

ANALISIS BUTIR SOAL ULANGAN AKHIR SEMESTER GANJIL DALAM MATA PELAJARAN PANCASILA KURIKULUM MERDEKA KELAS IV SD/MI

Moch. Farich Alfani

Universitas Islam Negeri Sunan Kalijaga Yogyakarta

Corresponding Author: farichalfani25@gmail.com

Abstract

Without the stage of analyzing item questions, the risk of irrelevant, inconsistent, or ineffective questions in measuring students' understanding can increase. Therefore, the stage of analyzing item questions is an important step that needs to be taken before the questions are tested to ensure good quality educational assessment in Elementary Schools/Madrasah Ibtidaiyah. The purpose of this study is to determine the output of the quality of item questions and the implementation of tests on the semester final exam questions. This research uses a quantitative-descriptive approach. The sampling methods used are random, stratified, and proportional sampling. Analysis is conducted using the ITEMAN 3.00 program. The results of the quantitative analysis show the following data: item questions are categorized as follows: not good (3%), fair (33%), and good (71%). The difficulty level of item questions obtained are categorized as too difficult (28.6%), moderate (60.1%), and too easy (21%). The effectiveness of distractors in the questions is effective (100%), not effective (0%). The reliability of the item questions globally reaches 0.386, which means that the questions have high consistency.

Keywords: *Item Analysis; assessment; multiple choices*

Abstrak

Tanpa tahapan analisis butir-butir soal, risiko terjadinya pertanyaan yang tidak relevan, tidak konsisten, atau tidak efektif dalam mengukur pemahaman siswa dapat meningkat. Sehingga, tahapan analisis butir-butir soal adalah langkah yang penting dan perlu dilakukan sebelum soal diujikan untuk memastikan kualitas evaluasi pendidikan yang baik khususnya di tingkat pendidikan Sekolah Dasar/Madrasah Ibtidaiyah. Tujuan penelitian ini adalah untuk mengetahui dari output kualitas butir soal dan pelaksanaan tes pada soal ujian akhir semester tersebut. Penelitian ini menggunakan pendekatan kuantitatif-deskriptif. Metode pengambilan sampel yang digunakan adalah random, stratified, dan proporsional sampling. Analisis dengan menggunakan program ITEMAN 3.00. Hasil analisis kuantitatif menunjukkan data sebagai berikut; soal kurang baik soal kurang baik (3%), cukup baik (33%), dan baik (71%). Tingkat kesulitan butir soal yang diperoleh dengan hasil, terlalu sulit (28,6%), sedang (60,1%), dan terlalu mudah (21%). Kefektifan distractor pada soal adalah efektif (100%), tidak efektif (0%). Realibilitas dalam butir soal secara global mencapai 0,386 yang artinya soal-soal tersebut memiliki konsistensi yang tinggi.

Kata Kunci: Analisis butir soal; asesmen; pilihan ganda

PENDAHULUAN

Dalam lingkungan Pendidikan, sebuah penilaian kerap disamaratakan pendefinisianya yakni dengan istilah kegiatan mengevaluasi. Evaluasi sendiri merupakan sektor dari wilayah proses dan secara global tidak dapat dipisahkan dari adanya kegiatan pembelajaran (Muhammad Affandi,

2023). Kegiatan evaluasi di lingkungan pemerintah telah diatur dalam Undang-undang Republik Indonesia No. 20 Tahun 2003 tentang sistem Pendidikan nasional bab XVI pasal 58 ayat 1 yang berbunyi bahwa evaluasi hasil belajar yang dilakukan peserta didik dilakukan oleh guru atau pendidik dengan maksud untuk memantau proses, kemajuan, dan perbaikan dari adanya hasil belajar peserta didik secara berkesinambungan. Oleh karena itu, kegiatan evaluasi dari hasil belajar bertujuan untuk menilai pencapaian kompetensi serta memperbaiki proses pembelajaran yang berpedoman pada penyusunan laporan kemajuan hasil belajar siswa. Salah satu komponen alat yang dapat diaplikasikan sebagai sarana kegiatan untuk menilai hasil belajar peserta didik yaitu tes. Tes merupakan himpunan pertanyaan yang harus dijawab, harus ditanggapi, atau tugas yang harus dilaksanakan oleh orang yang dites. Tes digunakan untuk mengukur sejauh mana seorang siswa telah menguasai pelajaran yang disampaikan terutama meliputi aspek pengetahuan dan ketrampilan (Resti et al., 2020).

Adapun definisi tes menurut Linn dan Granold (Slack & Priestley, 2023) adalah *“test is an instrument of systematic prosedur for measuring a sample of behafioring ny posing a set of questions in a uniform manner. Because a test a form of assessment, test a also answer the questions how well does the individual perform either in comparasion with other or in comparasion rwith a domain of performance task”*. Tes pada umumnya dimaksudkan untuk mengukur aspek-aspek perilaku manusia, seperti aspek pengetahuan (Kognitif), aspek sikap (afektif), maupun aspek ketrampilan (psikomotor). Hal yang hendak diukur adalah tingkat penguasaan peserta didik terhadap bahan pelajaran yang telah diajarkan (Sumarna, 2004). Dalam uraian ini, perlu dibedakan antara prestasi belajar dan hasil belajar. Prestasi belajar hanya mengukur dua aspek yaitu koginitif dan aspek psikomotorik. Hasil belajar meliputi aspek pembentukan watak seseorang peserta didik, dengan demikian mengukur tiga aspek utama hasil Pendidikan, yaitu aspek kognitif, psikomotor, dan afektif. Analisis dari butir-butir soal yang akan diujikan merupakan suatu kegiatan yang harus dilakukan guru dalam meningkatkan kualitas output soal yang telah ditulis.

Tujuan menganalisis butir – butir soal yaitu untuk meningkatkan kualitas setiap butir tes yang diujikan dan untuk mengetahui mengenai informasi diagnostic peserta didik (Ahmad & Sukiman, 2019). Ciri dari soal yang berkualitas yaitu soal-soal yang dapat menyajikan informasi setepat-tepatnya sehingga dapat diketahui peserta didik antara mereka yang telah menguasai materi dengan yang masih belum menguasai materi yang telah dibawakan oleh guru (Supriadi et al., 2023). Pertanyaan soal-soal ulangan akhir semester menjadi sangat krusial ketika peneliti mendapatkan beberapa pernyataan yang membahas berhwa soal-soal ujian akhir semester termasuk dikategorikan kedalam rancangan tes sumatif. Dan fungsi tes sumatif sendiri diantaranya untuk mengetahui sampai sejauh mana kompetensi pencapaian peserta didik dalam mata pelajaran tertentu (Resti et al., 2020). Adapun hasil tes sumatif peserta didik, selanjutnya masuk kedalam perbandingan dengan penetapan tujuan pembelajaran yang telah dirancang dan ditetapkan sehingga mencapai pada kriteria ketuntasan minimal atau sering dikenal dengan istilah KKM (Wiguna, 2022). Oleh karenanya, nilai dari sebuah tes sumatif dapat dijadikan sebagai catatan tidak maupun adanya kemajuan yang dirasakan peserta didik dan menjadi penentu seorang peserta didik dapat atau tidaknya menerima program berikutnya.

Lembaga sekolah yang notabene masih jauh dari pengawasan, masih banyaknya kasus-kasus butir soal yang akan digunakan dalam pelaksanaan ujian akhir semester yang tidak diawali dengan adanya tahap pengujian sebelumnya, diantaranya tanpa melalui tahapan analisis butir – butir soal yang telah disusun (Nurhabibah et al., 2021). Faktanya, Dalam kegiatan wawancara yang telah dilakukan pada tanggal 15 April 2023 seluruh peserta didik di lingkungan MIN 4 Brebes telah

menyelesaikan kegiatan ujian akhir semester ganjil dengan materi soal yang diujikan tanpa melalui tahapan dari analisis butir soal sendiri. Fakta dari data hasil wawancara tersebut, peneliti menginginkan adanya pengkajian dalam melakukan penelitian terkait analisis butir soal Ulangan Akhir Semester Ganjil Dalam Mata Pelajaran Pancasila Kurikulum Merdeka Kelas IV SD/MI”.

METODE

Dalam penelitian ini menggunakan metode deskripsi kuantitatif dengan analisis Iteman yang merupakan program untuk menganalisis butir-butir soal secara klasik. Program ini termasuk satu paket dalam Micro CATN yang telah dikembangkan oleh Assesment System Corporation mulai dari tahun 1982 dan mengalami tahap revisian pada tahun 1984, 1986 sampai dengan versi 3.50. Program ini dapat digunakan untuk; (1) menganalisis data file jawaban butir soal yang dihasilkan melalui manual entry data atau mesin scanner; (2) menskor serta menganalisis data soal pilihan ganda dan skala likert; (3) menganalisis sebuah tes yang terdiri dari 10 skala (subtes) dan memberikan informasi tentang validitas setiap butir (daya pembeda, tingkat kesukaran, proporsi jawaban pada setiap option, reliabilitas, standar error of measurement, mean, variance, standar deviasi skew, kurtosis, skor minimum, skor maksimum, skor median, dan frekuensi distributor skor (Sudaryono et al., 2013).

Populasi dalam penelitian ini yaitu sebanyak 20 peserta didik yang mengikuti ujian akhir sekolah semester ganjil mata pelajaran pendidikan kewarganegaraan kelas IV di MIN 4 Brebes tahun ajaran 2021/2022. Pengambilan sampel dilakukan dengan menggunakan teknik sampling dalam memperoleh data yang representatif dalam penelitian (Sugeng, 2022; Yusuf, 2016). Pengambilan sampel dalam penelitian ini memakai teknik sampling jenuh, sehingga penentuan sampel menggunakan semua anggota populasi sebagai sampel. Penulis memilih menggunakan teknik sampling jenuh, karena penulis ingin membuat generalisasi dengan kesalahan yang sangat kecil. Jadi, banyaknya sampel yang digunakan dalam penelitian ini cukup dengan memakai 20 siswa yang mengikuti tes. Teknik pengumpulan data merupakan cara yang penulis gunakan dalam memperoleh semua data yang dibutuhkan dalam penelitian. Teknik pengumpulan dalam penelitian ini dilakukan dengan memakai metode wawancara dan juga dokumentasi (Creswell, 2014; Creswell & Creswell, 2018).

Metode wawancara dalam penelitian ini digunakan sebagai literatur studi pendahuluan untuk menemukan titik permasalahan yang akan diteliti dan proses pelaksanaan tes pada ujian akhir semester. Adapun teknik wawancara yang digunakan dalam penelitian ini yaitu dengan konsep wawancara tak terstruktur. Dokumentasi dalam penelitian ini, dilakukuan dalam memperoleh kisi-kisi penulisan soal, butir soal ujian akhir semester, lembar jawab siswa, dan daftar nama siswa pada ujian akhir semester ganjil mata pelajaran pendidikan kewarganegaraan kelas IV di MIN 4 Brebes tahun ajaran 2021/2022. Peneliti menggunakan alat bantu untuk mempermudah dalam memperoleh data.

Alat pengumpulan data yang digunakan dalam penelitian ini yaitu: (1) Daftar pertanyaan wawancara yang didalamnya berisikan pertanyaan yang diajukan kepada tim penyusun soal dan juga guru kelas IV di MIN 4 Brebes; (2) Alat recording, sebagai alat bantu dalam pengumpulan data yang berupa suara, alat perekam digunakan agar data yang diperoleh tersajikan secara jelas serta (3) Daftar Chekklist yang digunakan sebagai alat dalam mengumpulkan data yang berupa dokumentasi untuk mengetahui daftar dokumen yang digutuhkan dalam penelitian (Sudaryono et al., 2013). Penelitian ini dengan menggunakan teknik analisis data, yaitu analisis secara kualitatif

dan kuantitatif. Analisis kualitatif dalam penelitian ini yakni ditinjau dengan melihat aspek analisis materi, konstruksi, bahasa, distribusi jenjang ranah kognitif, dan wawancara pelaksanaan tes. Analisis kuantitatif didalamnya mencakup pengukuran validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektifitas pengecoh. Analisis materi, konstruksi, serta bahasa dalam penelitian ini memiliki tujuan untuk mengetahui validitas isi pada soal pilihan ganda pada ujian akhir semester ganjil mata pelajaran pendidikan pancasila kelas IV MIN 4 Brebes.

Terdapat empat tahapan dalam penghitungannya, yaitu dengan menelaah butir soal menggunakan teknik panel, menentukan spesifikasi domain skala empat-point, menentukan model kesepakatan interrater, dan menghitung validitas isi. Analisis distribusi jenjang ranah kognitif merupakan sebuah kegiatan analisis yang bertujuan untuk mengetahui jenjang ranah kognitif pada soal (Muhammad Affandi, 2023). Analisis distribusi pada jenjang ranah kognitif merupakan kegiatan analisis yang bertujuan untuk mengetahui jenjang ranah kognitif pada butir-butir soal yang diujikan (Wahyuni & Mardapi, 2015).

Terdapat enam ranah pengetahuan kognitif taksonomi Bloom yang direvisi, yaitu mengetahui, memahami, menerapkan, menganalisis, menilai, dan menciptakan (Krathwohl, 2002). Ranah kognitif yang cocok diterapkan di sekolah dasar yaitu pengetahuan, pemahaman, dan aplikasi. Selain ranah kognitif tersebut dapat dilatihkan (Pratyca et al., 2023). Oleh karena itu, analisis distribusi jenjang ranah kognitif dalam penelitian ini dilakukan dengan melakukan cara mencocokkan butir soal yang memiliki 3 kriteria dalam ranah kognitif yang telah diterapkan pada jenjang sekolah dasar. Analisis secara kuantitatif dilakukan untuk menganalisis data empirik dari butir soal yang bersangkutan. Data empirik dengan diperoleh dari butir soal yang telah diujikan. Analisis dalam penelitian ini dengan menggunakan pendekatan klasik (Resti et al., 2020). Dengan berdasarkan pendekatan klasik aspek-aspek yang dianalisis yaitu aspek validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh. Analisis kuantitatif dalam riset penelitian ini dilaksanakan dengan menggunakan program iteman.

HASIL DAN PEMBAHASAN

Kegiatan menelaah dari butir-butir soal menggunakan teknik panel, menentukan spesifikasi domain pada skala empat-point, menentukan model kesepakatan interrater, dan menghitung validitas isi. Hasil dari analisis materi, konstruksi, dan bahasa menunjukkan bahwa butir-butir soal yang dianalisis memiliki validitas isi berkategori sangat tinggi, konstruksi berkategori sangat tinggi, dan bahasa dengan tingkat pengkategorian sangat tinggi. Jadi, dapat disimpulkan bahwa dalam penyajian soal-soal ujian akhir semester tersebut memiliki validitas isi yang baik, sehingga layak untuk diujikan. Kegiatan analisis pertama adalah dengan melakukan analisis secara kualitatif terhadap 20 butir soal yang hasilnya dapat dilihat sebagai berikut; a) sesuai segi materi 86,58 %, b.) Sesuai segi konstruksional 64%, c.) Sesuai segi bahasa: 74%. Secara keseluruhan butir soal dari segi materi sudah baik artinya pembuatan soal sudah mengacu pada kompetensi dasar yang telah diturunkan menjadi kisi-kisi dalam butir soal. Hal ini sesuai dengan langkah penulisan butir soal bahwa pembuatan soal dengan diawali pendahuluan yakni menganalisis kompetensi dasar. Setelah mencapai tahap analisis kualitatif butir soal kemudian dilakukan penganalisisan secara kualitatif, hasilnya dipaparkan sebagai berikut:

Validitas

Hasil dari perhitungan validitas dengan harga kritik 4% adalah 12 soal (68,57%) valid dan 8 soal (31,43%) tidak valid, secara rinci digambarkan seperti pada table berikut.

Tabel 1 Hasil Perhitungan Validitas Soal

No	Validitas	No butir soal	Jumlah	%
1	≥ 0.3291 (Valid)	3, 5, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19	14	88,58%
2	< 0.3291 (tidak valid)	1, 2, 4, 6, 11, 19, 18, 20	8	21,42%
			35	100%

Dengan adanya penyajian data dari table berikut secara umum soal memiliki validitas yang baik, 88,58% dengan demikian guru harus melakukan revisi pada setiap butir soal dengan jumlah peserta tes sebanyak 20 peserta didik, maka dapat diketahui bahwa hanya satu soal yang hasilnya tidak memiliki validitas yaitu pada soal bernomor 4. Ada 9 soal yang diterima dan 7 soal yang perlu untuk direvisi. Hal ini sangat penting karena validitas merupakan syarat krusial untuk terbentuknya soal yang akurat. Karena begitu pentingnya maka dikatakan oleh Drost (2012) bahwa validitas merupakan “validity is concerned with the meaningfulness of research components”.

Realibilitas

Reabilitas merupakan hasil rata-rata perhitungan reliabilitas 20 peserta didik adalah 0,386 masuk pada tingkat kategori tinggi, dan bisa dilihat pada table berikut:

Tabel 2 Hasil Reliabilitas Soal Pilihan Ganda

Score	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First- Last	S-B Odd- Even
Scored items	-0,659	1,590	-0,359	-0,217	-0,386	-1,120	-0,553	-1,255

Tabel 3 Reliabilitas Seluruh Peserta Didik Kelas IV MIN 4 Brebes

Item ID	P / Item Mean	R	Flag(s)
item03	0,750	-0,285	K, LR
item05	0,050	-0,174	K, LR
item07	0,650	-0,403	K, LR
item08	0,750	0,000	K
item09	0,150	-0,379	K, LR
item10	0,050	-0,174	K, LR
item13	0,100	-0,235	K, LR
item14	0,550	0,058	K
item15	0,150	-0,276	K, LR
item16	0,050	-0,355	K, LR
item17	0,100	-0,235	K, LR
item18	0,400	-0,209	K, LR
item19	0,250	-0,241	K, LR
item20	0,500	-0,133	K, LR

Dilihat dengan adanya table 2 diatas sehingga dapat disimpulkan secara global butir soal-soal pada soal ujian akhir semester sudah memiliki reliabilitas yang cukup tinggi dengan rata-rata

0,287. Oleh karena itu, butir-butir soal sudah layak digunakan dalam mengukur kemampuan peserta didik.

Daya Beda

Perhitungan daya beda dapat menghasilkan butir – butir soal yang jelek berjumlah 1 (3%), cukup baik 7 (33%) dan baik 14 (71%), baik sekali/sangat jelek 0(0%) secara rinci disajikan dalam table berikut:

Tabel 4 Hasil Daya Beda

Klasifikasi	Jml	%	No. Soal
<i>Poor/jelek (D<0,20)</i>	1	3%	8
<i>Satisfactory/Cukup (0,20<D<0,40)</i>	7	33%	1,2,4,6,11,19,18,20
<i>Good/Baik (0,40<D<0,70)</i>	14	71%	3, 5, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19
<i>Exceletn/Baik Sekali (D>0,70) Bertanda Negatif (-)/ Jelek Sekali</i>	-	-	-

Tabel 4 menunjukkan secara umum butir soal memiliki daya beda yang baik, akan tetapi di MIN 4 Brebes perlu mendapatkan perhatian karena daya beda tidak berfungsi dengan baik. Patut diduga disebabkan dengan adanya karakteristik peserta didik, antara satu dengan lainnya yang merupakan keunggulan yang dimiliki.

Tingkat Kesukaran

Dalam perhitungan adanya hasil tingkat kesukaran yang diperoleh data sebagai berikut:

Tabel 5 Tingkat Kesukaran

Option	N	0-20%	20-40%	40-60%	60-80%	80-100%	Color
A	10	0,667	0,000	0,000	0,625	0,500	Maroon
B	1	0,000	0,000	0,000	0,000	0,167	Green
C	4	0,000	0,667	0,000	0,125	0,167	Blue
D	5	0,333	0,333	0,000	0,250	0,167	Olive

Adapun hasil analisis tiap butir – butir soal adalah dapat dilihat pada tabel berikut.

Tabel 6 Tingkat Kesukaran Soal

Klasifikasi	Presentase
<i>Terlalu Sukar (P<0,30) Cukup / Sedang (0,30<P<0,70)</i>	8,6%
<i>Terlalu Mudah (P>0,70)</i>	60,1%
	21%

Pada rincian tabel 5 menjelaskan tingkat kesukaran dalam menunjukkan soal Sebagian besar dalam kategori sedang atau cukup, sehingga sudah layak untuk dapat digunakan. Hal tersebut senada dengan sebuah pernyataan dari Sindi Amelia (2023) dengan memaparkan hasil riset penelitiannya bahwa soal yang dikategorikan baik ialah soal-soal yang tergolong dalam kategori tingkat kesulitan yang tengah atau sedang, yakni tidak masuk kategori mudah maupun juga masuk pada kategori sulit.

Distraktor atau Pengecoh

Butir soal tersebut merupakan hasil dari perhitungan tingkat efektivitas soal pengecoh yang memiliki fungsi dengan jumlah 19 soal (90%) dan tidak memiliki tingkat keberfungsian sejumlah 1 soal (10%), maka secara perincian tingkat efektivitas pengecoh dalam soal disajikan pada tabel berikut:

Tabel 7 Efektivitas Distraktor Soal Kelas IV MIN 4 Brebes

Klasifikasi	Umlah	Presentase	Nomor Soal
<i>Berfungsi</i>	20	100%	1,2,3,4,5,6,7,8,9,10, 11,12,13,14,15,16,17,18,19,20
<i>Tidak Berfungsi</i>	0	0%	

N	P	Total Rpbis	Total Rbis	Alpha w/o
20	0,100	-0,235	-0,402	-0,512

Dari paparan data pada tabel 7 tersebut maka menunjukkan secara umum butir soal belum mempunyai distractor dari fungsi efektivitasnya. Hal tersebut dapat dianalisis dari adanya hasil perhitungan yaitu 87% pilihan jawaban tidak efektif sebagai pengecoh. Dugaan secara sementara adanya factor kesulitan guru dalam mengambil jawaban yang berifat homogen pada level *Higher Order Thinking Skill* (HOTS). Disamping hal tersebut dari adanya hasil penelitian yang memaparkan bahwa selama ini guru mengalami tingkat kesulitan dalam menganalisis butir soal, sehingga kegiatan analisis tidak pernah dapat dilakukan (Anggraeni, 2017). Oleh karena itu, terkait informasi ini dapat dijadikan sebagai pemicu dalam melakukannya analisis butir soal secara berkelanjutan supaya dapat menghasilkan butir-butir soal yang merupakan berkualitas.

SIMPULAN

Dengan melihat adanya hasil riset penelitian dan pembahasan analisis soal ujian akhir semester pada kelas IV di sekolah MIN 4 brebes bahwa hasil penganalisisan dalam analisis kualitatif menunjukkan berbagai pertanyaan yang telah memenuhi standar, namun perlu adanya beberapa tahapan perbaikan dalam hal materi, konstruksi, serta kesesuaian Bahasa/budaya di beberapa lingkup bagian. Sedangkan analisis kuantitatif diperoleh hasil 88,58 % Butir soal yang valid, dan mempunyai tingkat reliabilitas -0,386 tergolong dalam kategori tinggi, kemudian perhitungan dengan gaya pembeda yang diperoleh dengan hasil soal kurang baik (3%), cukup baik (33%), dan baik (71%). Tingkat kesulitan butir soal yang diperoleh dengan hasil, terlalu sulit (28,6%), sedang (60,1%), dan terlalu mudah (21%). Kefektifan distractor pada soal adalah efektif (100%), tidak efektif (0%). Dengan adanya hasil tersebut dapat disimpulkan bahwa butir soal yang diujikan telah memenuhi standar, namun masih diperlukan dalam menyangkut perbaikan materi, konstruksi, dan Bahasa/budaya, sehingga perlu adanya perbaikan dari opsi pilihan jawaban sehingga distractor akan lebih efektif.

DAFTAR PUSTAKA

- Akhwani, Akhwani, dan Dewi Widiana Rahayu. 2021. "Analisis Komponen TPACK Guru SD Sebagai Kerangka Kompetensi Guru Profesional Di Abad 21." *Jurnal Basicedu* 5(4):1918–25. doi: 10.31004/basicedu.v5i4.1119.
- Ahmad, I. F., & Sukiman, S. (2019). Analisis Higher Order Thinking Skills (HOTS) pada Soal Ujian Akhir Siswa Kelas 6 KMI dalam Kelompok Mata Pelajaran Dirasah Islamiyah di Pondok Tazakka Batang. *Jurnal Pendidikan Agama Islam*, XVI(2), Article 2.

- Amelia, S., Widiati, I., & Yadrika, G. (2023). Pengembangan Soal Numerasi untuk Peserta Didik Fase D. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 12(3), Article 3. <https://doi.org/10.24127/ajpm.v12i3.7236>
- Angel Pratyca, Arya Dharma Putra. (n.d.). Analisis Perbedaan Kurikulum 2013 dengan Kurikulum Merdeka. *Jurnal Pendidikan Sains dan Komputer*, 3(1), 58–64. <https://doi.org/10.47709/jpsk.v3i01.1974>
- Anggraeni, L.-. (2017). Peningkatkan Kompetensi Guru Menyusun Butir Soal Bermutu Melalui Program Workshop. *Jurnal Inovasi Pembelajaran Karakter*, 1(2), Article 2. <https://www.i-rpp.com/index.php/jipk/article/view/715>
- Creswell, J. W. (2014). *Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research* (Pearson New International Edition, Fourth Edition). Pearson.
- Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). Sage Publications.
- Drost, J., Van der Does, A. J. W., Antypa, N., Zitman, F. G., Van Dyck, R., & Spinhoven, Ph. (2012). General, Specific and Unique Cognitive Factors Involved in Anxiety and Depressive Disorders. *Cognitive Therapy and Research*, 36(6), 621–633. <https://doi.org/10.1007/s10608-011-9401-z>
- Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*. https://doi.org/10.1207/s15430421tip4104_2
- Muhammad Affandi. (2023). *Evaluasi Pembelajaran Sekolah Dasar (Pertama)*. UNISSULA Press.
- Nurhabibah, S., Ga, P. R., Koroh, T. R., Adoe, T., & Nawa, N. E. A. (2021). Pengaruh Pembelajaran Daring dengan Model Mind Mapping Berbasis Digital Android terhadap Hasil Belajar Mahasiswa PGSD UNDANA. *Haumeni Journal of Education*, 1(2), Article 2. <https://doi.org/10.35508/haumeni.v1i2.5379>
- Pratyca, A., Putra, A. D., Salsabila, A. G. M., Adha, F. I., & Fuadin, A. (2023). Analisis Perbedaan Kurikulum 2013 dengan Kurikulum Merdeka. *Jurnal Pendidikan Sains Dan Komputer*, 3(01), Article 01. <https://doi.org/10.47709/jpsk.v3i01.1974>
- Resti, Y., Zulkarnain, Z., Astuti, A., & Kresnawati, E. S. (2020). Peningkatan Kemampuan Numerasi Melalui Pelatihan Dalam Bentuk Tes Untuk Asesmen Kompetensi Minimum Bagi Guru SDIT Auladi Sebrang Ulu II Palembang. *Applicable Innovation of Engineering and Science Research (AVoER)*, 670–673.
- Slack, H. R., & Priestley, M. (2023). Online learning and assessment during the Covid-19 pandemic: Exploring the impact on undergraduate student well-being. *Assessment & Evaluation in Higher Education*, 48(3), 333–349. <https://doi.org/10.1080/02602938.2022.2076804>
- Sudaryono, S., Rahayu, W., & Margono, G. (2013). *Pengembangan Instrumen Penelitian Pendidikan*. Graha Ilmu.
- Sugeng, B. (2022). *Fundamental Metodologi Penelitian Kuantitatif (Eksplanatif)*. Deepublish.
- Supriadi, F., Wibowo, D. H., & Afgani, C. A. (2023). Analisis Implementasi Kurikulum Operasional Satuan Pendidikan Berbasis Kemaritiman di SMK Negeri 1 Alas Kabupaten Sumbawa. *JIIP - Jurnal Ilmiah Ilmu Pendidikan*, 6(1), 421–434. <https://doi.org/10.54371/jiip.v6i1.1456>
- Wahyuni, S., & Mardapi, D. (2015). Komparasi Metode Estimasi Kesalahan Pengukuran Soal Ujian Akhir Semester Matematika SMA di Kabupaten Lampung Tengah. *Jurnal Evaluasi Pendidikan*, 3(2), Article 2.
- Wiguna, I. K. W. (2022). Langkah Mempercepat Perkembangan Kurikulum Merdeka Belajar. *Edukasi: Jurnal Pendidikan Dasar*, 3(1), 17–26.
- Yusuf, M. (2016). *Metode Penelitian Kuantitatif, Kualitatif & Penelitian Gabungan*. Prenada Media.