

Pengembangan Instrumen Evaluasi Program Supervisi Akademik Berbasis CIPP (Context, Input, Process, Product)

Ratna Zahrotus Sania^{*1}, Edi Istiyono², Yolanderu Septiana³

^{1,2,3}Program Studi Penelitain dan Evaluasi Pendidikan, Sekolah Pascasarjana, Universitas Negeri
Yogyakarta, Yogyakarta, Indonesia
Email: ¹ratnazahrotus.2025@student.uny.ac.id,

Abstrak

Program supervisi akademik merupakan salah satu upaya penting dalam meningkatkan kinerja guru dan kualitas pembelajaran di sekolah. Namun, pelaksanaannya sering kali belum dievaluasi secara komprehensif karena keterbatasan instrumen yang valid dan reliabel untuk mengukur keberhasilan program secara menyeluruh. Penelitian ini bertujuan untuk mengembangkan instrumen evaluasi program supervisi akademik berdasarkan model evaluasi CIPP (*Context, Input, Process, Product*). Tahapan pengembangan instrumen meliputi analisis konstruk, penyusunan kisi-kisi instrumen, pengembangan butir instrumen, validasi ahli, revisi instrumen, serta uji coba instrumen. Instrumen yang dikembangkan terdiri dari angket, lembar wawancara, lembar observasi, dan lembar dokumentasi. Validitas isi instrumen dianalisis menggunakan indeks *Aiken's V* berdasarkan penilaian tiga validator. Validitas instrumen angket dianalisis menggunakan *Corrected Item Total Correlation* (CITC), sedangkan reliabilitas menggunakan koefisien *Cronbach Alpha* dengan *software R*. Hasil penelitian menunjukkan bahwa instrumen yang dikembangkan memiliki kualitas yang sangat baik. Hasil validasi isi menunjukkan nilai *Aiken's V* berkisar antara 0,778–1,000 dengan rata-rata sebesar 0,992. Sebanyak 51 dari 52 butir angket dinyatakan valid, sedangkan satu butir direvisi berdasarkan saran validator. Hasil uji validitas empiris menunjukkan seluruh butir memiliki nilai CITC di atas 0,30 dengan rentang nilai 0,657–0,829. Uji reliabilitas menghasilkan koefisien *Cronbach Alpha* sebesar 0,986 yang menunjukkan reliabilitas sangat tinggi. Temuan tersebut menunjukkan bahwa instrumen evaluasi program supervisi akademik berbasis model CIPP yang dikembangkan memenuhi kriteria validitas dan reliabilitas sehingga layak digunakan sebagai alat evaluasi program supervisi akademik di sekolah.

Kata kunci: CIPP, Evaluasi, Instrumen, Mixed Method, Reliabilitas, Supervisi Akademik, Validitas.

Development of a CIPP (Context, Input, Process, Product)-Based Evaluation Instrument for Academic Supervision Programs

Abstract

Academic supervision programs play a strategic role in improving teacher performance and enhancing the quality of classroom instruction. However, the effectiveness of these programs is often difficult to evaluate comprehensively due to the limited availability of valid and reliable evaluation instruments. This study aimed to develop and validate an evaluation instrument for academic supervision programs based on the CIPP (*Context, Input, Process, Product*) evaluation model. The development stages included construct analysis, instrument blueprint design, item development, expert validation, instrument revision, and instrument try-out. The developed product consisted of questionnaires, interview guidelines, observation sheets, and documentation checklists. Content validity was analyzed using *Aiken's V* index based on assessments from three experts. Empirical validity of the questionnaire was examined using *Corrected Item-Total Correlation* (CITC), while reliability was analyzed using *Cronbach's Alpha* coefficient with the assistance of *R* software. The results indicated that the developed instrument possessed excellent psychometric properties. Content validity analysis yielded *Aiken's V* values ranging from 0.778 to 1.000, with an average value of 0.992. A total of 51 out of 52 questionnaire items were declared valid, while one item was revised based on expert recommendations. The empirical validity test showed that all items had CITC values above 0.30, ranging from 0.657 to 0.829. Reliability testing produced a *Cronbach's Alpha* coefficient of 0.986, indicating very high reliability. These findings demonstrate that the CIPP-based academic supervision program evaluation instrument meets the criteria of validity and reliability and is suitable for evaluating academic supervision programs in schools.

Keywords: kata kunci sedapat mungkin menjelaskan isi tulisan, ditulis dengan huruf kecil kecuali singkatan, maksimum enam kata, masing-masing dipisahkan dengan koma, Times New Roman 10, italic dan ditulis urut

alphabetic. Academic Supervision, CIPP, Evaluation, Instrumental Development, Mixed Method, Validity And Reliability.

1. PENDAHULUAN

Pendidikan merupakan fondasi yang sangat penting dalam upaya pembangunan sumber daya manusia yang berkualitas. Pemerintah melalui Kemdikbud terus menerus berupaya melakukan berbagai pembaharuan sistem pendidikan nasional guna untuk menata dan memperbaiki mutu pendidikan di Indonesia. Salah satu upaya yang dapat dilakukan untuk meningkatkan mutu Pendidikan di Indonesia adalah dengan peningkatan kualitas pembelajaran. Kualitas proses pembelajaran sangat dipengaruhi oleh kinerja (*work performance*) guru [1]. Kinerja guru yang baik dapat dilihat berdasarkan kemampuan merencanakan proses pembelajaran, melaksanakan proses pembelajaran, serta merefleksi dan mengevaluasi proses pembelajaran yang dilakukan. Upaya yang dapat dilakukan untuk meningkatkan kinerja guru salah satunya yaitu dengan melaksanakan program evaluasi pendidikan melalui supervisi akademik. Supervisi akademik adalah serangkaian kegiatan membantu guru mengembangkan kemampuannya dalam proses pembelajaran untuk mencapai tujuan pembelajaran [1]. Program supervisi akademik bertujuan untuk mengoptimalkan kinerja guru, merangsang refleksi diri, dan mendukung perkembangan kompetensi guru [2] [3]. Program supervisi akademik mencakup berbagai praktik pengawasan, bimbingan, dan pembinaan yang bertujuan untuk memperbaiki kualitas pembelajaran di kelas [4]. Dengan adanya program supervisi akademik yang efektif, guru memiliki kesempatan untuk merefleksikan dan memperbaiki metode pengajaran, mengidentifikasi area pengembangan yang diperlukan, serta menerapkan strategi terbaik untuk memenuhi berbagai kebutuhan belajar murid [5].

Program supervisi akademik tidak hanya digunakan untuk mengukur keberhasilan pelaksanaan tugas guru, tetapi juga menjadi dasar dalam meningkatkan profesionalisme serta memperbaiki praktik pembelajaran secara berkelanjutan [6]. Namun pada pelaksanaan program supervisi akademik pada satuan pendidikan seringkali belum menunjukkan hasil yang optimal dalam peningkatan mutu pengajaran, dan hanya berhenti pada fungsi pengawasan proses pembelajaran saja. Berbagai kendala seperti kurangnya pemahaman guru terhadap tujuan dari program supervisi akademik, kurangnya sumber daya, serta keterbatasan sarana dan prasarana pendukung menjadi hambatan dalam mencapai tujuan program supervisi akademik. Program supervisi akademik hanya berjalan untuk memenuhi tugas administratif saja, belum menyentuh esensi substantif pembelajaran. Hasil dari program supervisi akademik juga seringkali berhenti hanya sebagai berkas administrasi tanpa proses refleksi dan tindak lanjut yang jelas dan terarah. Kondisi seperti itu menyebabkan program supervisi akademik belum memberikan dampak yang signifikan terhadap peningkatan kinerja guru dan kualitas proses pembelajaran.

Beberapa penelitian terdahulu menunjukkan bahwa supervisi akademik memiliki kontribusi positif terhadap peningkatan kompetensi guru dan kualitas pembelajaran. Choiriyah et al. (2024) menemukan bahwa supervisi akademik berperan dalam meningkatkan motivasi dan kinerja guru melalui proses pembinaan yang sistematis [3]. Fujiono et al. (2023) juga melaporkan bahwa pelaksanaan supervisi akademik yang efektif dapat mendukung pengembangan profesionalisme guru secara berkelanjutan [5]. Selain itu, Kasirin (2020) menunjukkan bahwa supervisi akademik yang dilakukan secara terencana mampu meningkatkan kualitas pelaksanaan pembelajaran di sekolah dasar [6]. Temuan-temuan tersebut mengindikasikan bahwa supervisi akademik merupakan program strategis dalam upaya peningkatan mutu pendidikan.

Di sisi lain, berbagai penelitian evaluasi program pendidikan telah memanfaatkan model CIPP sebagai kerangka evaluasi yang komprehensif karena mampu mengkaji aspek konteks, input, proses, dan produk secara terpadu. Zhang et al. (2011) menegaskan bahwa model CIPP dapat digunakan untuk mengevaluasi efektivitas suatu program secara menyeluruh [7], sedangkan Frye dan Hemmer (2012) menyatakan bahwa model ini membantu pengambil keputusan dalam menentukan perbaikan program berdasarkan bukti empiris [11]. Meskipun demikian, sebagian besar penelitian terdahulu berfokus pada implementasi supervisi akademik atau hasil evaluasi program, sedangkan penelitian yang secara khusus mengembangkan instrumen evaluasi supervisi akademik berbasis model CIPP yang teruji validitas dan reliabilitasnya masih terbatas.

Berdasarkan kajian tersebut, masih terdapat kesenjangan penelitian (*research gap*) berupa keterbatasan instrumen evaluasi supervisi akademik yang memiliki dasar teoritis kuat, mencakup seluruh dimensi CIPP, serta telah teruji validitas dan reliabilitasnya secara empiris. Sebagian instrumen yang digunakan dalam praktik supervisi akademik masih bersifat administratif dan belum mampu menghasilkan informasi evaluatif yang komprehensif sebagai dasar pengambilan keputusan program. Oleh karena itu, penelitian ini menawarkan kontribusi berupa pengembangan instrumen evaluasi program supervisi akademik berbasis model CIPP yang terdiri atas instrumen angket, wawancara, observasi, dan dokumentasi yang telah melalui proses validasi ahli serta pengujian karakteristik psikometrik secara empiris.

Untuk menjawab kebutuhan tersebut, diperlukan suatu kerangka evaluasi yang mampu menilai program supervisi akademik secara komprehensif dan sistematis. Evaluasi merupakan proses menggambarkan, memperoleh, dan menyediakan informasi yang berguna untuk menilai alternatif keputusan [7]. Evaluasi program merupakan serangkaian kegiatan yang dilakukan secara sistematis dan bertujuan untuk menilai tingkat keberhasilan suatu program [8]. Salah satu model evaluasi yang dapat diterapkan untuk mengevaluasi program supervisi akademik yaitu *CIPP* (*Context, Input, Process, Product*) yang dikembangkan oleh Stufflebeam. Model evaluasi *CIPP* memungkinkan evaluasi secara menyeluruh terhadap kebutuhan program, ketersediaan sumber daya, proses pelaksanaan program, dan hasil yang dicapai dari pelaksanaan program yang dievaluasi. Oleh karena itu diperlukan pengembangan instrumen evaluasi program supervisi akademik berdasarkan *CIPP* yang dapat memberikan gambaran menyeluruh mengenai kekurangan dan kelebihan program supervisi akademik, serta saran perbaikan yang mungkin untuk dilakukan. Penelitian ini bertujuan untuk menghasilkan instrumen evaluasi program supervisi akademik yang valid dan reliabel.

2. METODE PENELITIAN

Penelitian ini bertujuan untuk mengembangkan instrumen evaluasi program supervisi akademik berbasis model evaluasi *CIPP* (*Context, Input, Process, Product*). Model *CIPP* dipilih karena mampu mengevaluasi program secara komprehensif dan sistematis, tidak hanya pada hasil, tetapi juga pada aspek konteks, sumber daya, proses pelaksanaan dan hasil dari pelaksanaan program.

Pada aspek konteks (*context*) bertujuan untuk mengevaluasi latar belakang program, kebutuhan supervisi akademik, tujuan program, kebijakan dan visi misi sekolah dalam meningkatkan kualitas pembelajaran. Pada aspek sumber daya (*input*) bertujuan untuk mengevaluasi sumber daya yang digunakan dalam supervisi akademik, seperti kompetensi supervisor, kesiapan guru, sarana dan prasarana, instrumen supervisi, anggaran program, serta perencanaan program. Pada aspek proses pelaksanaan (*process*) bertujuan untuk mengevaluasi pelaksanaan supervisi akademik, observasi pembelajaran, pembinaan dan umpan balik, monitoring, dan tindak lanjut. Dan pada aspek hasil pelaksanaan (*product*) bertujuan untuk mengevaluasi hasil supervisi akademik, khususnya dampaknya terhadap kompetensi guru, peningkatan kualitas pembelajaran, serta peningkatan mutu pendidikan [9] [10].

Instrumen yang dikembangkan dalam penelitian ini meliputi angket, lembar wawancara, lembar observasi, dan lembar dokumentasi. Instrumen angket digunakan untuk memperoleh data persepsi guru terhadap pelaksanaan program supervisi akademik dan dikembangkan berdasarkan empat komponen *CIPP* dan terdiri dari 52 butir dengan skala *Likert* empat kategori, yaitu sangat tidak setuju (1), tidak setuju (2), setuju (3), dan sangat setuju (4). Penggunaan skala empat kategori bertujuan untuk menghindari kecenderungan responden memilih jawaban netral. Instrumen wawancara (*in-depth interview*) dikembangkan untuk kepala sekolah, guru, dan pihak terkait untuk memperoleh informasi mengenai pelaksanaan program supervisi akademik. Instrumen observasi dikembangkan berdasarkan indikator kualitas pembelajaran dan pelaksanaan program supervisi akademik. Observasi dilakukan untuk mengamati secara langsung pelaksanaan supervisi akademik, proses pembelajaran di kelas, dan interaksi antara guru dengan murid. Instrumen dokumentasi digunakan untuk melengkapi data penelitian meliputi instrumen supervisi akademik, hasil pelaksanaan observasi kelas, RPP yang digunakan dalam pelaksanaan program supervisi akademik.

Pengembangan instrumen dilakukan melalui beberapa tahapan yaitu (1) analisis kebutuhan dan kajian teori, (2) penyusunan konstruk instrumen, (3) penyusunan kisi-kisi instrumen, (4) pengembangan butir instrumen, (5) validasi ahli, (6) revisi instrumen, (7) uji coba instrumen. Tahap analisis kebutuhan dilakukan melalui kajian literatur mengenai supervisi akademik, evaluasi program pendidikan, dan model evaluasi *CIPP* yang dikembangkan oleh Stufflebeam. Hasil kajian digunakan sebagai dasar dalam merumuskan konstruk, indikator, dan subindikator instrumen evaluasi program supervisi akademik. Dari konstruk, indikator, dan subindikator instrumen yang telah dirumuskan, kemudian dikembangkan menjadi butir instrumen.

Validitas isi instrumen dievaluasi melalui *expert judgement* yang melibatkan tiga validator ahli yang memiliki kompetensi di bidang evaluasi pendidikan, supervisi akademik, dan pengembangan instrumen. Validitas isi dianalisis menggunakan indeks *Aiken's V*. Setelah proses validasi isi, instrumen direvisi berdasarkan masukan dan rekomendasi para validator.

Instrumen yang telah direvisi selanjutnya diuji cobakan kepada 100 guru SMA yang berasal dari beberapa sekolah menengah atas. Responden dipilih menggunakan teknik *purposive sampling* dengan kriteria telah mengikuti program supervisi akademik yang dilaksanakan oleh kepala sekolah atau pengawas sekolah dalam satu tahun terakhir. Karakteristik responden terdiri atas guru laki-laki dan perempuan dengan masa kerja yang bervariasi sehingga dianggap mampu memberikan informasi yang representatif mengenai pelaksanaan program supervisi akademik. Jumlah responden sebanyak 100 orang dinilai memadai untuk menguji karakteristik psikometrik instrumen pada tahap pengembangan awal.

Data hasil uji coba kemudian dianalisis untuk mengetahui validitas empiris dan reliabilitas instrumen. Validitas empiris dianalisis menggunakan *Corrected Item Total Correlation* (CITC), sedangkan reliabilitas dianalisis menggunakan koefisien *Cronbach Alpha*. Hasil analisis validitas dan reliabilitas digunakan sebagai dasar untuk menetapkan kelayakan instrumen evaluasi program supervisi akademik berbasis model CIPP yang dikembangkan. Produk akhir penelitian ini berupa seperangkat instrumen evaluasi program supervisi akademik yang terdiri atas angket, pedoman wawancara, lembar observasi, dan lembar dokumentasi yang telah memenuhi kriteria validitas dan reliabilitas untuk digunakan dalam evaluasi program supervisi akademik di sekolah.

3. HASIL

Penelitian ini bertujuan mengembangkan instrumen evaluasi program supervisi akademik berbasis model CIPP (*Context, Input, Process, Product*) yang valid dan reliabel. Instrumen yang dikembangkan terdiri atas angket, pedoman wawancara, lembar observasi, dan lembar dokumentasi yang disusun berdasarkan komponen *context, input, process, dan product*. Sebelum digunakan dalam evaluasi program supervisi akademik, seluruh instrumen terlebih dahulu diuji validitas isi (*content validity*) melalui *expert judgement* yang melibatkan tiga validator yang memiliki kompetensi di bidang supervisi akademik dan evaluasi pendidikan.

Analisis validasi isi instrumen dianalisis menggunakan indeks *Aiken's V*. Penilaian dilakukan menggunakan skala empat kategori 1 (tidak relevan), 2 (kurang relevan), 3 (relevan), dan 4 (sangat relevan). Nilai *Aiken's V* dihitung dengan rumus:

$$V = \frac{\sum s}{[n(c-1)]} \quad (1)$$

Instrumen angket yang berfungsi sebagai instrumen kuantitatif selanjutnya diuji validitas empiris menggunakan *Corrected Item Total Correlation* (CITC) dan reliabilitas menggunakan koefisien Cronbach Alpha. Analisis dilakukan dengan bantuan *software R-studio*. Butir instrumen dinyatakan valid jika memiliki nilai CITC $\geq 0,30$, sedangkan reliabilitas instrumen dinyatakan memadai jika nilai koefisien *Cronbach Alpha* $\geq 0,80$.

Instrumen wawancara, observasi, dan dokumentasi tidak dianalisis menggunakan CITC maupun *Cronbach Alpha* karena berfungsi sebagai instrumen kualitatif dalam penelitian evaluasi program supervisi akademik dengan pendekatan *mixed-methods*. Kualitas instrumen tersebut hanya diuji menggunakan validitas isi melalui *expert judgement* dengan indeks *Aiken's V*.

3.1. Uji Validitas dan Reliabilitas Instrumen Angket

Hasil uji validitas isi terhadap instrumen angket yang terdiri dari 52 butir pernyataan menunjukkan bahwa nilai *Aiken's V* berkisar antara 0,778 sampai dengan 1,000. Sebanyak 51 butir memperoleh nilai *Aiken's V* $\geq 0,80$ sehingga dinyatakan valid, sedangkan 1 butir (B39) memperoleh nilai *Aiken's V* sebesar 0,778 sehingga perlu revisi sesuai dengan saran dari validator.

Berikut hasil analisis nilai *Aiken's V* pada instrumen angket:

Tabel Hasil Uji Validasi Isi Instrumen Angket

Item	<i>Aiken's V</i>	Keputusan	Item	<i>Aiken's V</i>	Keputusan	Item	<i>Aiken's V</i>	Keputusan
B1	1,00	Valid	B19	1,00	Valid	B37	1,00	Valid
B2	1,00	Valid	B20	1,00	Valid	B38	1,00	Valid
B3	1,00	Valid	B21	1,00	Valid	B39	0,78	Revisi
B4	1,00	Valid	B22	1,00	Valid	B40	0,89	Valid
B5	1,00	Valid	B23	1,00	Valid	B41	1,00	Valid
B6	1,00	Valid	B24	1,00	Valid	B42	1,00	Valid
B7	1,00	Valid	B25	1,00	Valid	B43	1,00	Valid
B8	1,00	Valid	B26	1,00	Valid	B44	1,00	Valid
B9	1,00	Valid	B27	1,00	Valid	B45	1,00	Valid
B10	1,00	Valid	B28	1,00	Valid	B46	1,00	Valid
B11	1,00	Valid	B29	1,00	Valid	B47	1,00	Valid
B12	1,00	Valid	B30	1,00	Valid	B48	1,00	Valid
B13	1,00	Valid	B31	1,00	Valid	B49	1,00	Valid
B14	0,89	Valid	B32	1,00	Valid	B50	1,00	Valid
B15	1,00	Valid	B33	1,00	Valid	B51	1,00	Valid
B16	1,00	Valid	B34	1,00	Valid	B52	1,00	Valid
B17	1,00	Valid	B35	1,00	Valid			

Item	<i>Aiken's V</i>	Keputusan	Item	<i>Aiken's V</i>	Keputusan	Item	<i>Aiken's V</i>	Keputusan
B18	1,00	Valid	B36	1,00	Valid			

Butir B39 memperoleh nilai *Aiken's V* sebesar 0,778 sehingga memerlukan revisi sebelum dilakukan uji coba instrumen. Berdasarkan masukan dari validator, butir tersebut dinilai memiliki kemiripan substansi dengan butir B38 sehingga berpotensi menimbulkan redundansi dalam pengukuran. Oleh karena itu, dilakukan perbaikan redaksi dengan memperjelas fokus pernyataan agar lebih merepresentasikan indikator yang diukur dan memiliki perbedaan yang jelas dengan butir sebelumnya. Setelah dilakukan revisi, butir B39 diikutsertakan dalam uji coba instrumen dan memperoleh nilai *Corrected Item Total Correlation* (CITC) sebesar 0,684 sehingga dinyatakan valid secara empiris.

Setelah dilakukan revisi pada instrumen, selanjutnya instrumen diuji coba kepada 100 responden. Hasil analisis menunjukkan bahwa seluruh butir instrumen memiliki nilai *Corrected Item Total Correlation* (CITC) di atas 0,30 dengan rentang nilai antara 0,657 sampai dengan 0,829. Dengan demikian seluruh butir dinyatakan valid. Berikut hasil uji CITC pada instrumen angket:

Tabel Hasil Uji CITC Instrumen Angket

Item	CITC	Keputusan	Item	CITC	Keputusan	Item	CITC	Keputusan
B1	0,735	Layak	B19	0,696	Layak	B37	0,800	Layak
B2	0,736	Layak	B20	0,735	Layak	B38	0,805	Layak
B3	0,736	Layak	B21	0,796	Layak	B39	0,684	Layak
B4	0,672	Layak	B22	0,771	Layak	B40	0,756	Layak
B5	0,765	Layak	B23	0,812	Layak	B41	0,819	Layak
B6	0,657	Layak	B24	0,773	Layak	B42	0,757	Layak
B7	0,761	Layak	B25	0,750	Layak	B43	0,750	Layak
B8	0,726	Layak	B26	0,788	Layak	B44	0,767	Layak
B9	0,719	Layak	B27	0,829	Layak	B45	0,710	Layak
B10	0,728	Layak	B28	0,714	Layak	B46	0,780	Layak
B11	0,731	Layak	B29	0,749	Layak	B47	0,752	Layak
B12	0,713	Layak	B30	0,697	Layak	B48	0,769	Layak
B13	0,749	Layak	B31	0,758	Layak	B49	0,706	Layak
B14	0,754	Layak	B32	0,677	Layak	B50	0,773	Layak
B15	0,783	Layak	B33	0,808	Layak	B51	0,793	Layak
B16	0,730	Layak	B34	0,716	Layak	B52	0,806	Layak
B17	0,768	Layak	B35	0,739	Layak			
B18	0,753	Layak	B36	0,761	Layak			

Hasil uji validitas empiris menunjukkan bahwa seluruh butir instrumen memiliki nilai *Corrected Item Total Correlation* (CITC) di atas batas minimum 0,30. Nilai CITC berkisar antara 0,657 hingga 0,829 yang menunjukkan bahwa setiap butir memiliki korelasi yang memadai dengan skor total instrumen. Dengan demikian, seluruh butir angket dinyatakan valid secara empiris dan layak digunakan dalam pengumpulan data.

Uji reliabilitas menggunakan koefisien *Cronbach Alpha* menghasilkan nilai sebesar 0,986. Nilai tersebut berada jauh di atas batas minimum reliabilitas yang direkomendasikan sebesar 0,80. Hasil ini menunjukkan bahwa instrumen memiliki konsistensi internal yang sangat tinggi sehingga dapat menghasilkan data yang stabil dan konsisten ketika digunakan untuk mengevaluasi program supervisi akademik.

3.2. Uji Validitas Isi Instrumen Observasi

Lembar observasi pelaksanaan program supervisi akademik dalam penelitian ini digunakan sebagai instrumen kualitatif untuk memperoleh data pendukung dan melakukan triangulasi terhadap terhadap hasil angket dan hasil wawancara. Oleh karena itu instrumen observasi dijamin kualitasnya melalui validasi isi oleh tiga validator menggunakan indeks *Aiken's V*.

Tabel Hasil Uji Validasi Isi Instrumen Observasi

ITEM	<i>Aiken's V</i>	Keputusan
B1	1,00	Valid
B2	1,00	Valid
B3	1,00	Valid
B4	1,00	Valid

ITEM	<i>Aiken's V</i>	Keputusan
B5	1,00	Valid
B6	1,00	Valid
B7	1,00	Valid
B8	1,00	Valid
B9	1,00	Valid
B10	0,89	Valid
B11	1,00	Valid
B12	0,78	Revisi
B13	1,00	Valid
B14	1,00	Valid
B15	1,00	Valid

Satu indikator observasi, yaitu B12, memperoleh nilai *Aiken's V* sebesar 0,78 sehingga dilakukan revisi redaksi sesuai dengan masukan validator sebelum digunakan dalam pengumpulan data. Setelah dilakukan perbaikan, indikator tersebut tetap dipertahankan karena dinilai relevan dengan aspek yang diukur dalam evaluasi program supervisi akademik.

3.3. Uji Validitas Isi Instrumen Wawancara

Validitas isi instrumen wawancara kepala sekolah, wakil kepala sekolah bidang kurikulum, dan guru dianalisis menggunakan indeks *Aiken's V* berdasarkan penilaian dari tiga validator. Berikut hasil uji validitas isi menggunakan *Aiken's V* pada instrumen wawancara:

Tabel Hasil Uji Validasi Isi Instrumen Wawancara

ITEM	<i>Aiken's V</i>	Keputusan
B1	1,00	Valid
B2	1,00	Valid
B3	0,89	Valid
B4	1,00	Valid
B5	0,89	Valid
B6	1,00	Valid
B7	1,00	Valid
B8	1,00	Valid
B9	1,00	Valid
B10	0,89	Valid
B11	1,00	Valid
B12	1,00	Valid
B13	1,00	Valid
B14	0,89	Valid
B15	1,00	Valid

Hasil analisis menunjukkan bahwa seluruh pertanyaan memperoleh nilai *Aiken's V* di atas 0,80 sehingga seluruh pertanyaan dinyatakan valid dan layak digunakan untuk memperoleh data kualitatif terkait pelaksanaan program supervisi akademik.

3.4. Uji Validitas Isi Instrumen Dokumentasi

Instrumen dokumentasi yang digunakan untuk menelaah berbagai dokumen terkait supervisi akademik juga telah divalidasi oleh tiga validator. Berikut hasil uji validasi menggunakan *Aiken's V* pada instrumen dokumentasi:

Tabel Hasil Uji Validasi Isi Instrumen Dokumentasi

ITEM	<i>Aiken's V</i>	Keputusan
B1	1,00	Valid
B2	1,00	Valid
B3	0,89	Valid
B4	1,00	Valid
B5	0,89	Valid

ITEM	<i>Aiken's V</i>	Keputusan
B6	1,00	Valid
B7	1,00	Valid
B8	1,00	Valid
B9	1,00	Valid
B10	0,89	Valid
B11	1,00	Valid
B12	1,00	Valid
B13	1,00	Valid
B14	0,89	Valid
B15	1,00	Valid
B16	1,00	Valid
B17	1,00	Valid
B18	1,00	Valid
B19	1,00	Valid

Hasil analisis menunjukkan bahwa seluruh indikator dokumentasi memperoleh nilai *Aiken's V* di atas 0,80. Dengan demikian seluruh indikator dinyatakan valid dan layak digunakan sebagai pedoman dokumentasi penelitian.

4. PEMBAHASAN

Evaluasi pendidikan tidak hanya berfungsi sebagai alat monitoring, tetapi juga sebagai dasar pengambilan keputusan dan perbaikan berkelanjutan program sekolah. Oleh karena itu, instrumen evaluasi yang valid dan reliabel menjadi prasyarat penting dalam menghasilkan informasi evaluatif yang akurat. Hasil penelitian menunjukkan bahwa instrumen evaluasi program supervisi akademik berbasis model CIPP yang dikembangkan memiliki kualitas psikometrik yang sangat baik. Hal tersebut ditunjukkan oleh hasil validitas isi yang memperoleh nilai *Aiken's V* berkisar antara 0,778–1,000. Nilai tersebut menunjukkan bahwa sebagian besar indikator dan butir instrumen telah memiliki kesesuaian yang tinggi dengan konstruk yang diukur. Menurut Aiken (1985), indeks *Aiken's V* digunakan untuk mengukur tingkat kesepakatan para ahli terhadap relevansi suatu butir instrumen dengan konstruk yang diukur [12]. Semakin tinggi nilai *Aiken's V*, semakin tinggi pula tingkat representativitas butir terhadap konstruk yang dikembangkan. Hasil penelitian ini menunjukkan bahwa instrumen yang dikembangkan telah memenuhi aspek validitas isi sehingga layak digunakan untuk mengevaluasi program supervisi akademik.

Temuan ini sejalan dengan pendapat Azwar (2017) bahwa validitas isi merupakan tahapan penting dalam pengembangan instrumen karena menentukan sejauh mana butir instrumen mampu merepresentasikan domain konstruk yang diukur [13]. Temuan ini juga didukung oleh Nurhayati dan Mardapi (2024) yang menyatakan bahwa kualitas instrumen pendidikan sangat dipengaruhi oleh ketepatan konstruk dan kesesuaian indikator dengan tujuan pengukuran [14]. Tingginya nilai *Aiken's V* pada instrumen yang dikembangkan menunjukkan bahwa indikator context, input, process, dan product telah terwakili secara memadai dalam setiap instrumen yang disusun, baik angket, wawancara, observasi, maupun dokumentasi. Dengan demikian, instrumen yang dihasilkan memiliki landasan konseptual yang kuat sesuai dengan kerangka evaluasi CIPP yang dikembangkan oleh Stufflebeam dan Shinkfield (2007).

Hasil uji validitas empiris menunjukkan bahwa seluruh butir angket memiliki nilai *Corrected Item Total Correlation* (CITC) di atas 0,30 dengan rentang nilai 0,657–0,829. Hasil tersebut menunjukkan bahwa seluruh butir memiliki kemampuan yang baik dalam membedakan responden dan berkontribusi secara positif terhadap skor total instrumen. Menurut DeVellis (2017), nilai korelasi item-total yang memadai menunjukkan bahwa setiap butir memiliki konsistensi dengan konstruk yang diukur dan mampu memberikan informasi yang relevan terhadap pengukuran [15]. Dengan demikian, seluruh butir dalam instrumen ini dinyatakan valid secara empiris dan layak digunakan sebagai alat pengumpulan data.

Selain validitas, kualitas instrumen juga ditunjukkan oleh hasil uji reliabilitas yang menghasilkan koefisien *Cronbach Alpha* sebesar 0,986. Nilai tersebut menunjukkan tingkat konsistensi internal yang sangat tinggi. Menurut Hair et al. (2019), instrumen dengan nilai *Cronbach Alpha* di atas 0,80 dapat dikategorikan memiliki reliabilitas yang baik, sedangkan nilai di atas 0,90 menunjukkan reliabilitas yang sangat tinggi [16]. Oleh karena itu, instrumen evaluasi program supervisi akademik yang dikembangkan memiliki kemampuan menghasilkan data yang konsisten dan stabil ketika digunakan pada kelompok responden yang memiliki karakteristik serupa.

Keunggulan lain dari instrumen yang dikembangkan terletak pada penggunaan model evaluasi CIPP sebagai dasar penyusunan konstruk instrumen. Model CIPP memungkinkan evaluasi dilakukan secara komprehensif

karena tidak hanya berfokus pada hasil program, tetapi juga mempertimbangkan aspek kebutuhan, sumber daya, proses pelaksanaan, dan hasil yang dicapai. Zhang et al. (2011) menjelaskan bahwa model CIPP merupakan salah satu model evaluasi yang mampu memberikan informasi yang lengkap bagi pengambilan keputusan program [7]. Temuan tersebut sejalan dengan penelitian Sari dan Widoyoko (2024) yang menunjukkan bahwa model CIPP masih menjadi pendekatan evaluasi yang relevan karena mampu mengintegrasikan evaluasi konteks, sumber daya, proses, dan hasil secara sistematis [17]. Oleh karena itu, instrumen yang dikembangkan dalam penelitian ini berpotensi membantu sekolah memperoleh informasi yang lebih komprehensif mengenai efektivitas pelaksanaan supervisi akademik. Temuan ini juga sejalan dengan Khofifah et al. (2025) yang menunjukkan bahwa evaluasi supervisi akademik berbasis model CIPP mampu memberikan informasi yang komprehensif mengenai efektivitas pelaksanaan supervisi serta kontribusinya terhadap peningkatan profesionalisme guru [18].

Secara praktis, instrumen ini dapat digunakan oleh kepala sekolah, pengawas sekolah, maupun peneliti pendidikan untuk mengevaluasi pelaksanaan program supervisi akademik secara lebih sistematis. Hasil penelitian ini memperkuat temuan Suri (2024) yang menyatakan bahwa model CIPP merupakan salah satu model evaluasi yang paling banyak digunakan dalam evaluasi program pendidikan karena mampu mengintegrasikan evaluasi konteks, input, proses, dan produk secara sistematis [19]. Hasil evaluasi yang diperoleh dapat dimanfaatkan sebagai dasar dalam menyusun program pembinaan guru, meningkatkan kualitas pembelajaran, dan memperkuat pengambilan keputusan berbasis data di sekolah. Hasil penelitian ini juga mendukung temuan Prasetyo et al. (2023) yang menegaskan bahwa supervisi akademik yang dievaluasi secara sistematis dapat menjadi dasar dalam meningkatkan kualitas pembelajaran dan profesionalisme guru secara berkelanjutan [20]. Dengan tersedianya instrumen yang valid dan reliabel, evaluasi program supervisi akademik diharapkan tidak lagi terbatas pada pemenuhan administrasi, tetapi mampu menjadi sarana refleksi dan perbaikan berkelanjutan dalam peningkatan mutu pendidikan.

5. KESIMPULAN

Penelitian ini berhasil mengembangkan instrumen evaluasi program supervisi akademik berbasis model CIPP (Context, Input, Process, Product) yang terdiri atas instrumen angket, wawancara, observasi, dan dokumentasi. Hasil pengujian menunjukkan bahwa instrumen yang dikembangkan telah memenuhi kriteria validitas dan reliabilitas sehingga layak digunakan sebagai alat evaluasi program supervisi akademik di sekolah.

Instrumen yang dikembangkan mampu mengakomodasi evaluasi program supervisi akademik secara komprehensif melalui aspek context, input, process, dan product. Keberadaan instrumen ini diharapkan dapat membantu kepala sekolah, pengawas, maupun peneliti pendidikan dalam memperoleh informasi yang lebih sistematis dan objektif untuk mendukung pengambilan keputusan serta perbaikan berkelanjutan terhadap pelaksanaan supervisi akademik.

Penelitian selanjutnya disarankan untuk melakukan pengujian instrumen pada cakupan responden yang lebih luas dan beragam, serta menguji validitas konstruk menggunakan analisis faktor eksploratori maupun konfirmatori sehingga kualitas psikometrik instrumen dapat semakin diperkuat.

DAFTAR PUSTAKA

- [1] Daryanto. *Supervisi Pembelajaran*. Gava Media. 2015
- [2] Glickman, C. D., Gordon, S. P., & Ross-Gordon, J. M. *Supervision and instructional leadership: A developmental approach* (10th ed.). Pearson. 2018.
- [3] Choiriyah, N., Hariyadi, A., & Suadi. Supervisi akademik dalam meningkatkan motivasi guru. *Scientia: Social Sciences & Humanities*, 3(2), 220–223. 2024.
- [4] Sergiovanni, T. J. *The principalship: A reflective practice perspective*. Allyn & Bacon. 1987.
- [5] Fujiono, F., Hanayanti, C. S., Suwarna, A. I., Rais, R., & Ridani, A. Developing teacher professionalism through academic supervision activities of school principal. *Journal on Education*, 5(4), 14881–14885. 2023.
- [6] Kasirin, K. The efforts to improve teacher performance through academic supervision by the principal in primary school. *JP2D (Jurnal Penelitian Pendidikan Dasar) Untan*, 3(3), 112–118. 2020.
- [7] Zhang, G., Zeller, N., Griffith, R., Metcalf, D., Williams, J., Shea, C., & Misulis, K. Using the context, input, process, and product evaluation model (CIPP) as a comprehensive framework to guide the planning, implementation, and assessment of service-learning programs. *Journal of Higher Education Outreach and Engagement*. 15(4), 57–84. 2011.
- [8] Frye, A. W., & Hemmer, P. A. Program evaluation models and related theories: AMEE guide no. 67. *Medical Teacher*, 34(5), e288–e299. 2012. <https://doi.org/10.3109/0142159X.2012.668637>

-
- [9] Susanto, A. Teori belajar dan pembelajaran di sekolah dasar. *Kencana*. 2016.
- [10] Stufflebeam, D. L., & Shinkfield, A. J. Evaluation theory, models, and applications. *Jossey-Bass*. 2007.
- [11] Andri, S. J. A., & Rahmadani. Evaluasi program praktek kerja lapangan siswa SMK Negeri 1 Koto XI Tarusan dengan menggunakan logic model. *Journal of Sciencetech Research and Development*, 6(1), 333–342. 2024.
- [12] Aiken, L. R. Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. 1985. <https://doi.org/10.1177/0013164485451012>
- [13] Azwar, S. Validitas dan Reliabilitas. Edisi ke-4. Pustaka Pelajar. 2017.
- [14] Nurhayati, N., & Mardapi, D. Instrument development and psychometric evaluation in educational research: Recent perspectives. *International Journal of Evaluation and Research in Education*, 13(2), 987–996. 2024.
- [15] DeVellis, R. F. Scale Development: Theory and Applications (4th ed.). Sage Publications. 2017.
- [16] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. Multivariate Data Analysis (8th ed.). Cengage Learning. 2019.
- [17] Sari, D. P., & Widoyoko, E. P. Evaluation of educational programs using the CIPP model: Trends and challenges in Indonesian educational research. *Jurnal Cakrawala Pendidikan*, 43(1), 52–64. 2024.
- [18] Khofifah, J. M., Nurdin, D., & Herawan, E. *Enhancing Teacher Professionalism Through Academic Supervision: A CIPP Model Evaluation*. *Indonesian Journal of Educational Development*, 2025, 6(2), 380–392. DOI: 10.59672/ijed.v6i2.4727
- [19] Suri, S & Hariyatii, N. Literature Study: CIPP Evaluation Model in The Educational Evaluation. *Jurnal Cahaya Pendidikan*. 2024. 10(1), 10-19. <https://doi.org.10.33373/chypen.v10i1.5950>
- [20] Prasetyo, M. A. M., Bashori, B., & Wahyudi, W. Academic supervision in improving teacher performance: A systematic literature review. *Journal of Education and Learning*, 17(3), 395–405. 2023.