

HUBUNGAN ANTARA VALIDITAS ITEM DENGAN DAYA PEMBEDA DAN TINGKAT KESUKARAN SOAL PILIHAN GANDA PAS

Sri Nurhalimah¹, Yunin Hidayati², Irsad Rosidi³ dan Wiwin Puspita Hadi⁴

^{1,2,3,4} Pendidikan IPA, FIP, Universitas Trunojoyo Madura, Bangkalan, 69162, Indonesia
nhalimah226@gmail.com

Diterima tanggal: 10 Agustus 2021 Diterbitkan tanggal: 15 Maret 2022

Abstrak Penelitian ini bertujuan untuk mengetahui hubungan validitas, daya pembeda dan tingkat kesukaran pada butir soal pilihan ganda PAS ganjil mata pelajaran IPA kelas VIII UPTD SMPN 5 Bangkalan tahun ajaran 2019/2020. Penelitian ini merupakan penelitian kuantitatif deskriptif. Subjek dalam penelitian adalah peserta didik kelas VIII A, C, D, dan E UPTD SMPN 5 Bangkalan. Hasil penelitian ini menunjukkan bahwa validitas item dan daya pembeda soal memiliki nilai yang sama besar. Sehingga menunjukkan adanya hubungan antara validitas item dan daya pembeda soal, atau korelasi searah. Nilai indeks kesukaran yang besar tidak menunjukkan koefisien validitas item yang semakin besar sehingga tidak terjadi korelasi antara validitas item dengan indeks kesukaran, serta menunjukkan bahwa tidak adanya hubungan antara validitas item dengan tingkat kesukaran.

Kata Kunci: Analisis Soal, Kualitas Soal, Soal PAS ganjil.

Abstract *This study aims to determine the relationship between the validity, level of difficulty, and the distinguishing power of the multiple-choice items of odd PAS in the science subject of the 8th class of in UPTD SMPN 5 Bangkalan, which is reviewed based on the validity, reliability, the level of difficulty, the distinguishing, and the effectiveness of the deceive choice. This research used a descriptive quantitative method. The subjects of this research are the students of 8th class A, C, D, and F with the purposive sampling technique. The result of this research showed that the item validity and the distinguishing power of the questions have the same value. So that it shows a relationship between the validity of the item and the distinguishing power of the problem, or shows the direction of the direction. A large difficulty index value does not indicate a greater item validity coefficient, so there is no between item validity and difficulty index, and indicates that there is no relationship between item validity and difficulty level.*

Keywords: *Item Analysis, Quality of the question, Odd question PAS.*

Pendahuluan

Pendidikan mempunyai peranan penting dalam mencerdaskan bangsa menjadi generasi berkarakter dan berjiwa kompetitif (Rahmasari & Ismiyati, 2016). Pendidikan juga berperan dalam pembangunan nasional dalam pencapaian bangsa yang maju dan mandiri. Guru harus memiliki kompetensi dalam mewujudkan pendidikan yang bermutu. Faktor dalam mewujudkan tujuan pendidikan nasional merupakan kompetensi guru dalam mencetak lulusan yang kompeten. Sebagai tenaga pendidik, guru harus memiliki potensi dalam pengetahuan terkait perencanaan, pelaksanaan, serta evaluasi pembelajaran (Fitriatun & Sukanti, 2016).

Penilaian dalam pendidikan, disama artikan dengan istilah evaluasi. Evaluasi merupakan upaya yang dapat dilakukan dalam mengetahui pencapaian hasil belajar dan

perbaiki proses pembelajaran. Evaluasi mengacu pada proses sistematis dalam menentukan tujuan instruksional yang dicapai oleh peserta didik (Kusumawati & Hadi, 2018). Pencapaian tujuan pembelajaran dilakukan melalui suatu pengukuran dan penilaian. Dengan pengukuran dan penilaian, guru dapat mendiagnosis kemampuan peserta didik sehingga dapat mengetahui tindakan yang akan dilakukan untuk meningkatkan kemampuan peserta didik.

Penilaian dan pengukuran merupakan sesuatu yang sangat berkaitan dengan evaluasi (Wibawa, 2019). Pada umumnya penilaian dan evaluasi adalah sistem yang secara sistematis untuk mengetahui tingkat keberhasilan dan efisiensi proses pembelajaran. Evaluasi pendidikan sebagai kegiatan pemberian nilai dan makna terhadap hasil asesmen dalam pembelajaran yang sesuai dengan standar yang telah ditetapkan (Yusuf, 2015). Penilaian memiliki makna yang berbeda dengan evaluasi. Penilaian menjelaskan hasil pengukuran pada satu komponen, sedangkan evaluasi adalah penetapan nilai dan arti berupa komponen yang didasari pertimbangan dan pemberian keputusan (Arifin, 2016).

Salah satu alat yang digunakan sebagai sarana dalam penilaian hasil belajar yaitu tes. Tes merupakan alat ukur yang khusus digunakan untuk mengetahui kemampuan kognitif peserta didik (Yusrizal, 2016). Tes biasanya menggunakan soal objektif, terutama pada soal pilihan ganda. Soal pilihan ganda merupakan bentuk tes yang memiliki alternative jawaban sehingga peserta didik dapat memilih jawaban yang dianggap benar (Amelia, 2017). Soal pilihan ganda dapat mencakup lebih banyak materi, efisien dan mempermudah guru dalam penilaian. Tes dibagi menjadi dua segi yaitu *standardized test* dan *teacher-made test*. *Standardized test* adalah tes yang telah distandarisasi melalui proses validitas dan reliabilitas. Sehingga tes dinyatakan valid dan *reliable* untuk diujikan kepada peserta didik. Sedangkan *teacher-made test* adalah tes yang disusun untuk mengetahui ketercapaian hasil belajar peserta didik. Tes buatan guru terbatas pada satu pengajar yang sama atau satu sekolah (Septiana, 2016).

Menurut Nasir (2015) menyatakan bahwa kurangnya kompetensi guru dalam menganalisis soal yang melibatkan rumus-rumus yang komplis sehingga sulit dipahami serta belum ada alat bantu yang dapat mempermudah guru dalam hal menganalisis kualitas soal. Selain itu, materi yang diujikan guru juga belum mengetahui sesuai atau tidaknya dengan materi yang disampaikan dalam pembelajaran. Menganalisis butir soal merupakan upaya untuk mengetahui kualitas tes. Analisis butir soal merupakan cara dalam meningkatkan kualitas soal tes yang harus dilakukan oleh guru sebelum diujikan terhadap peserta didik (Kurniawan, 2015).

Analisis butir soal dapat digunakan dalam penentuan tindak lanjut terhadap soal yang telah diujikan. Secara umum, tindakan yang dapat dilakukan yaitu soal dapat digunakan kembali, soal perlu direvisi atau soal tidak dapat digunakan kembali (Zein, Fadillah, & Novianti, 2013). Namun, sampai saat ini belum ada acuan atau referensi yang akurat dalam menentukan syarat suatu soal dapat digunakan, direvisi atau tidak dapat digunakan dengan mempertimbangkan validitas item, daya pembeda dan tingkat kesukaran. Untuk itu, perlu mengetahui bentuk hubungannya dalam penentuan kualitas soal. Penelitian ini bertujuan untuk mengetahui relasi yang terdapat antara validitas item dengan daya pembeda dan tingkat kesukaran soal pilihan ganda PAS.

Metode Penelitian

Jenis penelitian yang dilakukan yaitu jenis penelitian analisis deskriptif kuantitatif. Penelitian ini dilaksanakan pada saat pelaksanaan PAS ganjil mata pelajaran IPA kelas VIII SMP tahun ajaran 2019/2020 di UPTD SMPN 5 Bangkalan. Terletak di JL. Mawar No. 177 Mlajah Bangkalan.

Populasi pada penelitian ini adalah seluruh siswa kelas VIII UPTD SMPN 5 Bangkalan, yang terdiri 8 kelas. Sampel yang digunakan dalam penelitian ini adalah kelas

VIII A, VIII C, VIII D, dan VIII E adalah 30 siswa, sehingga total sampel yang digunakan adalah 120 siswa. Bentuk tes yang dianalisis adalah pilihan ganda.

Teknik pengumpulan data pada penelitian ini adalah cara memperoleh data untuk mempermudah dalam melakukan penelitian. Observasi adalah pengamatan secara langsung ke lapangan untuk mengetahui permasalahan yang berhubungan dengan penelitian. Dokumentasi yaitu pengumpulan data melalui referensi tertulis berupa dataprimer yaitu kisi-kisi soal, kunci jawaban, soal dan lembar jawaban peserta didik PAS ganjil mata pelajaran IPA kelas VIII tahun ajaran 2019A/2020 di Kabupaten Bangkalan, dan wawancara dilakukan secara terbuka yang diajukan pada guru pengajara IPA kelas VIII di UPTD SMPN 5 Bangkalan dan penyusun soal PAS ganjil mata pelajaran IPA kelas VIII tahun ajaran 2019/2020.

Teknik Analisis data secara deskriptif kuantitatif pada butir soal PAS pilihan ganda melalui program ITEMAN untuk memperoleh hasil validitas, daya pembeda, dan tingkat kesukaran. Aspek validitas, daya pembeda dan tingkat kesukaran digunakan untuk mengetahui hubungan dari validitas terhadap daya pembeda dan tingkat kesukaran.

Validitas menunjukkan ukuran suatu instrumen dalam tingkat kevalidannya. Instumen dinyatakan valid apabila dapat mengukur apa yang seharusnya diukur. Tinggi rendahnya suatu instrumen menunjukkan sejauh mana data yang didapat sesuai dengan variabel yang dimaksud (Azra & Huda, 2017). Validitas butir soal menggunakan teknik *Korelasi Biserial*. Interpretasi indeks korelasi *Point Biserial* yang menunjukkan nilai validitas pada kriteria Tabel 1.

Tabel 1. Nilai Validitas

$0,80 \leq r_{pbi} < 1,00$	=	Sangat tinggi
$0,49 \leq r_{pbi} < 0,80$	=	Tinggi
$0,25 \leq r_{pbi} < 0,49$	=	Cukup
$0,00 \leq r_{pbi} < 0,25$	=	Rendah
Negatif	=	tidak valid

(Basuki & Hariyanto, 2016)

Daya pembeda dalam suatu tes bertujuan untuk mengetahui perbedaan peserta didik kemampuan peserta didik. Semakin tinggi indeks yang dimiliki oleh butir soal, maka semakin baik butir soal tersebut karena memiliki daya untuk membedakan kemampuan peserta didik yang pandai dan kurang pandai. Sebaliknya jika semakin rendah indeks yang dimiliki oleh butir soal, maka semakin rendah soal tersebut membedakan kemampuan peserta didik yang pandai dan kurang pandai (Amelia, 2017). Koefisien daya beda butir soal berkisar dari -1,00 sampai +1,00. Jika suatu butir soal memiliki tanda negatif maka dapat dinyatakan bahwa soal tersebut menyesatkan karena peserta dari kelompok pandai menjawab salah sehingga harus dihilangkan atau dibuang. Untuk menginterpretasikan koefisien daya pembeda table 2.

Tabel 2. Daya Beda

$0,40 \leq D < 1,00$	=	Sangat baik
$0,30 \leq D < 0,39$	=	Baik
$0,20 \leq D < 0,29$	=	Cukup
$0,00 \leq D < 0,19$	=	Jelek
Negatif	=	<i>No Discrimination</i>

(Arifin, 2016)

Tingkat kesukaran soal bertujuan untuk mengetahui seberapa sukar soal tersebut. Kualitas soal yang baik, disamping memenuhi validitas dan reliabilitas adalah adanya keseimbangan dari tingkat kesulitan soal tersebut. Soal yang seimbang sesuai dengan proporsi yaitu soal mudah 15%, sedang 75%, dan sukar 15% (Susanto, Rinaldi, & Novalia, 2015). Suatu soal hendaknya tidak terlalu sukar dan tidak terlalu mudah. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya (Solichin, 2017). Kriteria untuk menafsikan indeks kesukaran Tabel 3.

Tabel 3. Indeks Kesukaran

$P \leq 0,30$	=	Sulit
$0,30 \leq P < 0,70$	=	Sedang
$P \geq 0,70$	=	Mudah

(Modifikasi: Arifin, 2016)

Hasil Penelitian dan Pembahasan

Penelitian analisis butir soal dilakukan untuk mengetahui hubungan validitas terhadap tingkat kesukaran dan daya pembeda. Data yang digunakan adalah soal pilihan ganda PAS (Penilaian Akhir Semester) ganjil mata pelajaran IPA kelas VIII tahun ajaran 2019/2020 di UPTD SMPN 5 Bangkalan dengan sampel yang berjumlah 120 peserta didik. Butir soal berbentuk pilihan ganda berjumlah 40 soal. Analisis kualitas butir soal menggunakan metode analisis deskriptif kuantitatif.

Validitas menunjukkan kemampuan suatu alat ukur dalam mengukur suatu objek penelitian yang seharusnya diukur dan sesuai dengan variable yang dimaksud. Tinggi rendahnya suatu instrumen menunjukkan sejauh mana data yang didapat sesuai dengan variabel yang dimaksud.

Tabel 4 Hasil Validitas

Kriteria	Jumlah	Persentase	Nomor Soal
sangat tinggi	0	0%	0
tinggi	10	25%	6,9,13,20,23,24,29,32,38,39
cukup	21	53%	1,3,8,10,11,14,17,18,19,21,22,25,26,27,30,31,34,35,36,37,40
rendah	6	15%	2,4,5,12,15,16
tidak valid	3	8%	7.28,33

Berdasarkan hasil analisis validitas pada tabel 4 rata-rata validitas soal PAS memiliki nilai validitas cukup yaitu 21 butir soal, dan tidak terdapat soal dengan nilai validitas soal yang sangat tinggi. Sebanyak 10 butir soal dengan nilai koefisien validitas tinggi, 6 butir soal dengan kategori rendah. Nilai koefisien validitas negatif sebanyak 3 butir soal yang dinyatakan tidak valid. Butir soal yang tidak valid berarti soal tersebut tidak berfungsi sebagai alat ukur yang seharusnya, serta mempengaruhi daya beda antara peserta didik kemampuan tinggi dengan kemampuan rendah.

Indeks daya beda soal perbedaan skala antara kelompok atas dengan kelompok bawah yang menjawab benar suatu tes (Iskandar & Rizal, 2017). Daya beda bertujuan untuk mengetahui kemampuan setiap peserta didik, dan validitas sebagai alat ukur untuk mengetahui seberapa berfungsi butir soal tersebut dalam membedakan peserta didik yang memiliki kemampuan tinggi dan kemampuan rendah. Butir soal yang dapat membedakan peserta didik berkemampuan tinggi dengan peserta didik berkemampuan rendah jika memiliki

nilai korelasi ≥ 0.20 . Jika nilai korelasi ≤ 0.20 dinyatakan butir soal tidak dapat membedakan peserta didik berkemampuan tinggi dengan peserta didik berkemampuan rendah.

Tabel 5 Hasil Daya Pembeda

Kriteria	jumlah	Persentase	nomor soal
Sangat Baik	17	43%	6,9,11,13,18,19,20,21,22,23 24,29,30,32,37,38,39
Baik	10	25%	1,3,10,14,17,26,27,31,36,40
Cukup	6	15%	2,4,8,25,34,35
Jelek	4	10%	5,12,15,16
No Discrimination	3	8%	7,28,33

Berdasarkan hasil analisis daya pembeda pada tabel 5, terdapat 17 butir soal memiliki daya beda sangat baik menunjukkan soal mampu membedakan peserta didik berkemampuan tinggi dengan peserta didik kemampuan rendah, terdapat 10 butir soal memiliki daya beda soal baik dalam membedakan peserta didik kemampuan tinggi dengan kemampuan rendah secara signifikan. Daya beda cukup menunjukkan butir soal yang mampu membedakan peserta didik kemampuan tinggi dengan peserta didik kemampuan rendah yaitu 6 butir soal yang tidak signifikan sehingga perlu adanya revisi untuk perbaikan kualitas soal. Sebanyak 4 butir soal memiliki daya beda jelek sehingga tidak mampu membedakan peserta didik kemampuan tinggi dengan peserta didik kemampuan rendah, dan sebanyak 3 butir soal dengan nilai korelasi negatif yang menunjukkan bahwa tidak dapat membedakan peserta didik kemampuan tinggi dengan peserta didik kemampuan rendah. Menurut Hanifah (2014), butir soal dengan kategori *No Discrimination* berarti soal lebih banyak dijawab benar oleh peserta didik kemampuan rendah dari pada peserta didik kemampuan tinggi sehingga tidak layak digunakan kembali karena tidak dapat membedakan kemampuan peserta didik. Salah satu faktor penyebab nilai korelasi jelek dan negatif yaitu koefisien validitas soal yang tidak valid dan tingkat kesukaran tiap soal berbeda.

Tabel 6 Hasil Tingkat Kesukaran

Kriteria	Jumlah	Persentase	Nomor Soal
Sulit	8	20%	1,7,15,17,31,33,34,35
Sedang	29	73%	2,4,5,6,9,10,11,12,13,14,16,18,19,20,21, 22,23,24,25,26,27,28,29,30,32,37,38,39,40
Mudah	3	8%	3,8,36

Berdasarkan hasil analisis tingkat kesukaran pada tabel 6 menunjukkan soal dalam kategori sulit sebanyak 8 butir soal, soal dengan kategori sedang sebanyak 29 butir soal, dan sebanyak 3 butir soal dengan kategori soal mudah. Soal dengan tingkat kesukaran yang terlalu sulit dan mudah tidak layak digunakan. Soal yang terlalu sulit dapat membuat peserta didik putus asa dalam menyelesaikannya sedangkan soal yang terlalu mudah tidak merangsang kemampuan peserta didik dalam memecahkan masalah (Khaerudin, 2015). Butir soal PAS ganjil rata-rata berkategori sedang yang berarti soal PAS bertujuan merangsang kemampuan peserta didik dalam menyelesaikan masalah.

Tabel 7 Distribusi Indeks Validitas Item, Daya pembeda dan Tingkat Kesukaran

Nomor Soal	Tingkat Kesukaran	Validitas	Daya Pembeda
26	0,884	0,307	0,307
3	0,727	0,338	0,338
8	0,711	0,261	0,261
24	0,694	0,498	0,498
11	0,595	0,397	0,397
29	0,595	0,578	0,578
30	0,595	0,412	0,412
21	0,587	0,416	0,416
36	0,579	0,37	0,37
25	0,545	0,283	0,283
12	0,529	0,157	0,157
18	0,521	0,467	0,467
13	0,512	0,529	0,529
32	0,504	0,623	0,623
2	0,496	0,234	0,234
22	0,479	0,411	0,411
37	0,479	0,487	0,487
14	0,471	0,38	0,38
5	0,463	0,186	0,186
6	0,455	0,597	0,597
16	0,455	0,003	0,003
23	0,446	0,544	0,544
27	0,446	0,38	0,38
9	0,421	0,536	0,536
40	0,43	0,388	0,388
39	0,413	0,565	0,565
10	0,405	0,297	0,297
28	0,405	-0,218	-0,218
19	0,397	0,425	0,425
38	0,364	0,57	0,57
4	0,339	0,226	0,226
20	0,322	0,584	0,584
17	0,298	0,314	0,314
33	0,248	-0,012	-0,012
34	0,231	0,27	0,27
15	0,207	0,025	0,025
1	0,165	0,335	0,335
35	0,165	0,274	0,274
31	0,116	0,301	0,301
7	0,107	-0,039	-0,039

Daya beda dapat dihitung melalui statistik nilai korelasi point biserial. Korelasi point biserial juga untuk mengetahui kerelasian antara skor jawaban suatu soal dan pola jawaban peserta didik (Zein et al., 2013). Penentuan nilai daya pembeda soal dan validitas item

memiliki sistematis yang sama, yaitu dengan korelasi poin biserial. Hal ini yang menunjukkan bahwa terdapat kereliasian antara validitas item dengan daya pembeda (Zein et al., 2013). Bentuk korelasi yang ditunjukkan adalah korelasi searah. Dimana indeks daya pembeda yang semakin besar maka semakin besar pula validitas item. Bentuk korelasi searah ini dapat dilihat pada tabel 7. Berdasarkan analisis indeks validitas item dan daya pembeda pada tabel 4 soal memiliki nilai yang sama besar. Sehingga menunjukkan adanya hubungan antara validas item dan daya pembeda soal, atau korelasi searah.

Tingkat kesukaran merupakan rasio atau parameter yang mendeskripsikan seberapa sukar soal tes yang diberikan pada peserta ujian untuk memberikan jawaban yang benar terhadap suatu butir soal (Suryani, 2017). Koefisien korelasi point biserial pada dasarnya menunjukkan kereliasian antara skor jawaban suatu item. Nilai koefisien yang tinggi menunjukkan bahwa peserta didik dapat menjawab dengan benar suatu soal. Sebaliknya nilai koefisien yang rendah menunjukkan bahwa butir soal tidak dapat dijawab dengan benar. Nilai koefisien korelasi biserial menunjukkan yang semakin tinggi menunjukkan bahwa suatu item memiliki indeks kesukaran baik (Zein et al., 2013).

Koefisien validitas item menunjukkan seberapa tinggi skor total suatu item terhadap skor keseluruhan. Validitas merupakan kegiatan analisis untuk mengidentifikasi keakuratan dari keseluruhan tes dalam menjalankan fungsi ukurnya (Kurniawan, 2015). Jadi, validitas yaitu ukuran standar yang menunjukkan tingkat ketepatan suatu instrumen. Analisis butir soal dari segi validitas dilakukan untuk menyatakan apakah tes sudah tepat digunakan sebagai alat ukur atau tidak (Amalia, Kurniawan, & Antosa, 2017). Besarnya koefisien validitas item dapat ditentukan dengan korelasi poin biserial.

Bentuk korelasi yang berkaitan adalah korelasi searah. Semakin besar indeks kesukaran maka semakin besar koefisien validitas item. Bentuk korelasi searah antara validitas item dan indek kesukaran, dapat dilihat pada tabel 7. Berdasarkan analisis pada tabel 7 nilai indeks kesukaran yang besar tidak menunjukkan koefisien validitas item yang semakin besar sehingga tidak terjadi korelasi antara validitas item dengan indeks kesukaran. Hal ini selaras dengan penelitian Zein et al. (2013), menyatakan bahwa tidak terdapat korelasi antara indeks kesukaran dengan validitas item. Validitas item tidak bergantung pada banyaknya jumlah jawaban benar. Suatu butir soal yang dijawab benar atau salah oleh peserta didik belum tentu menghasilkan nilai koefisien yang semakin tinggi atau rendah. Sebaliknya pada indeks kesukaran soal yang semakin banyak jumlah yang dijawab benar, maka butir soal tersebut tergolong mudah.

Kesimpulan dan Saran

Berdasarkan hasil penelitian dapat disimpulkan bahwa daya pembeda soal kereliasian dengan validitas item. Validitas item dan daya pembeda soal memiliki nilai yang sama besar. Sehingga menunjukkan adanya hubungan antara validas item dan daya pembeda soal, daya pembeda soal dan validitas item juga memiliki korelasi searah, dimana semakin tinggi nilai daya beda maka validitas soal semakin tinggi. Pada tingkat kesukaran tidak memiliki kereliasian dengan validitas item, karena validitas item tidak dipengaruhi oleh banyaknya jumlah jawaban benar, dan juga tingkat kesukaran jika semakin banyak menjawab benar maka soal tersebut dinyatakan mudah. Serta menunjukkan bahwa tidak adanya hubungan antara validitas item dengan tingkat kesukaran.

Daftar Pustaka

Amalia, R., Kurniawan, O., & Antosa, Z. (2017). Analisis Butir Soal Ujiab Akhir Semester dalam Penerapan Kurikulum 2013 di Kelas III SD Negeri 111 Pekanbaru. *Jurnal Online*

Mahasiswa Fakultas Keguruan Dan Ilmu Pendidikan, 4(1), 1–10.

- Amelia, M. A. (2017). Analisis Soal Tes Hasil Belajar High Order Thinking Skills (HOTs) Matematika Materi Pecahan Untuk Kelas 5 Sekolah Dasar. *Jurnal Penelitian, 20(2)*, 123–131. Retrieved from <http://e-journal.usd.ac.id/index.php/JP/article/viewFile/869/685>
- Arifin, Z. (2016). *Evaluasi Pembelajaran*. Bandung: Rodaskarya.
- Azra, U., & Huda, I. (2017). Analisis Pilihan Ganda Materi Sistem Pernapasan Untuk Penrapan Pembelajaran Menggunakan Media Audio Visual Pada Pesantren Moderen Kota Banda Aceh. *Prosiding Seminar Nasional Biotik*, 428–433.
- Basuki, I., & Hariyanto. (2016). *Asesmen Pembelajaran*. Bandung: PT. Remaja Rodaskarya Offset.
- Fitriatun, A., & Sukanti. (2016). Analisis Validitas, Reliabilitas dan Butir Soal Latihan Ujian Nasional Ekonomi Akuntansi di MAN Maguwaharjo. *Jurnal Kajian Pendidikan Akuntansi Indonesia, 8(3)*, 1–11.
- Hanifah, N. (2014). Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi. *SOSIO E-KONS, 6(1)*, 41–55. <https://doi.org/http://dx.doi.org/10.30998/sosioekons.v6i1.1715>
- Iskandar, A., & Rizal, M. (2017). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan, 21(2)*, 12–23. <https://doi.org/10.21831/pep.v22i1.15609>
- Khaerudin. (2015). Kualitas Instrumen Tes Hasil Belajar. *Jurnal Madaniyah, 2(9)*, 212–235.
- Kurniawan, T. (2015). Analisis Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran IPS Sekolah Dasar. *Journal of Elementary Education, 4(1)*, 1–6. Retrieved from <http://journal.unnes.ac.id/sju/index.php/jee>
- Kusumawati, M., & Hadi, S. (2018). An analysis of Multiple Choice Questions (MCQs): Item and Test Statistics From Mathematics Assessments in Senior High School. *Research and Evaluation in Education, 4(1)*, 70–78. <https://doi.org/10.21831/reid.v4i1.20202>
- Nasir, M. (2015). Analisis Empirik Program Analisis Butir Soal Dalam Rangka Menghasilkan Soal yang Baik dan Bermutu Sebagai Alat Evaluasi Pembelajaran Fisika. *Prosiding Semirata, 336–347*.
- Rahmasari, D., & Ismiyati. (2016). Analisis Butir Soal Mata Pelajaran Pengantar Administrasi perkantoran. *Economic Education Analysis Journal, 5(1)*, 317–330.
- Septiana, N. (2016). Analisis Butir Soal Ulangan Akhir Semester (Uas) Biologi Tahun Pelajaran 2015/2016 Kelas X Dan Xi Pada Man Sampit. *Jurnal Pendidikan Sains Dan Matematika, 4(2)*, 115–121.
- Solichin, M. (2017). Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes,

- Interpretasi Hasil Tes dan Validitas Ramalan dalam Evaluasi Pendidikan. *Jurnal Manajemen Dan Pendidikan Islam*, 2(2), 192–213.
- Suryani, Y. E. (2017). Pemetaan Kualitas Empirik Soal Ujian Akhir Semester Pada Mata Pelajaran Bahasa Indonesia SMA di Kabupaten Klaten. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 21(2), 142–152.
- Susanto, H., Rinaldi, A., & Novalia. (2015). Analisis Validitas Reabilitas Tingkat Kesukaran dan Daya Beda pada Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika. *Jurnal Pendidikan Matematika*, 6(2), 203–216.
- Wibawa, E. A. (2019). Karakteristik Butir Soal Tes Ujian Akhir Semester Hukum Bisnis. *Jurnal Pendidikan Akutansi Indonesia*, 17(1), 87–96.
- Yusrizal. (2016). Analysis of Difficulty Level of Physics National Examination's Questions. *Jurnal Pendidikan IPA Indonesia*, 5(1), 140–149. <https://doi.org/10.15294/jpii.v5i1.5803>
- Yusuf, A. M. (2015). *Assesmen dan Evaluasi Pendidikan* (pertama). Jakarta: Prenadamedia Group.
- Zein, A., Fadillah, M., & Novianti, R. (2013). Hubungan Antara Validitas Butir , Reliabilitas , Tingkat Kesukaran Dan Daya Pembeda Soal Ujian Semester Genap Bidang Studi Biologi Kelas Xi Sma / Ma Negeri Di Kota Padang Tahun Pelajaran 2010 / 2011 *. *Semirata 2013 FMIPA Unila*, (2009), 39–47.