

Studi Komparatif Validitas dan Reliabilitas Soal Buatan Guru antara Sekolah Dasar dan Madrasah Ibtidaiyah

Oleh

Mujiansyah¹

mujiansyah@uin-antasari.ac.id

Halimatus Sa'diyah²

halimatussadiyahhh10@gmail.com

Ziyadatunni'mah³

ziyadatun232@gmail.com

Korespondensi:

¹Dosen PGMI UIN Antasari

Banjarmasin

²Mahasiswa PGMI UIN Antasari

Banjarmasin

³Mahasiswa PGMI UIN Antasari

Banjarmasin

ABSTRAK

Penelitian ini bertujuan menganalisis dan membandingkan validitas dan reliabilitas soal buatan guru pada mata pelajaran Bahasa Indonesia di SD Kuin Utara 5 kelas VI dan Seni Budaya dan Prakarya di MI Inayatushshibyan II Kelas V Banjarmasin. Penelitian menggunakan pendekatan kuantitatif dengan desain komparatif, melibatkan analisis 40 butir soal Bahasa Indonesia dan 40 butir soal Seni Budaya dan Prakarya. Pengumpulan data dilakukan melalui dokumentasi soal, lembar jawaban siswa, dan wawancara dengan guru pengampu mata pelajaran. Analisis validitas mencakup validitas isi, konstruk, dan kriteria menggunakan formula Aiken's V dan Point Biserial, sedangkan reliabilitas dianalisis menggunakan KR-20 dan ICC. Hasil penelitian menunjukkan soal Bahasa Indonesia memiliki validitas dan reliabilitas yang lebih tinggi (Aiken's V = 0.78; KR-20 = 0.82) dibandingkan soal Seni Budaya dan Prakarya (Aiken's V = 0.75; KR-20 = 0.78). Perbedaan ini dipengaruhi oleh karakteristik mata pelajaran, kualifikasi guru, dan intensitas program pengembangan profesional. Soal Bahasa Indonesia unggul dalam mengukur aspek kognitif dan literasi, sementara soal Seni Budaya dan Prakarya lebih baik dalam mengintegrasikan penilaian teori dan praktik. Penelitian merekomendasikan pengembangan program kolaboratif antar sekolah untuk meningkatkan kualitas instrumen evaluasi pembelajaran.

Kata kunci: validitas dan reliabilitas, soal buatan guru, evaluasi pembelajaran

PENDAHULUAN

Evaluasi pembelajaran merupakan komponen vital dalam sistem pendidikan dasar yang berfungsi sebagai alat ukur ketercapaian tujuan pembelajaran (Brookhart & McMillan, 2020). Khususnya pada mata pelajaran Bahasa Indonesia di SD dan Seni Budaya dan Prakarya di MI, evaluasi berperan penting dalam mengukur kompetensi siswa secara holistik. Proses evaluasi yang berkualitas memerlukan instrumen penilaian yang valid dan reliabel untuk menghasilkan data yang akurat tentang kemampuan siswa. Guru sebagai ujung tombak pendidikan memiliki tanggung jawab besar dalam mengembangkan instrumen evaluasi yang sesuai dengan karakteristik peserta didik. Menurut Zhang & Burry-Stock (2021), kualitas soal yang dibuat guru berpengaruh langsung terhadap akurasi pengukuran capaian belajar siswa.

Fenomena di lapangan menunjukkan bahwa masih terdapat variasi signifikan dalam kualitas soal buatan guru untuk mata pelajaran Bahasa Indonesia dan Seni Budaya dan Prakarya (Richardson et al., 2023). Beberapa penelitian mengindikasikan bahwa 45% guru sekolah dasar mengalami kesulitan dalam mengembangkan soal yang memenuhi kriteria validitas dan reliabilitas. Studi yang dilakukan Ahmad & Sutanto (2022) menemukan bahwa hanya 60% soal buatan guru yang memenuhi standar validitas konstruk. Kemampuan guru dalam mengembangkan instrumen evaluasi sering terkendala oleh pemahaman yang terbatas tentang prinsip-prinsip pengukuran pendidikan. Tantangan ini semakin kompleks dengan adanya tuntutan implementasi Kurikulum Merdeka.

SD Kuin Utara 5 dan MI Inayatushshibyan II Banjarmasin memiliki karakteristik pembelajaran yang berbeda dalam pendekatan evaluasi pembelajaran. Mata pelajaran Bahasa Indonesia di SD Kuin Utara 5 menekankan pada kemampuan literasi dan komunikasi, sementara Seni Budaya dan Prakarya di MI Inayatushshibyan II fokus pada pengembangan kreativitas dan apresiasi seni. Perbedaan orientasi pembelajaran ini mempengaruhi konstruksi soal evaluasi yang dikembangkan guru. Pendekatan pembelajaran berbasis literasi di SD membutuhkan soal-soal yang mengukur kemampuan membaca, menulis dan berbahasa secara komprehensif. Menurut Hamdani & Rahmat (2023), variasi dalam pendekatan pembelajaran membutuhkan instrumen evaluasi yang berbeda pula.

Validitas dan reliabilitas merupakan dua aspek fundamental yang menentukan kualitas instrumen penilaian (Anderson & Morgan, 2023). Untuk mata pelajaran Bahasa Indonesia dan Seni Budaya dan Prakarya, validitas soal harus mencakup aspek kognitif dan keterampilan secara seimbang. Studi Wijaya et al. (2024) mengungkapkan bahwa hanya 55% instrumen penilaian di tingkat sekolah dasar yang memenuhi standar validitas dan reliabilitas. Konsistensi hasil pengukuran menjadi perhatian utama dalam pengembangan instrumen evaluasi yang berkualitas. Pentingnya kedua aspek ini semakin meningkat dengan adanya tuntutan akuntabilitas pendidikan.

Data Dinas Pendidikan Kota Banjarmasin tahun 2023 menunjukkan terdapat 210 SD dan 45 MI yang tersebar di lima kecamatan. Rasio guru berkualifikasi S1 pengampu Bahasa Indonesia di SD mencapai 85%, sementara guru Seni Budaya dan Prakarya di MI sebesar 75%. Perbedaan kualifikasi ini berpotensi mempengaruhi kualitas instrumen evaluasi yang dikembangkan. Distribusi sekolah yang tidak merata juga mempengaruhi akses terhadap program pengembangan profesional guru. Menurut data Kemendikbudristek (2023), rata-rata nilai evaluasi pembelajaran di kedua mata pelajaran menunjukkan disparitas yang perlu dikaji lebih lanjut.

Penelitian terdahulu sebagian besar berfokus pada analisis validitas dan reliabilitas secara terpisah untuk masing-masing mata pelajaran. Rahman & Sulistyowati (2023) mengkaji validitas soal Bahasa Indonesia di SD tanpa membandingkan dengan mata pelajaran lain di MI. Penelitian Hidayat et al. (2024) tentang reliabilitas soal Seni Budaya dan Prakarya di MI belum memasukkan perspektif komparatif dengan SD. Kajian komparatif yang ada belum mempertimbangkan karakteristik unik masing-masing mata pelajaran. Kesenjangan ini menunjukkan perlunya studi komprehensif yang membandingkan kualitas soal di kedua sekolah.

SD Kuin Utara 5 menunjukkan upaya sistematis dalam pengembangan instrumen evaluasi Bahasa Indonesia. Program pengembangan profesional guru mencakup pelatihan pembuatan soal berbasis literasi. Monitoring evaluasi pembelajaran dilakukan secara berkala oleh tim supervisi akademik. Soal-soal yang dikembangkan guru telah melalui proses telaah internal untuk memastikan kesesuaian dengan tujuan pembelajaran. Namun, belum ada kajian mendalam tentang validitas dan reliabilitas soal yang dihasilkan.

MI Inayatushshibyan II memiliki keunggulan dalam mengembangkan instrumen evaluasi Seni Budaya dan Prakarya yang integratif. Guru-guru rutin mengikuti workshop pengembangan instrumen penilaian berbasis kreativitas dan apresiasi seni. Soal-soal yang dikembangkan memperhatikan aspek keterampilan dan pengetahuan secara seimbang. Proses evaluasi

pembelajaran mencakup penilaian teori dan praktik yang komprehensif. Aspek teknis validitas dan reliabilitas soal masih perlu pengkajian lebih lanjut untuk memastikan kualitas pengukuran.

Kesenjangan dalam penelitian sebelumnya terletak pada minimnya studi yang menganalisis faktor-faktor penyebab perbedaan kualitas soal antara kedua mata pelajaran tersebut. Penelitian Gunawan & Azizah (2023) berfokus pada aspek teknis tanpa mempertimbangkan karakteristik mata pelajaran. Studi yang dilakukan Prasetyo et al. (2024) belum mengeksplorasi pengaruh pendekatan pembelajaran terhadap kualitas soal. Aspek kontekstual seperti budaya sekolah dan karakteristik siswa belum mendapat perhatian memadai. Analisis komparatif yang komprehensif diperlukan untuk memahami dinamika pengembangan instrumen evaluasi di kedua mata pelajaran.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan menganalisis dan membandingkan validitas dan reliabilitas soal buatan guru pada mata pelajaran Bahasa Indonesia dan Seni Budaya dan Prakarya. Pemilihan SD Kuin Utara 5 dan MI Inayatushshibyan II sebagai lokasi penelitian didasarkan pada karakteristik kedua sekolah yang representatif. Studi ini mempertimbangkan faktor internal dan eksternal yang mempengaruhi kualitas instrumen evaluasi. Proses pengembangan soal di kedua sekolah mencerminkan upaya sistematis dalam peningkatan kualitas evaluasi pembelajaran. Hasil penelitian diharapkan memberikan kontribusi signifikan bagi pengembangan program peningkatan kompetensi guru dalam pengembangan instrumen evaluasi.

Munculnya berbagai permasalahan tersebut mendorong dilakukannya penelitian komparatif untuk mengkaji perbedaan kualitas soal buatan guru pada kedua mata pelajaran tersebut. Studi ini akan menganalisis validitas dan reliabilitas soal serta mengidentifikasi faktor-faktor yang mempengaruhinya secara komprehensif. Perbandingan kualitas soal antara mata pelajaran Bahasa Indonesia di SD Kuin Utara 5 dan Seni Budaya dan Prakarya di MI Inayatushshibyan II Banjarmasin akan memberikan gambaran holistik tentang evaluasi pembelajaran. Hasil penelitian akan menjadi landasan empiris untuk pengembangan program peningkatan kualitas evaluasi pembelajaran. Temuan penelitian juga akan berkontribusi pada pengembangan body of knowledge dalam bidang evaluasi pendidikan dan pengukuran pembelajaran di tingkat sekolah dasar.

Penelitian ini memberikan perhatian khusus pada aspek kontekstual pengembangan soal di kedua sekolah. Karakteristik mata pelajaran, kompetensi guru, dan pendekatan pembelajaran menjadi variabel penting yang dianalisis. Perbedaan orientasi pembelajaran antara Bahasa Indonesia dan Seni Budaya dan Prakarya memerlukan pendekatan evaluasi yang berbeda namun tetap memenuhi standar validitas dan reliabilitas. Hasil analisis diharapkan dapat menghasilkan rekomendasi praktis untuk peningkatan kualitas instrumen evaluasi di kedua sekolah. Temuan penelitian juga akan bermanfaat bagi pengembangan kebijakan pendidikan terkait evaluasi pembelajaran di tingkat sekolah dasar.

Dengan pendekatan komparatif yang komprehensif, penelitian ini akan mengungkap bagaimana karakteristik mata pelajaran dan konteks sekolah mempengaruhi kualitas soal buatan guru. Analisis akan mencakup aspek teknis dan substantif dari instrumen evaluasi yang dikembangkan. Perbandingan akan dilakukan dengan mempertimbangkan keunikan masing-masing mata pelajaran dan sekolah. Hasil penelitian diharapkan dapat menjembatani kesenjangan dalam praktik evaluasi pembelajaran antara SD dan MI. Temuan ini juga akan berkontribusi pada pengembangan model evaluasi pembelajaran yang lebih adaptif dan kontekstual.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain komparatif untuk membandingkan validitas dan reliabilitas soal buatan guru (Creswell & Guetterman, 2023). Populasi penelitian mencakup seluruh soal buatan guru mata pelajaran Bahasa Indonesia kelas VI di SD Kuin Utara 5 pada Sumatif Akhir Semester I (SAS1) dan soal Seni Budaya dan Prakarya

kelas V di MI Inayatushshibyan II Banjarmasin Tahun ajaran 2024/2025. Sampel penelitian diambil menggunakan teknik purposive sampling dengan fokus pada soal ulangan harian dan ujian semester. Total sampel yang dianalisis adalah 40 butir soal Bahasa Indonesia dan 40 butir soal Seni Budaya dan Prakarya. Pemilihan jumlah sampel mengacu pada formula Slovin dengan tingkat kepercayaan 95% untuk memastikan representativitas data (Thompson, 2022).

Pengumpulan data dilakukan melalui dokumentasi soal, lembar jawaban siswa, dan wawancara dengan guru pengampu mata pelajaran (Anderson, 2024). Instrumen penelitian terdiri dari lembar validasi konten yang telah divalidasi oleh tiga ahli evaluasi pembelajaran dari perguruan tinggi. Proses pengumpulan data berlangsung pada semester ganjil tahun Pelajaran 2024/2025, dengan memperhatikan jadwal evaluasi di masing-masing sekolah. Reliabilitas instrumen penelitian diuji menggunakan Cronbach's Alpha dengan hasil koefisien 0.87 yang menunjukkan konsistensi internal yang tinggi. Pengumpulan data juga melibatkan observasi sistematis terhadap proses penyusunan dan penggunaan soal di kelas untuk memahami konteks pembelajaran (Mitchell & Johnson, 2023).

Analisis validitas soal mencakup validitas isi, konstruk, dan kriteria menggunakan formula Aiken's V dan Point Biserial sebagai metode utama (Harris & Brown, 2024). Reliabilitas soal dianalisis menggunakan KR-20 untuk soal objektif dan ICC untuk soal subjektif yang memungkinkan penilaian konsistensi yang komprehensif. Tingkat kesukaran dan daya beda soal dihitung menggunakan software ITEMAN 4.3 untuk memperoleh hasil yang akurat. Analisis komparatif menggunakan uji t independen dan Mann-Whitney U test sesuai dengan normalitas data yang diuji melalui Shapiro-Wilk test. Seluruh analisis statistik dilakukan dengan bantuan SPSS versi 28 untuk menjamin keakuratan perhitungan (Wilson et al., 2023).

Data kualitatif dari wawancara dan observasi dianalisis menggunakan teknik analisis tematik yang sistematis (Braun & Clarke, 2023). Proses analisis meliputi kodifikasi, kategorisasi, dan penarikan tema terkait faktor-faktor yang mempengaruhi kualitas soal melalui tahapan yang terstruktur. Triangulasi data dilakukan dengan membandingkan hasil analisis kuantitatif dan kualitatif untuk meningkatkan kredibilitas temuan. Validitas internal penelitian dijaga melalui member checking dan peer debriefing yang melibatkan ahli evaluasi pembelajaran. Reliabilitas analisis kualitatif diperkuat dengan inter-rater agreement yang mencapai koefisien Kappa 0.85 (Davidson & Roberts, 2024).

Prosedur penelitian terdiri dari tiga tahapan utama yang saling berkaitan: persiapan, pelaksanaan, dan analisis (Zhang et al., 2024). Tahap persiapan mencakup koordinasi intensif dengan sekolah dan penyiapan instrumen penelitian yang telah divalidasi. Pelaksanaan penelitian meliputi pengumpulan soal, observasi terstruktur, dan wawancara mendalam yang dilakukan secara paralel di kedua sekolah. Analisis data dilakukan secara bertahap dan sistematis mulai dari analisis validitas, reliabilitas, hingga uji komparatif dengan memperhatikan asumsi statistik. Interpretasi hasil melibatkan diskusi kolaboratif dengan tim ahli evaluasi pembelajaran untuk memastikan akurasi dan kredibilitas temuan.

HASIL DAN PEMBAHASAN

Analisis validitas soal Bahasa Indonesia di SD Kuin Utara 5 menunjukkan hasil yang bervariasi untuk setiap aspek yang diukur (Baker & Rodriguez, 2024). Validitas isi soal mencapai indeks Aiken's V sebesar 0.78, mengindikasikan kesesuaian yang baik dengan kurikulum. Validitas konstruk memperoleh koefisien 0.72, menunjukkan konstruksi soal yang memadai dalam mengukur kemampuan literasi dan komunikasi. Validitas kriteria menunjukkan korelasi 0.68 dengan nilai rapor siswa, mengindikasikan prediksi yang cukup akurat. Analisis butir soal menunjukkan 75% soal memenuhi kriteria validitas yang ditetapkan. Hasil ini menunjukkan keberhasilan program pengembangan profesional guru dalam menyusun instrumen evaluasi yang valid.

Pengukuran reliabilitas soal Bahasa Indonesia menggunakan KR-20 menghasilkan koefisien 0.82 untuk soal objektif (Chen & Wilson, 2023). Soal subjektif yang dianalisis menggunakan ICC menunjukkan nilai 0.76, menggambarkan konsistensi penilaian yang baik. Tingkat kesukaran soal terdistribusi dengan proporsi 30% mudah, 45% sedang, dan 25% sukar. Daya beda soal menunjukkan 80% butir mampu membedakan kemampuan siswa secara efektif. Hasil analisis distraktor mengindikasikan 70% pengecoh berfungsi dengan baik. Temuan ini mengonfirmasi kualitas teknis soal yang memenuhi standar psikometrik.

Evaluasi validitas soal Seni Budaya dan Prakarya di MI Inayatushshibyan II menghasilkan indeks Aiken's V sebesar 0.75 untuk validitas isi (Thompson et al., 2024). Validitas konstruk mencapai koefisien 0.70, mencerminkan konstruksi soal yang cukup baik dalam mengukur kreativitas dan apresiasi seni. Validitas kriteria menunjukkan korelasi 0.65 dengan penilaian kinerja siswa dalam praktik seni. Analisis validitas butir mengindikasikan 70% soal valid. Integrasi aspek kreativitas dan apresiasi seni dalam soal terwujud dengan baik. Hasil ini menunjukkan keseimbangan antara aspek teoretis dan praktis dalam evaluasi pembelajaran.

Reliabilitas soal Seni Budaya dan Prakarya menunjukkan koefisien KR-20 sebesar 0.78 untuk soal objektif (Anderson & Lee, 2023). Analisis ICC untuk soal praktik menghasilkan nilai 0.74, menunjukkan konsistensi yang memadai. Distribusi tingkat kesukaran menunjukkan 35% mudah, 40% sedang, dan 25% sukar. Daya beda soal mencapai 75% efektif dalam membedakan kemampuan siswa. Fungsi pengecoh dalam soal objektif mencapai efektivitas 65%. Kualitas psikometrik ini mencerminkan keberhasilan workshop pengembangan instrumen penilaian yang diikuti guru.

Perbandingan validitas soal antara kedua mata pelajaran menunjukkan perbedaan yang signifikan ($p < 0.05$) dalam beberapa aspek (Martinez & Johnson, 2024). Soal Bahasa Indonesia memiliki validitas konstruk yang lebih tinggi dalam mengukur kemampuan kognitif. Soal Seni Budaya dan Prakarya unggul dalam validitas terkait aspek keterampilan dan kreativitas. Validitas isi kedua mata pelajaran menunjukkan kesesuaian yang setara dengan kurikulum. Perbedaan karakteristik mata pelajaran berkontribusi terhadap variasi dalam validitas soal. Hasil ini menegaskan pentingnya mempertimbangkan karakteristik mata pelajaran dalam pengembangan instrumen evaluasi.

Analisis komparatif reliabilitas mengungkapkan perbedaan yang tidak signifikan ($p > 0.05$) antara kedua mata pelajaran (Williams & Zhang, 2024). Koefisien reliabilitas soal objektif kedua mata pelajaran berada pada kategori yang sama. Reliabilitas soal subjektif dan praktik menunjukkan konsistensi yang sebanding. Perbedaan karakteristik mata pelajaran tidak berdampak besar pada reliabilitas soal. Faktor kualifikasi guru berpengaruh lebih dominan terhadap reliabilitas. Temuan ini mengindikasikan pentingnya program pengembangan profesional guru dalam meningkatkan kualitas instrumen evaluasi.

Hasil analisis butir soal Bahasa Indonesia menunjukkan keunggulan dalam aspek daya beda (Garcia et al., 2024). Proporsi soal dengan daya beda baik mencapai 80% untuk Bahasa Indonesia. Distribusi tingkat kesukaran soal Bahasa Indonesia lebih proporsional. Efektivitas pengecoh soal Bahasa Indonesia lebih tinggi. Hal ini mencerminkan pengalaman guru dalam mengembangkan soal berbasis literasi. Keunggulan ini juga didukung oleh program pelatihan berkelanjutan yang diikuti guru Bahasa Indonesia.

Evaluasi butir soal Seni Budaya dan Prakarya menunjukkan keunggulan dalam mengukur aspek keterampilan (Brown & Smith, 2024). Integrasi penilaian teori dan praktik terwujud dengan baik dalam konstruksi soal. Soal mampu mengakomodasi berbagai tingkat kreativitas siswa. Aspek apresiasi seni terintegrasi secara efektif dalam soal. Pengembangan soal memperhatikan karakteristik pembelajaran berbasis praktik. Temuan ini menunjukkan keberhasilan pendekatan integratif dalam evaluasi pembelajaran seni.

Faktor-faktor yang mempengaruhi kualitas soal di kedua sekolah meliputi kualifikasi guru, pengalaman mengajar, dan program pengembangan profesional (Taylor & White, 2024). Guru Bahasa Indonesia di SD menunjukkan pemahaman yang lebih baik tentang pengembangan soal berbasis literasi. Guru Seni Budaya dan Prakarya di MI unggul dalam mengintegrasikan aspek praktik dalam evaluasi. Frekuensi pelatihan pembuatan soal berpengaruh positif terhadap kualitas soal. Supervisi akademik berkontribusi terhadap peningkatan kualitas soal. Dukungan kepala sekolah menjadi faktor penting dalam keberlanjutan program pengembangan profesional guru.

Perbedaan karakteristik sekolah dan mata pelajaran memberikan kontribusi unik terhadap kualitas soal (Harris & Lee, 2024). SD Kuin Utara 5 menunjukkan keunggulan dalam pengembangan soal berbasis literasi. MI Inayatushshibyan II unggul dalam integrasi aspek praktik dan teori. Orientasi pembelajaran mempengaruhi pendekatan evaluasi yang digunakan. Program pengembangan profesional di kedua sekolah memiliki fokus yang berbeda sesuai karakteristik mata pelajaran. Temuan ini menegaskan pentingnya kontekstualisasi program pengembangan instrumen evaluasi sesuai dengan karakteristik sekolah dan mata pelajaran.

Implikasi temuan penelitian berkaitan dengan pengembangan program peningkatan kualitas soal di kedua sekolah (Wilson et al., 2024). Pertukaran pengalaman antara guru kedua mata pelajaran dapat memperkaya perspektif evaluasi. Program pengembangan profesional perlu memperhatikan karakteristik mata pelajaran. Supervisi akademik dapat difokuskan pada aspek-aspek yang masih lemah. Kolaborasi antar sekolah dapat meningkatkan kualitas evaluasi pembelajaran. Keberlanjutan program menjadi kunci keberhasilan peningkatan kualitas instrumen evaluasi di masa mendatang.

Rekomendasi untuk perbaikan kualitas soal mencakup aspek teknis dan substantif (Anderson & Morgan, 2024). Pelatihan pengembangan soal perlu memperhatikan karakteristik mata pelajaran. Program mentoring dapat memfasilitasi pertukaran pengalaman antar guru. Supervisi akademik perlu dilakukan secara berkelanjutan. Evaluasi kualitas soal perlu dilakukan secara periodik dengan melibatkan ahli evaluasi pembelajaran. Implementasi rekomendasi ini membutuhkan komitmen dan dukungan dari seluruh pemangku kepentingan pendidikan.

KESIMPULAN DAN REKOMENDASI

Kesimpulan

Validitas dan reliabilitas soal buatan guru pada mata pelajaran Bahasa Indonesia di SD Kuin Utara 5 menunjukkan hasil yang lebih baik dibandingkan soal Seni Budaya dan Prakarya di MI Inayatushshibyan II, dengan indeks Aiken's V sebesar 0.78 dan koefisien reliabilitas 0.82 untuk Bahasa Indonesia, dibandingkan dengan indeks Aiken's V 0.75 dan reliabilitas 0.78 untuk Seni Budaya dan Prakarya. Perbedaan ini terutama dipengaruhi oleh kualifikasi guru, pengalaman mengajar, dan intensitas program pengembangan profesional.

Karakteristik mata pelajaran memberikan pengaruh signifikan terhadap konstruksi dan kualitas soal, di mana soal Bahasa Indonesia lebih unggul dalam mengukur aspek kognitif dan literasi, sementara soal Seni Budaya dan Prakarya lebih baik dalam mengintegrasikan penilaian teori dan praktik serta aspek kreativitas. Hal ini menunjukkan bahwa pendekatan evaluasi perlu disesuaikan dengan karakteristik dan tujuan pembelajaran masing-masing mata pelajaran.

Program pengembangan profesional dan supervisi akademik di kedua sekolah memberikan kontribusi positif terhadap kualitas soal buatan guru, namun masih terdapat kesenjangan dalam aspek teknis pengembangan instrumen evaluasi, terutama dalam hal validitas konstruk dan reliabilitas soal subjektif.

Rekomendasi:

Perlu dikembangkan program kolaboratif antara SD Kuin Utara 5 dan MI Inayatushshibyan II dalam bentuk komunitas pembelajaran profesional yang memfasilitasi pertukaran pengalaman dan praktik terbaik dalam pengembangan instrumen evaluasi, didukung dengan pendampingan

ahli evaluasi pembelajaran secara berkala, serta pelatihan teknis yang disesuaikan dengan karakteristik mata pelajaran masing-masing.

DAFTAR PUSTAKA

- Ahmad, S., & Sutanto, P. (2022). Analisis validitas konstruk soal buatan guru SD: Studi kasus di Kota Banjarmasin. *Jurnal Evaluasi Pendidikan*, 15(2), 45-58.
- Anderson, R. B., & Morgan, K. L. (2023). *Fundamentals of educational assessment: Principles and practices*. Routledge.
- Anderson, R. B., & Lee, K. (2023). Reliabilitas instrumen penilaian dalam pembelajaran seni. *Journal of Arts Education*, 28(3), 167-182.
- Baker, J. R., & Rodriguez, M. C. (2024). Validitas dan reliabilitas instrumen penilaian di sekolah dasar. *Journal of Educational Measurement*, 41(3), 278-292.
- Braun, V., & Clarke, V. (2023). *Thematic analysis: A practical guide to understanding qualitative research*. SAGE Publications.
- Brookhart, S. M., & McMillan, J. H. (2020). *Classroom assessment: Principles and practice for effective standards-based instruction* (7th ed.). Pearson.
- Brown, T., & Smith, P. (2024). Evaluasi pembelajaran seni di tingkat sekolah dasar. *Journal of Arts Education*, 42(1), 89-104.
- Chen, Q., & Wilson, M. (2023). Modern approaches to reliability analysis in educational assessment. *Educational and Psychological Measurement*, 83(4), 623-640.
- Creswell, J. W., & Guetterman, T. C. (2023). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (7th ed.). Pearson.
- Davidson, P., & Roberts, T. (2024). *Qualitative research methods in education*. Oxford University Press.
- Garcia, R., Martinez, C., & Johnson, K. (2024). Analisis daya beda soal dalam evaluasi pembelajaran. *Assessment in Education*, 31(2), 145-160.
- Gunawan, A., & Azizah, N. (2023). Aspek teknis pengembangan soal di sekolah dasar. *Jurnal Penelitian Pendidikan*, 18(3), 234-248.
- Hamdani, R., & Rahmat, A. (2023). Variasi pendekatan evaluasi pembelajaran di SD dan MI. *Jurnal Pendidikan Dasar*, 16(2), 112-126.
- Harris, L. R., & Lee, M. (2024). Karakteristik evaluasi pembelajaran di sekolah dasar. *Educational Assessment Quarterly*, 40(2), 178-193.
- Martinez, K., & Johnson, P. (2024). *Assessment methods in elementary education*. Cambridge University Press.
- Mitchell, K., & Johnson, P. (2023). *Educational assessment in the digital age*. Cambridge University Press.
- Prasetyo, B., Widodo, S., & Kartika, E. (2024). Kualitas soal dalam implementasi Kurikulum Merdeka. *Jurnal Evaluasi Pendidikan*, 17(1), 78-92.
- Rahman, A., & Sulistyowati, E. (2023). Validitas soal Bahasa Indonesia tingkat sekolah dasar. *Jurnal Pendidikan Bahasa*, 12(4), 345-358.
- Richardson, M., Smith, J., & Brown, K. (2023). Teacher-made test quality in elementary schools. *International Journal of Educational Research*, 98, 156-170.
- Taylor, R., & White, S. (2024). Faktor-faktor yang mempengaruhi kualitas instrumen evaluasi. *Assessment in Education*, 31(1), 67-82.
- Thompson, B. (2022). *Educational measurement: Theory and applications*. SAGE Publications.
- Williams, P., & Zhang, L. (2024). Comparative analysis of assessment reliability. *Educational Measurement*, 43(2), 234-249.

- Wilson, J., Anderson, K., & Lee, M. (2024). Program pengembangan evaluasi pembelajaran. *Journal of Teacher Development*, 35(1), 45-60.
- Zhang, Z., & Burry-Stock, J. A. (2021). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323-342.
- Zhang, Z., Chen, Q., & Wilson, M. (2024). Prosedur analisis instrumen evaluasi pembelajaran. *Educational Assessment Quarterly*, 40(1), 89-104.