

ANALISIS BUTIR SOAL EVALUASI PEMBELAJARAN PKn KELAS IV SEKOLAH DASAR DENGAN ITEMAN DI KECAMATAN BANJARNEGARA

Pamujo¹, Dynar Rizqi Mara Romadhoni²
Program Studi PGSD, Fakultas Keguruan dan Ilmu Pendidikan
^{1,2}Universitas Muhammadiyah Purwokerto
Email: pamujopgsd61@gmail.com

Abstrak: Penelitian ini bertujuan untuk menganalisis soal ujian akhir mata pelajaran Pendidikan Kewarganegaraan Sekolah Dasar di Kabupaten Banjarnegara. Penelitian ini menggunakan pendekatan kuantitatif-deskriptif. Metode pengambilan sampel yang digunakan adalah random, stratified, dan proportional sampling. Dari sampling diperoleh 10 sekolah, 4 sekolah inti dan 6 sekolah imbas. Analisis kualitatif menggunakan review, data kuantitatif menggunakan program ITEMAN 3.00 dan metode manual. Hasil analisis kualitatif menunjukkan pertanyaan-pertanyaan telah memenuhi standar, namun perlu beberapa perbaikan pada materi, konstruksi, dan bahasa/budaya di beberapa bagian. Sedangkan hasil analisis kuantitatif diperoleh daya pembeda soal kurang baik (3%), cukup baik (11%), dan baik (86%). Tingkat kesulitan soal diperoleh hasil, terlalu sulit (8,5%), sedang (68,5%), dan terlalu mudah (23%). Keefektifan distraktor pada soal adalah efektif (57%), tidak efektif (43%). Reliabilitas butir soal secara keseluruhan mencapai 0,839, artinya soal-soal tersebut memiliki konsistensi yang tinggi. Dari hasil tersebut disimpulkan bahwa soal sudah memenuhi standar, namun masih perlu adanya perbaikan materi, konstruksi dan bahasa/budaya, perlu perbaikan pilihan jawaban sehingga distraktor cukup efektif. Selain itu, soal yang dibuat sudah sangat *reliable*. Dengan menggunakan perhitungan validitas item, diketahui ada satu pertanyaan yang tidak valid, yaitu item 8.

Kata Kunci: Analisis Butir; ITEMAN; Tingkat Kesukaran dan Daya Pembeda

Abstract: The study has an aim to analyze the multiple-choice questions in the final exam of Civic Education subject for the elementary school in Banjarnegara District. It applied a quantitative-descriptive approach. The sampling methods used here are random, stratified, and proportional sampling. From the sampling, it gave 10 schools, 4 core and 6 additional schools. The qualitative analysis applied a review and the quantitative one used the computer software of ITEMAN 3.00 and the manual method. The result of qualitative analysis shows the questions have agreed with the standard, but need some improvements of materials, construction, and language/culture in some parts. The quantitative analysis proves that difference impact of the questions is poor (3%), satisfactory (11%), and good (86%). The questions difficulty level is too difficult (8,5%), moderate (68,5%), and too easy (23%). The effectiveness of the distractor in the question is effective (57%), ineffective (43%). The reliability of the items in a whole reaches 0,839, meaning the questions have a high consistence. From those results, it can be concluded that the questions are matched to the standard, but they need some improvements of materials, construction and language/culture. The questions have a good difference impact, a moderate difficulty, and an effective distractor. In addition, the questions are also highly reliable. Using item validity calculation, it is known that there is one invalid question, i.e. the item 8.

Keywords: an Item Analysis; ITEMAN; Difficulty Level and Discrimination Power

PENDAHULUAN

Evaluasi merupakan bagian yang tak terpisahkan dari proses pembelajaran dan memiliki kedudukannya yang sentral. Ralph Tyler (Arikunto, 2017; dan Chellamani, 2013), menjelaskan bahwa evaluasi merupakan sebuah proses pengumpulan data untuk menentukan sejauh mana, dalam hal apa, dan bagian mana tujuan pendidikan sudah tercapai. Sementara Griffin dan Nix (Mardapi, 2017); Raymond, (2012); Lalaba, (2008); juga Bers, (2008), menekankan pentingnya penggunaan hasil evaluasi untuk perbaikan, sedangkan Harlen, Gip, Broadfoot dan Nutnally (Podlar, 2005) mengungkapkan penilaian sebagai umpan balik akan kemajuan yang telah dicapai. Oleh karena itu evaluasi merupakan proses sistematis tentang mengumpulkan, menafsirkan dan menganalisis informasi terkait dengan tercapainya tujuan. Tes merupakan sekumpulan pertanyaan yang harus dijawab, tugas yang harus dikerjakan atau pernyataan yang harus dipilih. Sedangkan Cronbach (Azwar, 2012) mendefinisikan *a systemic procedure for observing a person behavior and describing it with the aid of numerical scale or a category system*. Proses evaluasi bukan sekedar mengukur sejauh mana tujuan tercapai, tetapi digunakan untuk membuat keputusan (Stiggins & Cappisus, 2012). Jadi dalam evaluasi selalu ada objek yang akan dinilai, kriteria dan instrumen yang digunakan menilai yang sering dikenal dengan sebutan tes dalam bentuk butir soal. Evaluasi biasa dilakukan diakhir semester sekaligus merupakan salah satu alat untuk mengukur ketercapaian tujuan dan dalam kurun waktu tertentu. Dalam satu tahun Evaluasi Akhir Semester dilakukan dua kali, yaitu Semester Gasal dan Semester Genap.

Hal penting yang harus diperhatikan dalam proses penilaian adalah hasil penilaian menggambarkan kemampuan siswa yang sebenarnya. Suatu tes yang baik hasilnya dapat dipercaya dan mampu mengungkapkan apa yang diukur, (Sopiah, 2019). Terkait hal ini penyusunan instrumen tes harus berdasar kriteria validitas, reliabilitas, kompetensi, obyektif dan mendidik (Puskur, 2008; Kubiszyn & Borich, 2013, Arikunto, 2013); dan kemendiknas (Fitri, 2019). Selain itu instrumen tes harus melalui proses analisis butir untuk menjamin butir soal tersebut memenuhi syarat. *Item analyses a somewhat more important role in construct and predictive validation* (Nurmaly & Bernstein, 1994). Dengan demikian identifikasi dan analisis terhadap butir soal perlu dilaksanakan secara rutin dan berkelanjutan sebelum butir soal tersebut diujikan. Kegiatan ini diharapkan dapat

memperbaiki dan menyempunakan yang pada akhirnya memeningkatkan kualitas butir soal. Dengan cara ini diharapkan tes yang disusun dapat menjalankan fungsinya alat ukur.

Analisis soal bertujuan untuk mengidentifikasisoal yang baik dan soal yang jelek (Arikunto, 2013); dan juga mengetahui kualitas intrimen butir soal yang disusun Zhou, Huang & Tian, 2013). Sedangkan Kunandar (2015), menjelaskan analisis butir soal merupakan prosedur yang sistimatis yang memberikan informasi khusus tentang soal yang disusun. Salah satu cara menganalisis butir soal adalah dengan teori klasik. Menurut Qassem, *classical test theory introduces three concepts test score, true score, and error score*.

Proses analisis butir soal ini dilakukan dengan memperhatikan tingkat kesukaran, daya beda dan fungsi distraktor atau pengecoh (Mardapi, 2017) Tingkat kesukaran soal ditunjukkan dengan indeks kesukaran yang memiliki rentang antara 0,00-1,00. Semakin besar indeks yang diperoleh semakin mudah soal tersebut. Soal yang memiliki indeks pada interval 0,30-0,70 adalah soal yang baik (Kaplan & Sacuso, 2005; Sujana, 2011). Daya beda suatu soal ialah bagaimana kemampuan soal itu untuk membedakan siswa-siswa yang termasuk kelompok upper group/ kemampuan tinggi dengan siswa-siswa yang termasuk kelompok lower group/ kemampuan rendah. Indekks daya pembeda memiliki rentang antar 0,00-1,00. Butir soal yang baik adalah yang memiliki indeks sama dengan 0,3 atau lebih (Nunnally & Berstein, 2009; Azwar, 2012), butir soal dengan indeks cukup baik 0,20-0,29 (Crocker & Algina, 2008). Distrakor berfungsi baik apabila mampu menarik perhatian peserta tes yang tidak menguasai materi untuk memilihnya. Butir soal yang baik bila meiliki indeks lebih besar dari 5% (Mardapi, 2017; dan Ailen & Yen 1979; dan Mahjabeen (2017).

Analisis soal yang dilakukan berupa analisis kualitatif dan kuantitatif. Soal-soal yang akan dianalisis atau diteliti adalah soal Evaluasi Akhir Semester Genap atau sering disebut dengan Ulangan Kenaikan Kelas (UKK) yaitu tes yang dilaksanakan secara bersama-sama oleh setiap Sekolah Dasar di kecamatan Banjarnegara pada setiap akhir semester. Bentuk tes yang dianalisis adalah butir soal soal pilihan ganda.atau multiple choise. Soal ini merupakan salah satu bentuk soal dengan pertanyaan yang dilengkapi dengan pilihan jawaban yang terdiri dari kunci jawaban dandistraktor atau pengecoh (Burud, 2019). Pelaksanaan evaluasi akhir semester menjadi gambaran kualitas pesertadidik dan ketercaipan tujuan pembelajaran. Oleh karena itu pada pelaksanaanya dibutuhkan butir soal yang kredibel sehingga mampu memberikan informasi yang akurat.

Program yang digunakan untuk menganalisis secara kuantitatif adalah program IteMan versi 3,0. Program ini termasuk satu paket program dalam MicroCAT™ yang dikembangkan oleh Assessment System Corporation mulai tahun 1982 dan mengalami revisi pada 1984, 1986, 1988, dan 1993; mulai dari versi 2.00 sampai dengan versi 3.50. Pemilihan program iteman dikarenakan program tersebut tepat untuk menganalisis soal pilihan ganda. Soal yang akan dianalisis adalah mata pelajaran PKn kelas IV SD. Analisis kualitatif sering juga disebut sebagai validitas logis (*logical validity*) yaitu berupa penelaahan yang dimaksudkan untuk menganalisis soal ditinjau dari segi teknis, isi, dan editorial, (Depdiknas, 2008) Aspek yang diperhatikan di dalam penelaahan secara kualitatif ini adalah setiap soal ditelaah dari segi materi, konstruksi, bahasa/budaya, dan kunci jawaban/pedoman penskorannya.

Sesuai dengan perumusan masalah, maka penelitian ini bertujuan untuk mendeskripsikan karakteristik soal pilihan ganda Ulangan Kenaikan Kelas (UKK) mata pelajaran Pendidikan Kewarganegaraan kelas IV SD di Kecamatan Banjarnegara.

Dengan demikian artikel ini memberikan informasi kepada guru di Kecamatan Banjarnegara yaitu : a. apakah butir soal berfungsi tepat seperti dimaksudkan guru; b. umpan balik bagi guru tentang kesulitan belajar siswa; c. perbaikan butir soal.

METODE PENELITIAN

Penelitian ini menggunakan metode penelitian deskriptif kuantitatif. Diskriptif diartikan untuk memaparkan karakteristik butir soal PKn secara klasik. Subyek penelitian adalah butir soal mata pelajaran PKn siswa SD Kelas IV di Kecamatan Banjarnegara, Kabupaten Banjarnegara, yang diambil secara sampel dari 38 sekolah Dasar. Teknik pengambilan sampel berupa *purposive sampling* (sampel tujuan) yaitu pengambilan sampel berdasarkan pada karakteristik dan pertimbangan (Iskandar, 2010). Pemilihan teknik ini disebabkan populasinya heterogen dan berstrata. Jumlah sampel diambil secara proporsional yaitu SD inti (4 SD) dan SD imbas (6 SD). Teknik pengumpulan data awal menggunakan metode dokumentasi. Data atau dokumen soal diambil dari lembaga atau instansi sekolah yaitu berupa soal pilihan ganda Ulangan Kenaikan Kelas (UKK) termasuk kunci jawaban dan lembar jawab siswa. Analisis tingkat kesukaran daya beda dan efektifitas distraktor digunakan rumus tertentu yang sudah terstandar, sedangkan perhitungan validitas dan reliabilitas menggunakan program SPSS. Hasil perhitungan validitas dibandingkan dengan r tabel. Soal memiliki validitas baik jika r hitung lebih

besar dari r tabel, sedang hasil perhitungan dikonsultasikan dengan harga kritik reliabilitas. Kriteia harga kritik yang dipakai adalah 0,7 mengacu pendapat Kaplan (Widoyoko, 2019).

HASIL DAN PEMBAHASAN

Sebelum melakukan penelitian, lebih dahulu dilakukan kegiatan pra penelitian. Hal ini dilakukan untuk menggali permasalahan yang dihadapi guru ketika menyusun butir soal dan mengolah hasilnya. Informasi yang diperoleh menunjukkan guru mengalami kesulitan dalam menganalisis butir soal. Kondisi ini terjadi juga di daerah lain seperti diungkap Mukson, (2013). Oleh karena itu hasil penelitian diharapkan memberi masukan pada penyusunan butir soal diwaktu lain.

Kegiatan analisis pertama adalah melakukan analisis secara kualitatif terhadap 35 butir soal yang hasilnya dapat dilihat sebagai berikut : a, sesuai segi materi 88,57 %, b. sesuai segi konstruksisoal 67%, c. sesuai segi bahasa : 72%. Secara keseluruhan butir soal dari segi materi sudah baik artinya pembuatan soal sudah mengacu pada kompetensi dasar yang diturunkan menjadi kisi -kisi soal. Hal ini sesuai dengan langkah penulisan butir soal bahwa pembuatan soal didahului dengan analisis kompetensi dasar (Depdikbud, 2008).

Setelah melalui analisis kualitatif butir soal kemudian dianalisis secara kualitatif, hasilnya dipaparkan sebagai berikut :

a. **Validitas**, hasil perhitungan validitas dengan harga kritik 5% adalah 22 soal (68, 58%) valid dan 11 soal (31,42 %) soal tidak valid, secara rinci digambarkan pada tabel berikut

Tabel 1. Hasil Perhungan Validitas Soal

No	Validitas	No butir soal	Jumlah	%
1	≥ 0.3291 (Valid)	1,2,3,4,6,7,8,10,13,14,15,16,18,19,20,2 1,25,23,25,26,27,29,30,33,35	24	68,58%
2	< 0.3291 (tidak valid)	5,9,11,12,17,22, 24,28,31,32, 34	11	31,42%
			35	100%

Dari tabel tersebut diatas secara umum soal memiliki valitas yang baik, 68,58%, oleh karena itu guru harus melakukan revisi pada soal uyang belum valid. Berdasarkan penghitungan validitas setiap soal dengan jumlah peserta tes sebanyak 462 peserta, maka dapat diketahui bahwa hanya 1 soal yang hasilnya tidak valid yaitu soal nomor 8. Ada 19 soal yang diterima dan ada 16 soal yang perlu direvisi. Hali ini sangat penting karena

validitas merupakan syarat penting bagi soal yang akurat. Karena begitu pentingnya maka mengatakan validitas oleh Drost (2012), dikatakan “*validity is concerned with the meaningfulness of research components*”.

- b. Reliabilitas**, hasil rata rata perhitungan reliabilitas dari 10 Sekolah Dasar adalah 0,839 masuk pada kategori tinggi, dan bisa dilihat pada tabel berikut :

Tabel 2. Hasil Reliabilitas soal pilihan ganda

Analisis Reliabilitas	Nilai	Kriteria
	0,839	tinggi

Tabel 3. Reliabilitas seluruh Sekolah kelas IV SD di Kecamatan Banjarnegara

Sekolah	Analisis reliabilitas	
	Nilai	Kriteria
SD sampel 1	0,789	Tinggi
SD sampel 2	0,834	Sangat tinggi
SD sampel 3	0,664	Tinggi
SD sampel 4	0,774	Tinggi
SD sampel 5	0,963	Sangat tinggi
SD sampel 6	0,698	Tinggi
SD sampel 7	0,739	Tinggi
SD sampel 8	0,452	Cukup
SD sampel 9	0,750	Tinggi
SD sampel 10	0,833	Sangat tinggi

Dilihat dari tabel 2 diatas disimpulkan secara umum butir soal sudah memiliki reliabilitas yang tinggi dengan rata rata 0,839. Oleh karena itu butir soal sudah layak digunakan untuk mengukur kemampuan siswa.

- c. Daya beda**, perhitungan daya beda menghasilkan soal yang jelek berjumlah 1 (3%), cukup baik 3 (11%) dan baik 30 butir soal (86%), baik sekali/sangat jelek 0(0%) secara rinci disajikan dalam tabel berikut :

Tabel 4. Hasil Daya beda

Klasifikasi	Jml	%	No. Soal
<i>Poor</i> /jelek ($D < 0,20$)	1	3%	8
<i>Satisfactory</i> /cukup ($0,20 < D < 0,40$)	4	11%	10, 15, 28, 30
<i>Good</i> /baik ($0,40 < D < 0,70$)	30	86%	1, 2, 3, 4, 5, 6, 7, 9, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 29, 31, 32, 33, 34, 35
<i>Excellent</i> /baik sekali ($D > 0,70$)	-	-	-
Bertanda negatif (-)/jelek sekali	-	-	-

Tabel 5. Daya beda analisis setiap sekolah kelas IV SD di Kecamatan Banjarnegara

Sekolah	Excellent		Good		Satisfactory		Poor		Jelek sekali	
	Σ	%	Σ	%	Σ	%	Σ	%	Σ	%
SD sampel 1	5	14	19	54	6	17	1	3	4	12
SD sampel 2	8	23	21	60	5	14	1	3	0	0
SD sampel 3	0	0	14	40	12	34	9	26	0	0
SD sampel 4	3	9	19	54	9	26	3	9	1	3
SD sampel 5	25	71	4	11	1	3	0	0	5	14
SD sampel 6	2	6	16	46	11	31	6	17	0	0
SD sampel 7	1	3	24	69	5	14	4	11	1	3
SD sampel 8	1	3	6	17	3	9	4	11	21	60
SD sampel 9	4	11	17	49	8	23	1	3	5	14
SD sampel 10	18	51	5	14	2	6	0	0	10	29

Tabel 4 dan tabel 5 menunjukkan secara umum soal memiliki daya beda yang baik, tetapi di SD sampel 3 dan SD sampel 5 perlu mendapat perhatian karena daya beda tidak berfungsi dengan baik. Patut diduga disebabkan oleh karakteristik siswa, yang satu merupakan SD unggulan sedang yang lain adalah SD imbas dan terpencil.

d. Tingkat kesukaran, perhitungan hasil tingkat kesukaran diperoleh data sebagai berikut, terlalu sukar 3 soal (8,5%), sedang 24 soal (68,5%), mudah 8 soal(23%), secara rinci dapat dilihat pada tabel berikut :

Tabel 6. Tingkat kesukaran soal

Klasifikasi	Presentase
Terlalu Sukar ($P < 0,30$)	8,5%
Cukup/ sedang ($0,30 < P < 0,70$)	68,5%
Terlalu Mudah ($P > 0,70$)	23%

Adapun hasil analisis tiap sekolah adalah dapat dilihat pada tabel berikut :

Tabel 7. Tingkat kesukaran setiap sekolah soal kelas IV SD di Kecamatan Banjarnegara

Sekolah	Terlalu Sukar		Cukup / sedang		Mudah	
	Σ	%	Σ	%	Σ	%
SD sampel 1	6	17	21	60	8	23
SD sampel 2	2	6	15	43	18	51
SD sampel 3	14	40	19	54	2	6
SD sampel 4	8	23	20	57	7	20
SD sampel 5	1	3	1	3	33	94
SD sampel 6	9	26	25	71	1	3
SD sampel 7	3	9	20	57	12	34
SD sampel 8	6	17	4	11	25	71
SD sampel 9	8	23	21	60	6	17

SD sampel 10	9	26	22	63	4	11
--------------	---	----	----	----	---	----

Tabel 7. tingkat kesukaran menunjukkan soal sebagian besar pada kategori sedang atau cukup, sehingga sudah layak digunakan. Hal ini senada dengan pernyataan Amalia & Widayati (2012) dalam hasil penelitiannya bahwa soal yang baik adalah soal yang masuk kategori tingkat kesulitan sedang, yakni tidak terlalu mudah atau tidak terlalu sulit.

- e. **Distraktor atau pengecoh**, hasil perhitungan efektifitas pengecoh adalah berfungsi 20 soal (57%) dan tidak berfungsi 15 soal (43%), secara rinci efektifitas pengecoh disajikan pada tabel berikut :

Tabel 7. Efektifitas *distractor* soal kelas IV SD di Kecamatan Banjarnegara

Klasifikasi	Jumlah	Presentase	Nomor Soal
Berfungsi	20	57%	3, 4, 5, 6, 9, 12, 14, 15, 16, 18, 22, 23, 25, 26, 29, 30, 31, 32, 33, 35
Tidak berfungsi	15	43%	1, 2, 7, 8, 10, 11, 13, 17, 19, 20, 21, 24, 27, 28, 34

Tabel 8. Efektifitas *distractor* analisis setiap sekolah kelas IV SD di Kecamatan Banjarnegara

Sekolah	Berfungsi		Tidak berfungsi	
	Σ	%	Σ	%
SD sampel 1	16	46	19	54
SD sampel 2	9	26	26	74
SD sampel 3	31	89	4	11
SD sampel 4	14	40	21	60
SD sampel 5	1	3	34	97
SD sampel 6	23	66	12	34
SD sampel 7	14	40	21	60
SD sampel 8	5	14	30	86
SD sampel 9	15	43	20	57
SD sampel 10	18	51	17	49

Dari tabel 7 dan 8 tersebut diatas maka secara umum soal belum memiliki distraktor fungsi efektifitas. Hal ini dapat dilihat dari hasil perhitungan yaitu 43% pilihan jawaban tidak efektif sebagai pengecoh. Dugaan sementara faktor kesulitan guru dalam mencari pilihan jawaban yang homogen pada level HOST. Disamping itu dari hasil penelitian Susilowati, (2014); Rahmawati (2018), dipaparkan bahwa selama ini guru mengalami kesulitan dalam analisis butir soal, sehingga kegiatan analisis tidak pernah dilakukan. Oleh karena itu informasi ini bisa dijadikan sebagai pemicu untuk

dilakukannya analisis soal secara berkelanjutan agar dapat menghasilkan butir soal yang berkualitas.

KESIMPULAN DAN SARAN

Melihat pada hasil penelitian dan pembahasan analisis soal UKK PKn kelas IV SD di Kecamatan Banjarnegara, maka dapat disimpulkan bahwa soal evaluasi semester PKn kelas IV SD di Kecamatan Banjarnegara bahwa hasil analisis kualitatif menunjukkan pertanyaan-pertanyaan telah memenuhi standar, namun perlu beberapa perbaikan pada materi, konstruksi, dan bahasa/budaya di beberapa bagian. Sedangkan analisis kuantitatif diperoleh hasil 68,58% soal valid, dan memiliki reliabilitas 0,839 masuk kategori tinggi. Perhitungan daya pembeda diperoleh hasil soal kurang baik (3%), cukup baik (11%), dan baik (86%). Tingkat kesulitan soal diperoleh hasil, terlalu sulit (8,5%), sedang (68,5%), dan terlalu mudah (23%). Keefektifan distraktor pada soal adalah efektif (57%), tidak efektif (43%). Dari hasil tersebut disimpulkan bahwa soal sudah memenuhi standar, namun masih perlu adanya perbaikan materi, konstruksi dan bahasa/budaya, perlu perbaikan pilihan jawaban sehingga distraktor cukup efektif.

DAFTAR PUSTAKA

- Allen & Yen.1979. *Introduction to Measurement Theory*. Belmont, California: Wadsworth, Inc.
- Arikunto, S. (2016). *Dasar-dasar evaluasi pendidikan* (2nd ed.). Jakarta: Bumi Aksara.
- Azwar, S. (2012). *Tes prestasi fungsi pengembangan pengukuran prestasi belajar*. Yogyakarta: Pustaka Pelajar.
- Bers, T.H. 2008. The Role of Institutional Assessment in Assessing Student Learning Outcomes. *New Di-rections for Higher Education*, 141: 31-39.
- Crocker, L. & Algina, J. 2008. *Introduction to Classical and Modern Test Theory*. Ohio: Cengage Learning
- Chellamani, C. &. (2013). Analisis of Test Items on Difficulty Level and Discrimination Index The Test For Reseach in Education. *International Journal of Social Science & Interdisciplinary Research*, 189-193.
- Drost, E.A. Validity and Reliability in Social Science Research. *Education Research and Perspectives*, Vol.38, No.1, 2012.
- Fitri Alfarisa, Ima Ni'mah Chudari, & Firman Robiansyah. Analisis Butir Soal IPS Kelas V Sekolah Dasar Menggunakan Software ITEMAN. *EduBasic Journal: Jurnal Pendidikan Dasar*, 1(2), (2019): 100-106

- Ismail, Burud, K. N. (2019). Impact of distractors in item analysis of multiple choice questions. *International Journal of Research in Medical Sciences*, 1136-1139.
- Kaplan, R.M. & Saccuzzo, D.P. 2005. *Psychological Testing: Principles, Applications, and Issues* (6th edition). Belmont, CA: Thomson Wadsworth.
- Kubiszyn, T. & Borich, G.D. 2013. *Educational Testing and Measurement: Classroom Application and Practice (10th edition)*. Hoboken, NJ: John Wiley & Sons, Inc.
- Kunandar. (2015). *Penilaian Autentik*. Jakarta: PT Raja Grafindo.
- Junaidi, Lababa. Analisis butir soal dengan teori tes klasik: Sebuah pengantar. *Jurnal Iqra' Volume 5 Januari - Juni 2008*.
- Mardapi, D. (2012). *Pengukuran, penilaian dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S., & Rizvi, M. (2017). Difficulty index, discrimination index and distractor efficiency in multiple choice questions. *Annals of Pakistan Institute of Medical Sciences*, 13(4), 310–315
- Muhson, A. (2015). Kelayakan anbuso sebagai software analisis soal bagi guru. *Jurnal Kependidikan*, 210.
- Nunnally, J.C. & Bernstein, I.H. 1994. *Psychometric Theory (Third Edition)*. New York: McGraw-Hill, Inc.
- Pollard and J. Collins, *Reflective teaching*. A&C Black, 2005.
- Pusat Penilaian Pendidikan. 2005. *Panduan Penulisan Soal Pilihan Ganda*. Jakarta: Balitbang-Depdikbud
- Puskur 2008. *Model Penilaian Kelas Kurikulum Berbasis Kompetensi*. Jakarta: Balitbang Depdiknas.
- Qasem, M.A.N. A Comparative Study of Classical Theory (Ct) and Item Response Theory (Irt) In Relation To Various Approaches of Evaluating the Validity and Reliability of Research Tools. *IOSR Journal of Research & Method in Education (IOSR-JRME) Volume 3, Issue 5, pp 77-81, 2013.*
- Rahmawati. (2012). *Analisis Butir Soal Matematika Pada UKA PLPG LPTK Fakultas Tarbiyah IAIN Antasari Banjarmasin*.
- Raymond, J.E., Homer, C.S.E., Smith, R. & Gray, J.E.2012. Learning through Authentic Assessment: An Evaluation of a New Development in the Un-dergraduate Midwifery Curriculum. *Nurse Edu-cation in Practice*, 30: 1-6.
- Stiggins, R.J. & Chappuis, J. 2012. *An Introduction to Stu-dent Involved Assessment for Learning* (Sixth edi-tion). Boston: Pearson assessment training institute.

Sudjana, N. 2011. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya.

Susilowati. 2013. *Kurikulum 2013, 87 Persen Guru Kesulitan Cara Penilaian*, (Online), (<http://unnes.ac.id>), diakses 18 November 2013.

Zhou,Q., Huang, Q., & Tian, H. (2013). *Developing students ' critical thinking skills by task-based learning in chemistry experiment teaching*, 4(12), 40–45.