D., Solihin, R. R., Fauziyah, E. P., Yanti, N. V. I., & Ramadhania, A. P. (2024). Upaya Pemerintah Indonesia dalam Meningkatkan Kualitas Guru Ditinjau dari Hasil PISA 2022: Sebuah Kajian Pustaka. *PERSPEKTIF Ilmu Pendidikan*, *38*(1). https://doi.org/10.21009/PIP.381

Syaifuddin, S., Sarwi, S., Hartono, H., & Nuswowati, M. (2023). Prosiding Seminar Nasional Pascasarjana Karakterisasi Instrumen Tes Keterampilan Proses Sains pada Bahan Kajian Fluida Dinamis Berdasarkan Teori Tes Klasik. *Prosiding Seminar Nasional Pascasarjana UNNES 2023*, 818–825. http://pps.unnes.ac.id/pps2/prodi/prosiding-pascasarjana-unnes

Vasilev, Y., Vasileva, P., Batova, O., & Tsvetkova, A. (2024). Assessment of Factors Influencing Educational Effectiveness in Higher Educational Institutions. *Sustainability*, *16*(12), 4886. https://doi.org/10.3390/su16124886

Yusuf, M. (2023). Evaluasi Metode Penilaian dalam Pendidikan Islam dalam Upaya Meningkatkan Ketepatan dan Objektivitas Penilaian Siswa. *Jurnal Pendidikan Agama Islam*, *2*(1), 77–82. https://doi.org/10.56854/sasana.v2i1.218