

Analisis Butir Soal Bahasa Mandarin Kelas 2 SD JAC Surabaya

Anggi Gratia Putri Tatuwo¹, H. Ruffi'i²

^{1, 2} Program Studi Teknologi Pendidikan, Sekolah Pasca Sarjana, Universitas PGRI
Adibuana Surabaya

e-mail: gratiaputri7@gmail.com

Abstrak

Pembelajaran bahasa Mandarin telah diterapkan sebagai muatan lokal di sekolah dasar (SD). Keterampilan berbahasa asing seperti mendengarkan atau membaca harus diukur untuk mengetahui penguasaan materi pada siswa. Pendidik/evaluator harus mengembangkan tes dalam item pertanyaan. Butir soal dianalisis dari segi validitas, reliabilitas, daya pembeda dan tingkat kesukaran. Tujuan penelitian ini adalah menganalisis butir soal bahasa Mandarin kelas 2 SD JAC Surabaya. Penelitian ini berbentuk penelitian deskriptif kuantitatif. Teknik analisis data dilakukan secara kuantitatif dengan menggunakan aplikasi IBM SPSS 22 dan aplikasi Microsoft Excel, serta secara deskriptif untuk menarik kesimpulan. Hasil penelitian ini meliputi empat hal: (1) Validitas 45% soal tidak valid. Sedangkan 50% soal dinyatakan valid dengan nilai validitas yang berbeda-beda; (2) Reliabilitas soal dinyatakan sebesar 0,686; (3) Daya pembeda pada kategori buruk sebanyak 12 soal; dan (4) Tingkat kesulitan soal cenderung mudah sebanyak 17 soal. Diharapkan dapat menjadi bahan evaluasi dan pengambilan keputusan pada latihan soal selanjutnya.

Kata kunci: *Butir Soal, Bahasa Mandarin, Sekolah Dasar*

Abstract

Chinese language learning has been implemented as a local content in elementary schools (SD). Foreign language skills must be measured to determine the mastery of the material in students. Educators/evaluators should develop tests in question items. Question items are analyzed in terms of validity, reliability, differentiating power and level of difficulty. The purpose of this study was to analyze the 2nd grade Chinese question items at SD JAC Surabaya. This research is in the form of quantitative descriptive research. Data analysis techniques are carried out quantitatively using IBM SPSS 22 applications and Microsoft Excel, and descriptively to draw conclusions. The results: (1) The validity of 45% of the questions is invalid. While 50% are declared valid with different validity values; (2) The reliability of the question is stated at 0.686; (3) Discriminating power in the bad category of 12 questions; and (4) The difficulty level of the questions tends to be easy with 17 questions. It is expected to be a material for evaluation and decision making for the next practice.

Keywords : *Test Items, Mandarin Language, Elementary School*

PENDAHULUAN

Pembelajaran bahasa asing pada pendidikan formal di Indonesia merupakan salah satu upaya dalam memberikan peserta didik kemampuan untuk dapat mengembangkan informasi dan ilmu pengetahuan secara mengglobal (Buan, 2021). Pada pendidikan formal sekolah dasar pembelajaran bahasa asing diimplementasikan pada pembelajaran muatan lokal. Salah satu bahasa asing yang menjadi pembelajaran di beberapa sekolah dasar adalah bahasa Mandarin. Ivana (Ivana, 2022) menjelaskan bahwa seseorang dengan kemampuan bahasa asing, seperti bahasa Mandarin akan lebih memiliki nilai tambah dalam dunia pekerjaan.

Pada pembelajaran bahasa Mandarin, peserta didik akan memiliki empat aspek kemampuan berbahasa, antara lain: menyimak, berbicara, menulis dan membaca. Kemampuan tersebut dapat diukur setelah pembelajaran dengan menggunakan evaluasi. Vega dkk (Ardana et al., 2023) menyatakan bahwa salah satu tujuan evaluasi dilakukan adalah untuk menilai pencapaian kompetensi peserta didik. Musyarwanda Warsah menyimpulkan bahwa evaluasi adalah sebuah proses memberikan nilai kepada peserta didik berdasarkan suatu kriteria tertentu (Musarwan & Warsah, 2022). Karena itu peneliti ingin menyatakan bahwa evaluasi dalam pembelajaran bahasa Mandarin diperlukan untuk mengetahui penguasaan materi pada peserta didik. Sehingga tenaga pendidik sebagai evaluator harus dapat melakukan evaluasi atau penilaian terhadap seluruh proses pembelajaran peserta didik (Band. Purwanto, 2009).

Yuliawati (Yuliawati et al., 2022) menyebutkan evaluasi pembelajaran dapat dilakukan dengan dua teknik, yaitu: Tes dan Non Tes. Teknik Tes dilakukan secara tertulis dan tidak tertulis. Sedangkan teknik Non Tes dilakukan berdasarkan pengamatan untuk menilai sikap, keterampilan dan tingkah laku. Pada teknik tes dapat dikelompokkan lagi berdasarkan objektivitas penskoran, yaitu: Tes Objektif berupa tes salah-benar, tes menjodohkan, tes melengkapi, tes isian singkat dan tes pilihan ganda; dan Tes Subjektif berupa tes uraian dan tes essay. Berdasarkan macam-macam teknik evaluasi tersebut, pendidik harus dapat menentukan jenis tes yang tepat bagi peserta didik sesuai dengan tujuannya. Setelah menentukan teknik penskoran tes, serta jenis soal yang akan digunakan, maka langkah berikutnya adalah pendidik menyusun butir soal.

Dalam penyusunan butir soal, Rasyid mengungkapkan bahwa soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sukar, sehingga peserta didik dapat mengukur kemampuannya dengan pengukuran yang benar (Rasyid et al., 2022). Soal yang dapat digunakan untuk mengukur kompetensi peserta didik dengan tepat adalah soal yang sudah diuji reabilitas dan validitasnya. Nurliani juga menyatakan bahwa menganalisis butir soal merupakan kegiatan mengkaji dan menelaah setiap butir soal, sehingga dapat meningkatkan mutu soal yang ditulis (Siregar et al., 2021). Kegiatan ini berupa proses pengumpulan, peringkasan dan penggunaan informasi dari jawaban siswa untuk membuat keputusan setiap penilaian. Salah satu penilaian yang dilakukan setelah pembelajaran suatu topik adalah latihan soal. Heterion dalam penelitiannya menyatakan bahwa latihan adalah suatu kegiatan pengulangan dalam proses pembelajaran (Heterion et al., 2020). Tujuan dari adanya latihan adalah untuk memotivasi siswa dalam memahami materi pelajaran agar mendapatkan hasil yang lebih baik (Rusman et al., 2021).

SD JAC Surabaya merupakan salah satu sekolah di Surabaya, Jawa Timur yang memiliki pelajaran Bahasa Mandarin sebagai muatan lokal. Pada kelas 2, pembelajaran Bahasa Mandarin menggunakan materi dari buku *keben* dan *lianxice* 千岛华语 *Qiandao Huayu jilid 2*. Buku pelajaran ini terdiri dari 15 bab yang digunakan selama 1 tahun ajaran. Latihan soal yang diberikan oleh pendidik pada peserta didik berupa latihan soal dari buku *lianxice* (Fang & Hengeveld, 2022) yang berisi 6 jenis latihan menulis. Pendidik juga memberikan latihan soal pada peserta didik setelah menempuh 3 bab pembelajaran. Fokus dalam penelitian ini adalah latihan soal 1 yang berisi materi bab 1 sd. 3 pada buku pelajaran. Sehingga berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis butir soal dari latihan soal 1 pelajaran Bahasa Mandarin untuk kelas 2 SD JAC Surabaya. Analisis yang dilakukan pada butir soal adalah untuk mengetahui validitas (Yusup, 2018), reliabilitas, daya pembeda (Band. Dewi et al., 2019) serta tingkat kesukaran.

METODE

Penelitian ini berupa penelitian deskriptif kuantitatif. Menurut Boeren penelitian deskriptif kuantitatif merupakan penelitian yang berfokus pada analisis data berdasarkan data perhitungan atau data statistik yang diperoleh (Boeren, 2018). Kemudian data tersebut akan dijelaskan secara deskriptif. Data yang digunakan pada penelitian ini berupa latihan soal yang berisi 20 nomor pilihan ganda beserta jawaban dari 30 peserta didik kelas 2 SD JAC Surabaya.

Penelitian ini dilakukan pada bulan Oktober 2023. Prosedur penelitian ini (Whidmurni, 2017), antara lain: (1) Menyusun instrumen atau hal-hal yang berkaitan dengan pengukuran, (2) Mengumpulkan data, (3) Mendeskripsikan data yang telah dikumpulkan, (4) Menganalisis data, (5) Merumuskan kesimpulan dari hasil analisis. Pengumpulan data dilakukan dengan mengumpulkan jawaban latihan soal peserta didik. Teknik analisis data dilakukan dengan kuantitatif menggunakan aplikasi IBM SPSS 22 untuk mengukur validitas, reliabilitas. Sedangkan untuk menganalisis daya pembeda soal dan tingkat kesukaran menggunakan aplikasi Microsoft Excel. Hasil perhitungan melalui IBM SPSS 22 maupun Microsoft Excel, kemudian akan dijelaskan secara deskriptif untuk mengambil kesimpulan.

HASIL DAN PEMBAHASAN

Analisis butir soal pada latihan soal bahasa Mandarin untuk kelas 2 SD di JAC Surabaya terdiri dari 4 macam analisis, yaitu analisis validitas, reliabilitas, daya pembeda dan tingkat kesukaran. Hasil penelitian dan pembahasan ini disajikan dalam empat sub-bagian sesuai dengan macam penelitian. Penyusunan pembahasan berdasarkan penyajian data kuantitatif menggunakan tabel, kemudian analisisnya berupa penjelasan deskriptif.

Validitas Latihan Soal

Banyak praktisi penelitian sepakat bahwa validitas latihan soal berbentuk pilihan ganda merupakan validitas isi. Ida Varida menekankan bahwa Validitas isi bertujuan untuk mengukur bagaimana sebuah soal dapat mengukur kemampuan kognitif peserta didik yang disesuaikan dengan tujuan pembelajaran (Ida & Musyarofah, 2021). Mengukur validitas dapat dilakukan dengan menggunakan aplikasi IBM SPSS 22 dengan analisis korelasi biserial yang menghasilkan sebuah tabel korelasi. Dari tabel korelasi, kemudian diinterpretasikan menjadi tabel sederhana, agar analisis dapat terfokus pada hasil Pearson Correlation dan Sig. (2-tailed) dari N_{Total} . Sehingga dapat menentukan valid atau tidak validnya sebuah soal.

Kemudian seturut dengan Febriyanti (Febriyanti & Elfina, 2021) sebuah instrumen dapat dinyatakan valid adalah saat instrumen dapat mengukur objek yang akan diukur dengan benar. Dalam artikel yang singkat, Al Hakim menyatakan bahwa tinggi rendahnya suatu nilai instrumen dapat menentukan seberapa sesuainya instrumen terhadap variabel atau objek yang dia ukur. Oleh karena itu diperlukan sebuah patokan standart pengukuran (Al Hakim et al., 2021). Lebih jauh Sarucu menekankan (Sürücü & Maslakçi, 2020) dengan menyatakan bahwa cara menentukan kesimpulan validitas dari sebuah soal adalah dengan melihat pada nilai Sig. (2-tailed). Jika nilai Sig. (2-tailed) < 0.05, maka soal tersebut dinyatakan valid. Sedangkan untuk interpretasi kevalidan soal, dilihat dari nilai Pearson Correlation yang disesuaikan dengan Nilai Validitas. Nilai validitas yaitu sebagai berikut: 0.80 – 1.00 sangat tinggi, 0.49 – 0.80 tinggi, 0.25 – 0.49 cukup, 0.00 – 0.25 rendah, dan nilai negatif menyatakan tidak valid.

Tabel 1. Interpretasi N_{Total}

No Soal	Pearson Correlation	Sig. (2-tailed)	Kesimpulan	Interpretasi
2	0.217	0.248	Tidak Valid	
3	0.499	0.005	Valid	Tinggi
4	0.383	0.37	Tidak Valid	
5	0.232	0.217	Tidak Valid	
6	-	-	-	-
7	0.112	0.556	Tidak Valid	
8	0.705	0.000	Valid	Sangat Tinggi
9	0.161	0.395	Tidak Valid	
10	0.154	0.416	Tidak Valid	
11	0.112	0.556	Tidak Valid	
12	0.217	0.248	Tidak Valid	
13	0.425	0.019	Valid	Cukup

14	0.546	0.002	Valid	Tinggi
15	0.274	0.143	Tidak Valid	
16	0.402	0.028	Valid	Cukup
17	0.442	0.014	Valid	Cukup
18	0.784	0.000	Valid	Sangat Tinggi
19	0.388	0.034	Valid	Cukup
20	0.528	0.003	Valid	Tinggi
21	0.402	0.028	Valid	Cukup

Berdasarkan hasil analisis validitas di atas, maka dapat disimpulkan bahwa latihan soal bahasa Mandarin ini memiliki 2 soal yang valid dengan nilai sangat tinggi, 3 soal valid dengan nilai tinggi, 5 soal valid dengan nilai cukup, dan tidak ada soal valid dengan nilai rendah. Sedangkan terdapat 9 soal yang tidak valid dan 1 soal yang tidak terdeteksi penilaiannya oleh SPSS.

Reliabilitas Latihan Soal

Duke dkk menyimpulkan reliabilitas sebagai sebuah koefisien yang menunjukkan sejauh mana suatu soal dapat digunakan untuk mengukur sesuatu yang sama secara berulang-ulang dengan responden yang berbeda dikurun waktu yang berbeda pula (Duke et al., 2020). Karena itu reliabilitas sebuah soal dapat terlihat dari hasil yang dimunculkan relatif stabil dan konsisten. Tinggi rendahnya sebuah reliabilitas ditunjukkan oleh koefisien reliabilitas. Besarnya rentan nilai koefisien reliabilitas antara 0 sampai 1. Sehingga semakin tinggi angka reliabilitas, maka semakin besar konsistensi hasil pengukuran. Demikian pula sebaliknya.

Uji reliabilitas pada latihan soal bahasa Mandarin ini bertujuan untuk mengetahui kebiasaan latihan soal untuk dipakai kembali dengan responden yang berbeda. Uji reliabilitas ini menggunakan aplikasi IBM SPSS Statistic 22 dengan uji *Cronbach's Alpha*. Fanani, dkk menjelaskan bahwa uji *Cronbach's Alpha* akan membandingkan nilai *Cronbach's Alpha* sebuah instrumen penilaian dengan nilai koefisien reliabilitas minimal yang dapat diterima, yakni 0.6. Jika nilai *Cronbach's Alpha* > 0.6, maka instrumen penilaian adalah reliabel. Sedangkan jika nilai *Cronbach's Alpha* < 0.6, maka instrument penilaian tidak reliabel.

Tabel 2. Validitas dan Reliability

Case Processing Summary

		N	%
Cases	Valid	30	100.0
	Excluded ^a	0	.0
	Total	30	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
0.686	20

Berdasarkan tabel Reliability Statistic, nilai *Cronbach's Alpha* yang didapat adalah 0.686. Sehingga jika dibandingkan dengan nilai koefisien reliabilitas minimal, maka didapat $0.686 > 0.6$. Oleh karena itu, latihan soal bahasa

Mandarin untuk kelas 2 SD ini dinyatakan reliabel. Sehingga latihan soal ini dapat digunakan kembali sebagai instrumen penilaian dengan responden yang berbeda.

Daya Pembeda Latihan Soal

Setyawarno menyatakan bahwa daya pembeda soal adalah kemampuan soal yang dapat membedakan peserta tes menjadi dua kelompok (Setyawarno, 2017), yaitu: kelompok tinggi dan kelompok rendah. Sehingga semakin tinggi daya pembeda soal, maka akan semakin terlihat kelompok tinggi dapat menjawab benar dan kelompok rendah tidak dapat menjawab dengan benar. Nilai daya pembeda soal dapat dilihat dari indeks daya pembeda soal. Semakin tinggi prosentasi yang didapat berdasarkan indeks, maka makin mudah soal tersebut membagi peserta menjadi kelompok tinggi dan kelompok rendah (Nasobandi, 2022). Untuk mengetahui daya pembeda soal, maka perlu diketahui rumus dan indeks daya pembeda.

Rumus Daya Pembeda
$$D = \frac{BA}{JA} - \frac{BB}{JB}$$

Keterangan:	
D	: Daya Beda
BA	: Jumlah yang menjawab benar kelompok atas
JA	: Jumlah subjek kelompok atas
BB	: Jumlah yang menjawab benar kelompok bawah
JB	: Jumlah subjek kelompok bawah

Indeks Daya Pembeda

Indeks	Kategori
0.70 – 1.00	Baik Sekali
0.40 – 0.69	Baik
0.20 – 0.39	Cukup
0.00 – 0.19	Jelek

Indeks	Kategori
0.70 – 1.00	Baik Sekali
0.40 – 0.69	Baik
0.20 – 0.39	Cukup
0.00 – 0.19	Jelek

Pengujian daya pembeda soal menggunakan rumus secara manual, sehingga menggunakan aplikasi Ms Excel. Pengujian dilakukan dengan mengurutkan terlebih dahulu responden berdasarkan nilai akhirnya. Setelah itu, memecah kolom tersebut menjadi dua bagian, yakni kelompok atas dan kelompok bawah. Berikutnya adalah menghitung sesuai dengan rumus dan mengambil kesimpulan.

Tabel 3. Hasil Analisis Daya Pembeda Latihan Soal

Kriteria	Jumlah Soal	Nomor Soal
Baik Sekali	-	-
Baik	2	8, 18
Cukup	6	4, 5, 14, 16, 19, 20
Jelek	12	2, 3, 6, 7, 9, 10, 11, 12, 13, 15, 17, 21

Berdasarkan tabel-tabel analisis daya pembeda soal, maka didapat ada 2 soal berkategori baik, 6 soal berkategori cukup, dan 12 soal berkategori jelek. Dari hasil ini maka dapat disimpulkan bahwa daya pembeda dari latihan soal Mandarin untuk kelas 2 SD ini masih rendah. Hal ini menunjukkan bahwa latihan soal ini perlu diperbaiki.

Tingkat Kesukaran Latihan Soal

Tingkat kesukaran soal menurut Sopacua berbicara tentang seberapa mudah atau susah nya sebuah soal bagi peserta didik yang mengerjakannya. Semakin besar presentase peserta didik menjawab benar sebuah soal, maka soal tersebut dinyatakan mudah (Sopacua & Rahardjo, 2021). Demikian sebaliknya, semakin kecil presentase peserta didik menjawab benar sebuah soal, maka soal tersebut dinyatakan susah.

Rumus Tingkat Kesukaran
$$P = \frac{Np}{N}$$

Keterangan:
 P : Proporsi atau angka indeks kesukaran butir soal
 Np : Banyaknya peserta yang menjawab benar
 N : Jumlah peserta tes

Indeks Tingkat Kesukaran

Indeks	Kategori
< 0.29	Sukar
0.30 – 0.69	Sedang
> 0.70	Mudah

Pengujian tingkat kesukaran menggunakan rumus secara manual, sehingga menggunakan aplikasi Ms Excel. Pengujian dilakukan dengan mengurutkan terlebih dahulu responden berdasarkan nilai akhirnya. Setelah itu, menghitung sesuai dengan rumus dan mengambil kesimpulan. Berikut ini adalah analisis tingkat kesukaran.

Tabel 4. Tingkat Kesukaran

No Soal	Np	N	P	Kategori		
				Sukar	Sedang	Mudah
2	26	30	0,87			√
3	26	30	0,87			√
4	23	30	0,77			√
5	23	30	0,77			√
6	28	30	0,93			√
7	27	30	0,90			√
8	14	30	0,47		√	
9	26	30	0,87			√
10	24	30	0,80			√
11	27	30	0,90			√
12	26	30	0,87			√
13	27	30	0,90			√
14	18	30	0,60		√	
15	26	30	0,87			√
16	24	30	0,80			√
17	26	30	0,87			√
18	18	30	0,60		√	
19	25	30	0,83			√
20	25	30	0,83			√
21	24	30	0,80			√

Berdasarkan tabel di atas tingkat kesukaran, maka didapat ada 3 soal dengan tingkat kesukaran sedang dan 17 soal dengan tingkat kesukaran mudah. Oleh karena itu, disimpulkan bahwa latihan soal Mandarin memiliki tingkat kesukaran yang rendah. Hal ini terlihat dari banyaknya soal yang mudah untuk dijawab benar oleh peserta didik. Sehingga perlu sekali adanya perbaikan pada soal yang diberikan.

SIMPULAN

Hasil penelitian analisis butir soal pada latihan soal bahasa Mandarin untuk kelas 2 SD telah ditemukan. Karena itu diharapkan perbaikan soal untuk Latihan di waktu mendatang dapat dilakukan dengan mengganti pertanyaan, mengubah bahan *hanzi* dalam pertanyaan dan menambahkan pertanyaan seputar keterampilan mendengar.

DAFTAR PUSTAKA

- Al Hakim, R., Mustika, I., & Yuliani, W. (2021). VALIDITAS DAN RELIABILITAS ANGKET MOTIVASI BERPRESTASI. *FOKUS (Kajian Bimbingan & Konseling Dalam Pendidikan)*, 4(4), 263. <https://doi.org/10.22460/fokus.v4i4.7249>
- Ardana, L. N., Vega, N., Kirana, P., & Sunaryati, T. (2023). Peran Penting Evaluasi Pembelajaran pada Mata Pelajaran Pendidikan Kewarganegaraan di Sekolah Dasar Berbasis Pendidikan Karakter. *Jurnal Pendidikan Tambusai*, 7(2), 15814-15819. <https://doi.org/10.31004/jptam.v7i2.8871>
- Boeren, E. (2018). The Methodological Underdog: A Review of Quantitative Research in the Key Adult Education Journals. *Adult Education Quarterly*, 68(1), 63–79. <https://doi.org/10.1177/0741713617739347>
- Buan, Y. A. L. (2021). *Guru dan Pendidikan Karakter: Sinergitas Peran Guru Dalam Menanamkan Nilai-Nilai Pendidikan Karakter di Era Milenial*. Penerbit Adab.
- Dewi, S. S., Hariastuti, R. M., & Utami, A. U. (2019). ANALISIS TINGKAT KESUKARAN DAN DAYA PEMBEDA SOAL OLIMPIADE MATEMATIKA (OMI) TINGKAT SMP TAHUN 2018. *Transformasi: Jurnal Pendidikan Matematika Dan Matematika*, 3(1), 15–26. <https://doi.org/10.36526/tr.v3i1.388>
- Duke, C., Hamidi, S., & Ewing, R. (2020). Validity and Reliability. In R. Ewing & K. Park (Eds.), *Basic Quantitative Research Methods for Urban Planners* (1st ed., pp. 88–106). Routledge. <https://doi.org/10.4324/9780429325021-6>
- Fang, H., & Hengeveld, K. (2022). SENTENCE-FINAL Particles in Mandarin *. *Studia Linguistica*, 76(3), 873–913. <https://doi.org/10.1111/stul.12198>
- Febriyanti, R., & Elfina, E. (2021). Hubungan Pengetahuan, Sikap, Tingkat Religiusitas, dan Faktor Lain Terhadap Perilaku Seksual Remaja Di SMK X Bogor Tahun 2016. *Jurnal Pendidikan Dan Konseling (JPDK)*, 3(2), 220–233. <https://doi.org/10.31004/jpdk.v3i2.7823>
- Hetarion, B. D. S., Hetarion, Y., & Makaruku, V. (2020). Implementasi Pendidikan Karakter Berbasis Kearifan Lokal Cuci Negeri dalam Pembelajaran IPS. *Tehnologi Pendidikan*, 22(1), 1–12. <https://doi.org/10.21009/jtp.v22i1.15328>
- Ida, F. F., & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *AL-MU'ARRIB: JOURNAL OF ARABIC EDUCATION*, 1(1), 34–44. <https://doi.org/10.32923/al-muarrib.v1i1.2100>
- Ivana, I. (2022). Peningkatan Minat Belajar dengan Materi Kebudayaan Cina dalam Belajar Bahasa Mandarin. *Jurnal Pendidikan Tambusai*, 6(2), 11469–11472. <https://doi.org/10.31004/jptam.v6i2.4264>
- Musarwan, M., & Warsah, I. (2022). Evaluasi Pembelajaran (Konsep, Fungsi dan Tujuan) Sebuah Tinjauan Teoritis. *Jurnal Kajian Pendidikan Islam*, 186–199. <https://doi.org/10.58561/jkpi.v1i2.35>
- Nasobandi, A. (2022). Pengaruh Model Pembelajaran Project Based Learning Terhadap Kemampuan Berfikir Kreatif. *INTELEKTUUM*, 3(2). <https://doi.org/10.37010/int.v3i2.1071>
- Purwanto. (2009). *Evaluasi Pembelajaran*. Pustaka Belajar.

- Rasyid, A. H. A., Yunitasari, B., Susila, I. W., Dewanto, D., Yunus, Y., & Santoso, D. I. (2022). PENGEMBANGAN MODEL EVALUASI PEMBELAJARAN BERBASIS OBE. *Jurnal Pendidikan (Teori Dan Praktik)*, 7(1), 8–17. <https://doi.org/10.26740/jp.v7n1.p8-17>
- Rusman, R., Kurniawan, D., & Riyan, C. (2021). Pembelajaran Berbasis Teknologi Informasi dan Komunikasi. *Raja Grafindo Persada*, 169.
- Setyawarno, D. (2017). Penggunaan Aplikasi Software Iteman (Item and Test Analysis) untuk Analisis Butir Soal Pilihan Ganda Berdasarkan Teori Tes Klasik. *Jurnal Ilmu Fisika Dan Pembelajarannya (JIFP)*, 1(1), 11–21. <https://doi.org/10.19109/jifp.v1i1.866>
- Siregar, N., Siregar, H., & Hutahaean, H. (2021). Application of the Picture and Picture Type of Cooperative Learning Model in Improving Student Learning Creativity. *TP -Jurnal Teknologi Pendidikan*, 23(1), 23–36. <https://doi.org/10.21009/JTP2001.6>
- Sopacua, E. E. D., & Rahardjo, M. M. (2021). ANALISA PEMBELAJARAN TEMATIK DALAM PENDIDIKAN ANAK USIA DINI. *Satya Widya*, 36(1), 64–76. <https://doi.org/10.24246/j.sw.2020.v36.i1.p64-76>
- Sürücü, L., & Maslakçı, A. (2020). VALIDITY AND RELIABILITY IN QUANTITATIVE RESEARCH. *Business & Management Studies: An International Journal*, 8(3), 2694–2726. <https://doi.org/10.15295/bmij.v8i3.1540>
- Whidmurni, W. (2017). Penelitian Kuantitatif. *Pemaparan Metode Penelitian Kuantitatif*, 2(16), 1–16.
- Yuliawati, D. D., Bintang, K. A., Fath A., M. S., Fitrotul, R., Gheafitri, Z., & Hanatan, A. (2022). ANALISIS KEMAMPUAN BERPIKIR KRITIS PESERTA DIDIK DENGAN PENILAIAN TES DAN NON TES. *JURNAL PEMBELAJARAN FISIKA*, 11(2), 65. <https://doi.org/10.19184/jpf.v11i2.31048>
- Yusup, F. (2018). Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif. *Jurnal Tarbiyah : Jurnal Ilmiah Kependidikan*, 7(1). <https://doi.org/10.18592/tarbiyah.v7i1.2100>