



Terakreditasi S4 – SK No. 36/E/KPT/2019

Penerbit: Program Studi Pendidikan Geografi, FISE, Universitas Hamzanwadi



ANALISIS KUALITAS BUTIR SOAL TENGAH SEMESTER GEOGRAFI SMA ISLAM AL AZHAR 14 KOTA SEMARANG

Moch. Arifien¹, Pradika Adi Wijayanto^{2*}

^{1,2}Jurusan Geografi, Fakultas Ilmu Sosial, Universitas Negeri Semarang, Semarang, Indonesia

*Email Koresponden: pradikawijaya@mail.unnes.ac.id

Diterima: 22-06-2021, Revisi: 23-10-2021, Disetujui: 26-12-2021

©2021 Program Studi Pendidikan Geografi, FISE, Universitas Hamzanwadi

Abstrak Penelitian ini bertujuan untuk menganalisis kualitas butir soal geografi agar layak digunakan dalam proses pembelajaran. Data yang digunakan berupa hasil pengerjaan soal pilihan ganda ulangan tengah semester geografi kelas XI IS 1 & 2 SMA Islam Al Azhar 14 Kota Semarang. Data diambil secara *purposive sampling* dan angket tertutup untuk mengetahui persepsi validator tentang kualitas soal yang dibuat. Data dianalisis dengan teknik deskripsi persentase. Berdasarkan hasil penelitian dapat disimpulkan bahwa: 1) aspek materi/substansi terdapat revisi pada indikator pertama dan kedua; 2) aspek konstruksi terdapat revisi pada indikator pertama, kedua, keempat, dan kelima; 3) aspek bahasa terdapat revisi pada indikator pertama, kedua, ketiga, dan keempat; 4) Uji validitas terdapat 18 soal dapat dipertimbangkan untuk digunakan; 5) Uji reliabilitas butir mempunyai tingkat kepercayaan sangat tinggi; 6) Uji taraf kesukaran terdapat 23 soal dapat dipertimbangkan untuk digunakan; 7) Uji daya beda soal terdapat 37 soal dapat dipertimbangkan untuk digunakan; dan 8) Uji efektivitas pengecoh/*distractor* terdapat 14 soal dapat dipertimbangkan untuk digunakan. Berdasarkan uji secara aspek kualitatif dan aspek kuantitatif dapat disimpulkan bahwa instrumen tes yang dapat digunakan adalah soal nomor 1, 3, 4, 5, 6, 9, 11, 13, 14, 17, 20, 21, 22, 24, 26, 27, 28 dan 37 karena lebih berkualitas.

Kata Kunci: analisis, kualitas butir soal, geografi.

Abstract This study aims to analyze the quality of geography questions so that they are suitable for use in the learning process. The data used is the result of working on multiple choice questions in the mid-semester geography test for class XI IS 1 & 2 SMA Islam Al Azhar 14 Semarang City. The data were taken by *purposive sampling* and closed questionnaires to determine the validator's perception of the quality of the questions made. The data were analyzed by using percentage description technique. Based on the results of the study, it can be concluded that: 1) the material/substance aspect has revisions to the first and second indicators; 2) the construction aspect contains revisions to the first, second, fourth, and fifth indicators; 3) the language aspect has revisions to the first, second, third, and fourth indicators; 4) The validity test contains 18 questions that can be considered for use; 5) The item reliability test has a very high level of confidence; 6) The difficulty level test contains 23 questions that can be considered for use; 7) The test of differentiating power of questions, there are 37 questions that can be considered for use; and 8) The distractor effectiveness test contains 14 questions that can be considered for use. Based on the qualitative and quantitative aspects of the test, it can be concluded that the test instruments that can be used are questions number 1, 3, 4, 5, 6, 9, 11, 13, 14, 17, 20, 21, 22, 24, 26, 27, 28 and 37 because they are of higher quality.

Keywords: analysis, item quality, geography

PENDAHULUAN

Evaluasi pada hasil atau produk sangat diperlukan sebagai dasar untuk meningkatkan kualitas. Salah satu hasil atau produk dalam pembelajaran yang dapat digunakan dalam kegiatan evaluasi adalah instrumen tes. Instrumen tes seringkali digunakan oleh guru sehingga instrumen tes tidak bisa dipisahkan dalam aktivitas pembelajaran untuk mengukur kualitas siswa dari proses pembelajaran yang dilakukan guru (Wijayanto, 2017). Fungsi tes dapat bermanfaat untuk mengukur kemampuan siswa secara kognitif sehingga dalam proses pembuatan dan pengembangan harus memperhatikan prosedur atau pedoman ciri instrumen yang baik dan berkualitas. Hal ini sejalan dengan pendapat Rahayu et al.,

(2014) yang menyatakan bahwa penyusunan soal pilihan ganda untuk ujian harus sesuai dengan pedoman pembuatan soal yang baik agar soal tersebut benar-benar mampu mengukur tingkat pengetahuan siswa sesuai standar kompetensi yang telah ditetapkan. Instrumen tes tersebut haruslah berkualitas baik dan handal sehingga menurut Aliman (2020) instrumen yang andal dibutuhkan agar mampu mengukur secara detail kemampuan siswa. Mewujudkan instrumen yang handal dan berkualitas menurut Sudijono (2015) bahwa tes dikatakan berkualitas apabila memperhatikan aspek validitas, reliabilitas, daya pembeda, dan taraf kesukaran. Selain itu Arikunto (2018) juga menguatkan bahwa instrumen tes yang baik dapat diketahui dari segi validitas, reliabilitas, objektivitas, praktibilitas dan ekonomis.

Untuk mewujudkan soal yang berkualitas, guru geografi masih kesulitan terutama dalam melakukan evaluasi terhadap instrumen tes karena dalam proses evaluasi khusus analisis kualitas butir soal memiliki prosedur yang sistematis dan ilmiah. Proses tersebut juga dikuatkan oleh Kunandar, (2014) bahwa analisis butir soal adalah prosedur sistematis yang berguna untuk memberikan informasi-informasi sangat khusus terhadap butir soal yang disusun. Berdasarkan wawancara dengan guru kelas XI IS SMA Islam Al Azhar 14 Kota Semarang, guru mata pelajaran Geografi jarang melakukan analisis butir soal, sehingga kualitas tes belum diketahui apakah sudah baik atau sebaliknya. Analisis butir soal tersebut dianggap cukup rumit karena guru harus melibatkan banyak siswa untuk dijadikan sebagai subjek penelitian sehingga membutuhkan banyak waktu. Selain itu dalam apabila dilakukan analisis soal secara sejawat (*peer assessment*) jumlah guru terbatas dan kebanyakan masih belum memahami secara teoritis maupun implementasinya dalam proses analisis dan olahdata. Secara analisis butir soal kualitatif guru masih kesulitan mencari indikator penilaian instrumen, menyusun instrumen untuk mevalidasi soal hingga mencari validator yang mumpuni dibidangnya agar kualitas butir soal menjadi lebih baik sehingga analisis ini sering terlupakan. Berbeda dengan analisis butir soal secara kuantitatif Guru kelas XI IS SMA Islam Al Azhar 14 Kota Semarang sudah mampu menggunakan *software-software* bersifat otomatis untuk bisa menganalisis butir soal, namun dengan adanya kemudahan ini guru seringkali kurang mantap dalam menggunakan teknik atau cara menganalisis butir soal. Dibuktikan dengan guru tidak paham dalam menggunakan jenis rumus-rumus tertentu sehingga guru hanya membaca hasil akhir kualitas butir soal dari *software-software* tersebut tanpa mengetahui bagaimana proses analisis soal dilakukan.

Mengetahui hal tersebut maka sangat diperlukan adanya proses analisis butir soal tengah semester di sekolah tersebut karena manfaatnya yang besar terlebih pertimbangan apabila soal digunakan sebelum diujikan akan menghasilkan data yang belum akurat tentang hasil belajar dan kualitas soalnya. Mendukung pentingnya hal tersebut menurut Asmawi & Nasution dalam Widoyoko (2016) menjelaskan bahwa analisis butir soal diperlukan karena bermanfaat untuk: 1) mengetahui kekuatan dan kelemahan butir tes agar dibuat kriteria mana yang akan digunakan atau dapat direvisi, 2) bermanfaat bagi guru untuk menyusun perangkat soal sesuai kebutuhan berdasarkan informasi butir soal yang spesifik, 3) berguna bagi guru agar dapat mengetahui masalah yang terdapat pada butir soal setelah dianalisis, 4) sebagai dasar guru untuk menilai dan menyimpan soal tersebut dalam bentuk bank soal. Nur & Palobo (2018) juga menguatkan bahwa analisis butir soal merupakan kegiatan guru yang sangat bermanfaat untuk menafsirkan hasil belajar peserta didik dan mengetahui kualitas soal. Nurjanah & Marlianingsih (2015) berpendapat bahwa kegiatan analisis butir soal atau analisis item sangat penting untuk memperoleh kualitas soal yang baik, dan memperoleh gambaran tentang prestasi siswa yang sebenarnya. Solichin (2017) menjelaskan bahwa analisis soal sesungguhnya bermanfaat untuk memperoleh informasi mengenai kejelekan sebuah soal dan sebagai dasar untuk mengadakan perbaikan. Berdasarkan pendapat tujuan penelitian ini adalah mengevaluasi kualitas soal geografi kelas XI IS SMA Islam Al Azhar 14 Kota Semarang. Hal ini karena analisis kualitas soal sangat penting dilakukan agar guru mendapatkan manfaat untuk mengetahui hasil belajar siswa yang lebih akurat, mengetahui kualitas soal yang dibuat, dapat dijadikan sebagai dasar pertimbangan untuk mengadakan perbaikan soal apabila diperlukan dan guru dapat memilih soal yang akan disimpan dalam bank soal agar dapat dimanfaatkan di waktu-waktu yang akan datang.

METODE PENELITIAN

Penelitian ini adalah penelitian evaluatif yang menggunakan pendekatan evaluasi yang berorientasi pada pengguna. Pendekatan ini bertujuan untuk mengevaluasi suatu hasil atau produk agar dapat ditingkatkan kualitasnya. Hasil atau produk tersebut berupa instrumen tes soal ulangan tengah semester siswa geografi kelas XI di SMA Islam Al Azhar 14 Semarang berbentuk pilihan ganda. Soal tersebut dievaluasi pada segi aspek kualitatif (materi/substansi, konstruksi, dan bahasa) dengan menggunakan indikator instrumen menurut Subali (2014) serta secara kuantitatif (validitas, reliabilitas, taraf kesukaran, daya beda, dan analisis pengecoh/*distractor*). Sampel penelitian ini menggunakan teknik *purposive sampling* karena guru geografi hanya mengajar dua dari tiga kelas dan rata-rata nilai ulangan dua kelas tersebut baik sehingga sampel yang digunakan adalah siswa kelas XI IS 1 dan XI IS 2 tahun ajaran 2018/2019.

Teknik pengumpulan data dilakukan dengan menggunakan metode dokumentasi, angket tertutup dan wawancara. Dokumentasi dilakukan dengan mengumpulkan data-data berupa dokumen soal ujian, silabus, Rencana Pelaksanaan Pembelajaran (RPP), daftar nama siswa, semua jawaban siswa dan kunci jawaban soal di SMA Islam Al Azhar 14. angket tertutup digunakan untuk menelaah kualitas soal secara kualitatif. Wawancara dilakukan untuk mendapatkan informasi tentang pelaksanaan tes mata pelajaran geografi dan informasi tentang analisis soal yang sudah dilakukan oleh guru mata pelajaran geografi.

Teknik analisis data yang digunakan untuk menunjukkan kualitas instrumen tes secara kualitatif yaitu dengan melakukan pertimbangan (*judgement*) dari validator kemudian dihitung persentasenya. Sedangkan analisis data kuantitatif menggunakan rumus korelasi *product moment* angka kasar untuk menghitung validitas, menghitung reliabilitas tes menggunakan rumus Hoyt, penggunaan rumus tingkat kesukaran, menghitung masing-masing tingkat kesukaran soal perbutir dan menghitung daya beda soal menggunakan rumus dari Arikunto (2018).

TEMUAN DAN PEMBAHASAN

Aspek Materi/Substansi

Analisis pada aspek materi/substansi termasuk dalam analisis butir soal secara kualitatif yang bertujuan menunjukkan apakah soal dan pilihan jawaban yang dibuat sudah relevan dengan kurikulum kompetensi dasar, tujuan, indikator berdasarkan silabus dan Rencana Pelaksanaan Pembelajaran (RPP). Analisis tersebut menggunakan 4 indikator, sebagaimana ditampilkan pada Tabel 1. (Subali, 2014).

Tabel 1. Hasil Analisis Aspek Materi/Substansi

No	Indikator/Aspek Materi(Substansi)	Persentase Jumlah Soal	
		Sesuai (%)	Tidak Sesuai (%)
1	Item sesuai dengan indikator	92,5	7,5
2	Pernyataan/pertanyaan dan jawaban kunci terumuskan dengan benar	85	15
3	Materi/substansi yang ditanyakan sesuai dengan tujuan pengukuran (untuk tujuan pengukuran hasil belajar, tujuan pengukuran untuk seleksi, atau tujuan pengukuran untuk konfirmatori/mengukur status)	100	0
4	Materi/substansi yang ditanyakan sesuai dengan jenjang, jenis sekolah, dan tingkat kelas	100	0
Jumlah		94,4	5,6

Sumber: Hasil olahan data primer, 2019.

Berdasarkan Tabel 1 dapat diketahui bahwa pada indikator pertama butir soal yang sesuai sebesar 92,5% & 7,5% tidak sesuai. pada indikator kedua butir soal yang sesuai sebesar 85% & 15% tidak sesuai, pada indikator ketiga butir soal yang sesuai sebesar 100% dan pada indikator keempat butir soal yang sesuai sebesar 100% sehingga total persentase kesesuaian butir soal dengan aspek materi/substansi adalah 94,4 % dan tidak sesuai sebesar 5,6%.

Aspek Konstruksi

Analisis konstruksi merupakan analisis butir soal secara kualitatif dengan cara mencocokkan pembuatan tiap butir soal dengan memperhatikan singkat, jelas dan tegas materinya, tidak menimbulkan

multitafsir dan harus mengandung satu persoalan untuk yang ditanyakan dalam setiap nomor soal. Aspek kontruksi dapat dianalisis berdasarkan 6 indikator menurut Subali (2014). Hasil analisis aspek kontruksi ditunjukkan pada Tabel 2:

Tabel 2. Hasil Analisis Aspek Konstruksi

No	Indikator/Aspek Konstruksi	Persentase Jumlah Soal	
		Sesuai (%)	Tidak Sesuai (%)
1	Pokok soal harus dirumuskan dengan jelas dan tegas	80	20
2	Semua pernyataan yang dijadikan pilihan dirumuskan secara lugas sehingga jelas dapat dibedakan apakah setiap tersebut benar atau salah	85	15
3	Pokok soal tidak mengarah ke jawaban yang benar, misalnya menggunakan kata salah satu, tiga hal atau yang sejenisnya	100	0
4	Seluruh pernyataan pada alternatif jawaban harus homogen, baik dalam bentuk bunyi kata ataupun istilah	92,5	7,5
5	Gambar, grafik, tabel ataupun wacana benar-benar berfungsi	90	10
6	Antar item tidak bergantung satu sama lain	100	0
Jumlah		91,25	8,75

Sumber: Hasil olahan data primer, 2019.

Berdasarkan Tabel 2 tersebut diketahui bahwa pada indikator pertama butir soal yang sesuai sebesar 80% & 20% tidak sesuai. Persentase indikator kedua sebesar 85% sesuai & 15% tidak sesuai. Indikator ketiga persentase butir soal yang sesuai indikator sebanyak 100%. Persentase indikator keempat yang sesuai adalah 92,5% dan 7,5% tidak sesuai. Indikator kelima sebesar 90% butir soal yang sesuai dan 10% tidak sesuai. Indikator keenam mempunyai persentase sebanyak 100% sesuai, jadi total persentase kesesuaian butir soal dengan aspek konstruksi adalah 91,25% dan tidak sesuai sebesar 8,75%.

Aspek Bahasa

Analisis bahasa dalam penelitian ini dilakukan dengan cara menelaah penggunaan bahasa sesuai dengan aturan Ejaan Yang Disempurnakan (EYD) berdasarkan indikator penilaian bahasa. Aspek bahasa dapat dianalisis berdasarkan 5 indikator (Subali, 2014). Hasil analisis aspek bahasa ditunjukkan Tabel 3 seperti di bawah ini:

Tabel 3. Hasil Analisis Aspek Bahasa

No	Indikator/Aspek Bahasa	Persentase Jumlah Soal	
		Sesuai (%)	Tidak Sesuai (%)
1	Rumusan kalimat soal komunikatif	60	40
2	Kalimat menggunakan bahasa yang baik dan benar, sesuai dengan jenis bahasanya	57,5	42,5
3	Rumusan kalimat tidak menimbulkan penafsiran ganda atau salah pengertian	97,5	2,5
4	Menggunakan bahasa/kata yang umum (bukan bahasa lokal atau bahasa serapan baru yang belum dikenal oleh seluruh testi)	97,5	2,5
5	Rumusan kalimat soal komunikatif	100	0
Jumlah		82,5	17,5

Sumber: Hasil olahan data primer, 2019.

Berdasarkan Tabel 3 tersebut persentase butir soal sesuai dengan indikator pertama sebesar 60% & 40% tidak sesuai. Indikator kedua sebesar 57,5% butir soal sesuai & 42,5% tidak sesuai. Persentase butir soal sesuai dengan indikator ketiga sebesar 97,5% soal sesuai & 2,5% tidak sesuai. Indikator keempat sebesar 97,5% butir soal sesuai & 2,5 % tidak sesuai. Persentase butir soal sesuai dengan indikator kelima sebesar 100%, jadi total persentase kesesuaian butir soal dengan aspek bahasa adalah 82,5% dan tidak sesuai sebesar 17,5%.

Uji Validitas

Uji validitas digunakan untuk mengetahui tingkat kevalidan atau mengetahui valid tidaknya item pada soal pilihan ganda tersebut. Uji validitas menggunakan korelasi product moment angka kasar berdasarkan Arikunto (2018).

Tabel 4. Hasil Uji Validitas

No	Nomor Soal	Keterangan	Jumlah Soal	Persentase
1	0	Validitas sangat tinggi	0	0 %
2	5,6,20	Validitas tinggi	3	7,5 %
3	1,3,4,9,11,13,14,17,21,22,24,26,27,28,37	Validitas cukup	15	37,5%
4	7,8,10,12,18,23,30,31,32,36	Validitas rendah	10	25 %
5	2,15,16,19,25,29, 33,34,35,38,39,40	Validitas sangat rendah	12	30 %
Jumlah soal			40	100%

Sumber : Hasil olahan data primer, 2019.

Berdasarkan tabel diatas maka 40 soal yang diujikan di kelas XI IS 1 dan X IS 2 diperoleh hasil yang berbeda. Soal yang diuji menunjukkan kriteria soal yang mempunyai validitas sangat tinggi sebanyak 0 atau 0%, Validitas tinggi sebanyak 3 soal atau 7,5% , Validitas cukup 15 soal atau 37,5%, Validitas rendah sebanyak 10 soal atau 25% dan validitas sangat rendah sebanyak 12 soal atau 30%. Dominasi validitas terdapat pada kriteria cukup sebesar 30% dengan jumlah 15 soal.

Uji Reliabilitas

Tujuan penggunaan uji reliabilitas agar instrumen tes dapat diketahui tingkat kepercayaannya, jadi semakin tinggi skor maka tingkat reliabilitasnya semakin baik. Perhitungan reliabilitas tes menggunakan rumus Hoyt berdasarkan Arikunto (2018). Hasil analisis uji reliabilitas diperoleh hasil perhitungan 0,929. Hasil tersebut jika dikonsultasikan berdasarkan Guilford (1956) maka dapat disimpulkan tingkat reliabilitas dari soal tersebut adalah sangat tinggi.

Uji Taraf Kesukaran

Uji berfungsi untuk mengetahui kriteria kesulitan dari soal yang diujikan sehingga dapat diketahui tingkat kesulitan apakah soal tersebut mudah, sedang maupun sulit. Perhitungan uji taraf kesukaran menggunakan rumus berdasarkan Arikunto (2018).

Tabel 5. Hasil Uji Taraf Kesukaran

No	Nomor Soal	Keterangan	Jumlah Soal	Persentase
1	10,18,19,25,29,30,34	Sukar	7	17,5 %
2	2,5,6,8,9,11,12,13,14,16,17,20,21, 22,23,26,27,31,33,36,37,39,40	Sedang	23	57,5 %,
3	1,3,4,7,15,24,28,32,35,38	Mudah	10	25 %
Jumlah soal			40	100%

Sumber: Hasil olahan data primer, 2019.

Berdasarkan Tabel 5 dapat disimpulkan bahwa terdapat 7 soal bertaraf sukar dengan persentase 17,5%, 23 soal bertaraf sedang dengan persentase 57,5% dan 10 soal bertaraf mudah dengan persentase 25%. Soal yang diujikan lebih didominasi oleh taraf kesukaran sedang, hal tersebut bisa terjadi karena soal yang dibuat berdasarkan taksonomi Bloom pada tingkat tertentu. Dominasi taraf kesukaran terdapat pada kriteria sedang sebesar 57,5% dengan jumlah 23 soal.

Uji Daya Pembeda

Uji daya pembeda bermanfaat untuk mengetahui perbedaan tingkat kemampuan siswa dari hasil mengerjakan soal yang diberikan. Uji tersebut juga bertujuan untuk mengetahui kemampuan siswa melalui pengelompokan nilai tiap siswa dari kelompok atas dan kelompok bawah sehingga diketahui siapa saja yang mendapat nilai baik dan buruk. Perhitungan dan interpretasi data dilakukan berdasarkan rumus dan kriteria dari Arikunto (2018).

Tabel 6. Hasil Uji Daya Pembeda

No	Nomor Soal	Keterangan	Jumlah Soal	Persentase
1	1,3,6,8,9,19,20,21, 23,26,27	Baik sekali	11	27,5 %
2	4,5,11,12,14,15,17,18,22,24,31,37	Baik	12	30 %
3	2,7,10,13,16,28,29,30,32,33,34,36,39,40	Cukup	14	35 %
4	35,38	Jelek	2	5 %
5	25	Negatif	1	2,5 %
Jumlah soal			40	100%

Sumber: Hasil olahan data primer, 2019.

Berdasarkan Tabel 6 menunjukkan bahwa masing-masing soal memiliki daya beda yang tidak seragam. Terdapat 11 soal (daya beda baik sekali) dengan persentase 27,5%, 12 soal (daya beda baik) dengan persentase 30%, 14 soal (daya beda cukup) dengan persentase 35%, 2 (daya beda jelek) dengan persentase 5% dan 1 soal (daya beda negatif) dengan persentase 2,5%. Dominasi uji daya pembeda terdapat pada kriteria cukup sebesar 35% dengan jumlah 14 soal.

Uji Efektivitas Pengecoh (*Distractor*)

Uji ini bertujuan untuk mengetahui seberapa efektifkah butir soal yang dibuat dapat mengecoh siswa untuk memilih opsi selain jawaban yang benar. Pengecoh dapat dianggap efektif jika dijawab oleh minimal 5% peserta tes. Uji efektivitas tersebut menggunakan software AnBuso dengan menggunakan kriteria berdasarkan (Muhson, 2017).

Tabel 7. Hasil Uji Efektivitas Pengecoh (*Distractor*)

No	Nomor Soal	Keterangan	Jumlah Soal	Persentase
1	8,13,17,20,21, 31,37	Baik	7	17,5 %
2	4,5,6,7,9,14,22,23,26,27,35,38	Revisi Pengecoh	12	30%
3	1,3,10,11,12,18,24	Cukup Baik	7	17,5 %
4	2,15,16,19,25,28,29,30,32,33,34,36,39,40	Tidak Baik	14	35 %
Jumlah soal			40	100%

Sumber: Hasil olahan data primer, 2019.

Berdasarkan Tabel 7 menunjukkan bahwa masing-masing soal memiliki efektivitas pengecoh yang tidak seragam. Terdapat 7 soal (Baik) dengan persentase 17,5%, 12 soal (Revisi pengecoh) dengan persentase 30%, 7 soal (Cukup baik) dengan persentase 17,5% dan 14 (Tidak baik) dengan persentase 35%. Dominasi efektivitas pengecoh terdapat pada kriteria tidak baik sebesar 35% dengan jumlah 14 soal.

Kelayakan Soal Berdasarkan Analisis Kualitatif dan Kuantitatif Butir Soal

Analisis butir soal secara kualitatif dilakukan dengan mempertimbangkan aspek materi/substansi, kontruksi dan bahasa. Sedangkan secara kuantitatif dilakukan dengan menggunakan uji validitas, reliabilitas, uji taraf kesukaran, uji daya beda dan uji efektivitas pengecoh (*distractor*) yang telah dilakukan sebelumnya dan merupakan langkah untuk menentukan kelayakan soal. Analisis materi/substansi ini juga disebut sebagai uji validitas kurikuler atau isi dengan dikuatkan secara teori oleh Arifin (2013) yang menjelaskan bahwa validitas kurikuler berkenaan dengan pertanyaan apakah materi tes relevan dengan kurikulum yang sudah ditentukan. Arikunto (2018) validitas isi digunakan untuk mengukur tujuan khusus tertentu yang sejajar dengan materi atau isi pelajaran yang diberikan dan Widoyoko (2016) tes dikatakan mempunyai validitas isi apabila dapat mengukur kompetensi yang dikembangkan berdasarkan indikator dan materi pembelajarannya. Hal di atas juga dikuatkan penelitian oleh Wijayanto, Allifah & Amirrudin (2016) bahwa pengujian validitas kurikuler (isi) soal instrumen tes ini dilakukan dengan mencocokkan materi soal dengan kurikulum, kompetensi dasar dan indikator pembelajaran yang akan dicapai pada silabus dan Rencana Pelaksanaan Pembelajaran (RPP). Berdasarkan analisis yang telah dilakukan terdapat beberapa soal yang perlu direvisi. Butir-butir soal yang perlu direvisi dikarenakan tidak memenuhi indikator pertama dan kedua. Sedangkan pada indikator ketiga dan keempat semua butir soal sudah memenuhi.

Indikator pertama terdapat 3 butir soal yang perlu direvisi, yaitu butir soal nomor 7, nomor 11, dan nomor 14. Soal nomor 7 terdapat ketidaksesuaian ranah yang digunakan dalam soal dengan indikator pencapaian kompetensi yang telah ditentukan pada indikator pencapaian kompetensi, siswa diharapkan dapat menganalisis mobilitas penduduk. Indikator ini berada pada ranah C4, sedangkan soal berada pada ranah C2. Hal ini juga terjadi pada soal nomor 11, dan nomor 14. Revisi pada indikator kedua terdapat 6 soal, yaitu soal nomor 2, nomor 12, nomor 21, nomor 22, nomor 25, dan nomor 30. Soal nomor 2 kunci jawaban tidak terumuskan dengan benar. Pada soal nomor 12 kunci jawaban juga tidak terumuskan dengan benar. Revisi butir soal nomor 21 perlu dilakukan karena kunci jawaban tidak terumuskan dengan benar dan tidak terdapat pilihan jawaban yang tepat pada soal. Soal nomor 22 kunci jawaban juga tidak terumuskan dengan benar dan terdapat dua pilihan jawaban yang benar pada pilihan jawaban. Hal ini juga terjadi pada soal nomor 25 dan nomor 30.

Aspek kontruksi yang dimaksud di atas relevan berdasarkan validitas konstruk menurut Widoyoko (2016) yang menyebutkan validitas ini mengacu pada sejauh mana suatu instrumen mengukur konsep dari suatu teori kemudian dikuatkan Arikunto (2018) bahwa sebuah tes dikatakan memiliki validitas kontruksi apabila butir-butir soal yang membangun tes tersebut mengukur setiap aspek berfikir seperti yang disebutkan dalam tujuan instruksional khusus. Berdasarkan pendapat tersebut aspek kontruksi ini juga bertujuan untuk mengukur konsep berfikir yang ada di setiap indikator soal. Analisis pada aspek kontruksi menggunakan enam indikator yang sudah ditentukan sebelumnya.

Berdasarkan analisis yang dilakukan pada indikator/aspek 1. Pokok soal harus dirumuskan dengan jelas dan tegas, 2. Semua pernyataan yang dijadikan pilihan dirumuskan secara lugas sehingga jelas dapat dibedakan apakah setiap tersebut benar atau salah, 3. Pokok soal tidak mengarah ke jawaban yang benar, misalnya menggunakan kata salah satu, tiga hal atau yang sejenisnya, 4. Seluruh pernyataan pada alternatif jawaban harus homogen, baik dalam bentuk bunyi kata ataupun istilah, 5. Gambar, grafik, tabel ataupun wacana benar-benar berfungsi, 6. Antar item tidak bergantung satu sama lain yang telah dilakukan terdapat beberapa soal yang perlu direvisi. Butir-butir soal yang perlu direvisi dikarenakan tidak memenuhi indikator pertama, kedua, keempat, dan kelima. Sedangkan pada indikator ketiga dan keenam semua butir soal sudah memenuhi. Indikator pertama terdapat 8 butir soal yang perlu direvisi, yaitu nomor 3, nomor 4, nomor 5, nomor 18, nomor 21, nomor 22, nomor 24, dan nomor 26. Pada butir soal nomor 3 terdapat penjelasan berlebihan pada soal sehingga soal sulit dimengerti oleh siswa. Soal nomor 4 terdapat pernyataan yang tidak sesuai dengan konsep yang ditanyakan pada soal. Soal nomor 5 terdapat pengulangan kata yang membuat soal menjadi tidak efektif, berdasarkan Ayudia (2016) menyatakan bahwa pemilihan kata sangat berpengaruh terhadap makna yang terkandung dalam sebuah kalimat sehingga diksi yang dipilih harus baku, tepat dan tidak bermakna ganda agar tidak ambigu, berdasarkan hal tersebut maka revisi diperlukan untuk membuat soal menjadi lebih efektif. Soal nomor 18 tidak tertulis satuan pada angka, seharusnya satuan ditulis sehingga siswa lebih mudah memahami maksud soal. Soal nomor 21 terdapat gambar piramida penduduk yang kurang jelas terbaca sehingga sulit dipahami. Hal ini juga terjadi pada soal nomor 22. Soal nomor 24 tidak tertulis siswa harus memilih pernyataan yang disediakan sedangkan pada pilihan jawaban siswa harus memilih pernyataan yang ada. Hal ini dapat membuat siswa bingung dalam menjawab soal. Sedangkan pada soal nomor 26 bahasa yang digunakan tidak tepat sehingga sulit dipahami.

Revisi pada indikator kedua terdapat 6 butir soal, yaitu nomor 8, nomor 9, nomor 13, nomor 15, nomor 16, dan nomor 35. Pada butir soal nomor 8 terdapat pilihan jawaban yang tidak homogen dengan pilihan jawaban yang lain. Hal ini juga terjadi pada butir soal nomor 9. Soal nomor 13 tidak dituliskan satuan pada pilihan jawaban sehingga membingungkan siswa. Butir soal nomor 15 terdapat pernyataan yang tidak homogen dengan pernyataan yang lain. Hal ini juga terjadi pada soal nomor 16, sedangkan pada soal nomor 35 terdapat pernyataan yang mudah ditebak oleh siswa sehingga siswa dengan mudah mengeliminasi pernyataan tersebut. Indikator keempat terdapat 3 soal yang perlu direvisi, yaitu butir soal nomor 8, nomor 9, dan nomor 34. Butir soal nomor 8 terdapat pilihan jawaban yang tidak homogen dengan pilihan jawaban yang lain. Hal ini juga terjadi pada butir soal nomor 9 dan nomor 34. Indikator kelima terdapat 4 soal yang perlu direvisi, yaitu butir soal nomor 3, nomor 20, nomor 21, dan nomor 22. Pada soal nomor 3 terdapat rumusan kalimat yang berlebihan sehingga membingungkan siswa. Pada soal nomor 20, 21 dan 22 terdapat gambar piramida penduduk yang

kurang jelas terbaca sehingga membingungkan siswa. Berdasarkan analisis diatas maka soal yang baik secara kontruksi harus memenuhi syarat pokok soal, pernyataan ilustrasi dan konten pelengkap berupa gambar,table,grafik dan sebagainya harus jelas, tegas serta terdapat pengecoh agar tidak mudah mengarahkan ke jawaban, selain itu tentunya pertanyaan soal dan jawaban harus singkron (terhubung).

Analisis dalam aspek bahasa sangat penting karena menjadi salah satu faktor yang mempengaruhi mudah dan sulitnya dalam pengerjaan soal oleh siswa. Hal ini selaras dengan pendapat Wijayanto et al., (2016) bahwa bahasa soal akan mempengaruhi taraf kesukaran soal, apabila bahasa soal kurang mudah untuk dipahami maka soal akan sukar dikerjakan oleh siswa. Begitu juga sebaliknya, apabila bahasa soal mudah untuk dipahami maka siswa dapat mengerjakan soal dengan baik. Berdasarkan analisis yang telah dilakukan terdapat beberapa soal yang perlu direvisi. Butir-butir soal yang perlu direvisi dikarenakan tidak memenuhi indikator pertama, kedua, ketiga, dan keempat, sedangkan pada indikator kelima semua butir soal sudah memenuhi.

Indikator pertama terdapat 16 soal yang perlu direvisi, yaitu nomor 1, 3, 6, 14, 15, 16, 21, 22, 23, 26, 30, 32, 33, 34, 35, dan 40. Soal nomor 1, 3, 6, 14, 15, 16, 23, 30, 32, 33, dan 35 perlu direvisi dengan memberikan tambahan kata perintah pada soal. Soal nomor 21 dan 22 diperlukan penambahan kata sehingga soal lebih komunikatif. Soal nomor 26 terdapat kalimat yang tidak efektif sehingga membingungkan siswa. Soal nomor 34 dan 40 terdapat penggunaan kata yang berlebihan sehingga kalimat menjadi tidak efektif. Revisi pada indikator kedua terdapat 17 butir soal, yaitu nomor 2, 3, 4, 8, 9, 10, 15, 17, 18, 20, 21, 22, 26, 32, 34, 39 dan 40. Soal-soal tersebut terdapat kosa kata yang kurang tepat sehingga harus direvisi, selain itu juga terdapat penggunaan bahasa asing yang tidak dicetak miring. Indikator ketiga terdapat 1 soal yang perlu direvisi, yaitu butir soal nomor 6. Soal tersebut terdapat kalimat yang menimbulkan salah pengertian sehingga perlu direvisi. Indikator keempat juga terdapat 1 butir soal yang perlu direvisi, yaitu nomor 27. Soal tersebut terdapat penggunaan bahasa daerah yang dapat membuat siswa bingung.

Uji validitas ini menggunakan perhitungan statistik sehingga termasuk dalam kategori validitas empiris Arikunto (2018) dan Arifin (2013), sedangkan Widoyoko (2016) menyebutnya dengan validitas butir. Pendapat ketiganya relevan pada dasarnya prosesnya instrumen sudah diuji dari pengalaman lapangan. Berdasarkan hasil perhitungan statisik Uji Validitas yang dilakukan pada soal pilihan ganda di SMA Islam Al Azhar 14 secara persentase soal dengan validitas cukuplah yang paling mendominasi yaitu sebanyak 37,5% sedangkan persentase sisanya terdiri dari 0 % (validitas sangat tinggi), 7,5 % (validitas tinggi), 25 % (validitas rendah) dan 30 % (validitas sangat rendah) sehingga dapat disimpulkan bahwa total persentase soal yang layak digunakan adalah sebesar 45 % yang didapatkan dari total validitas sangat tinggi sampai cukup sehingga hendaknya guru menggunakan dan mempertahankan soal yang berkualitas sebesar 18 soal agar dapat berfungsi baik dalam mengukur tingkat kognitif siswa dalam pembelajaran geografi. Opsi lain apabila guru merasa soal tersebut berjumlah kurang banyak, soal dengan validitas rendah dapat direvisi kembali kemudian diujikan ulang kepada siswa tersebut untuk diketahui tingkat validitasnya sehingga dapat menghasilkan instrumen soal yang berkualitas baik.

Uji reliabilitas ini dihitung menggunakan rumus Hoyt sehingga menggunakan pendekatan reliabilitas internal jadi prosedurnya cukup satu kali instrumen diujikan pada sekelompok siswa sebagai subjek. Kelebihan pendekatan ini diperkuat Widoyoko (2016) karena mempunyai nilai kepraktisan dan efisiensi yang tinggi karena menganalisis data dari satu kali pengumpulan data. Berdasarkan Hasil uji reliabilitas yang telah dilakukan mendapatkan hasil sebesar 0,929. Hasil dikonsultasikan berdasarkan Guilford (1956) sehingga soal pilihan ganda yang dibuat memiliki kriteria sangat tinggi, hasil tersebut sudah tidak perlu dipermasalahkan karena kriteria tersebut berada di posisi puncak.

Uji taraf kesukaran digunakan untuk mengetahui sebaran tingkat kesulitan yang dihadapi siswa dalam mengerjakan instrumen tes sehingga dapat diketahui apakah soal tersebut baik, sedang atau sukar, uji ini juga disebut sebagai uji tingkat kesulitan. Persentase uji taraf kesukaran didominasi oleh soal dengan kriteria sedang sebesar 57,5 %, sisanya 17,5 % (kriteria sukar) dan 25 % (kriteria mudah). Hasil tersebut dikatakan sudah baik karena soal yang baik adalah didominasi oleh soal dengan kriteria sedang yang artinya tidak gampang atau sulit saat dikerjakan oleh siswa. Hal ini didukung oleh Arikunto (2018) bahwa soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar, bahkan Arifin (2013) menyebutkan bahwa jika suatu soal memiliki taraf kesukaran seimbang (proporsional) maka dapat

dikatakan bahwa soal tersebut baik. Berdasarkan hal tersebut maka hendaknya guru menggunakan dan mempertahankan soal yang berkualitas sedang sebesar 25 soal. Guru juga dapat merevisi soal yang mempunyai taraf kesukaran sukar dan mudah kemudian diujikan ulang sehingga menghasilkan instrumen soal yang berkualitas baik.

Uji daya pembeda bertujuan untuk membedakan antara siswa yang mempunyai kemampuan tinggi sehingga dapat menguasai kompetensi belajar dan siswa yang tidak pandai karena belum menguasai kompetensi belajar. Pendapat diatas relevan dengan Arikunto (2018) bahwa daya pembeda merupakan kemampuan soal untuk membedakan siswa yang pandai (kemampuan tinggi) dan siswa yang bodoh (kemampuan rendah) Hasil uji daya pembeda persentase yang telah dilakukan didominasi oleh kriteria cukup dengan persentase 35 % sedangkan sisanya terdapat 27,5 % (daya beda baik sekali), 30 % (daya beda baik), 5 % (daya beda jelek) dan 2,5 % (daya beda negatif). Jika ditotal persentase mulai daya beda baik sekali sampai cukup maka sebesar 92,5 % dari 40 soal yang dibuat dapat digunakan sebesar 37 untuk menguji kemampuan siswa. Sebagaimana uji validitas, jika uji daya pembeda mempunyai kriteria jelek maka dapat direvisi terlebih dahulu kemudian diujikan ulang kepada siswa sampai memperoleh kriteria yang lebih baik daripada sebelumnya sedangkan kriteria negatif sebaiknya dibuang sesuai dengan Arikunto (2018).

Uji efektivitas pengecoh bertujuan untuk menguji apakah pola jawaban soal yang dibuat dapat mengecoh siswa untuk memilih jawaban lain selain kunci jawaban yang sudah ada sehingga dibutuhkan kriteria tertentu dalam menentukan baik atau tidaknya pengecoh. Menurut Arifin (2013) butir soal yang baik, pengecohnya akan dipilih secara merata oleh peserta didik yang menjawab salah. Sebaliknya, butir soal kurang baik, pengecohnya akan dipilih secara tidak merata, kemudian dikuatkan oleh Muhson (2017) bahwa kriteria pengecoh alternatif butir soal yang baik adalah jika alternatif tersebut dijawab oleh minimal 5% peserta tes, sehingga alternatif itu dianggap efektif. Berdasarkan uji efektivitas pengecoh (*distractor*) dengan berbantuan AnBuso terlihat hasilnya didominasi oleh soal yang perlu direvisi sebesar 30 % dan sisanya 17,5 % dengan kriteria baik, 17,5 %, (cukup baik) dan 35 % (tidak baik), mengetahui hal tersebut maka sebaiknya soal sebanyak 14 kriteria baik dan cukup baik perlu dipertahankan sedangkan kriteria revisi pengecoh dan tidak baik harus segera diperbaiki untuk melengkapi jumlah soal agar lebih banyak digunakan dalam mengukur kompetensi siswa. pengecohnya akan dipilih secara tidak merata, kemudian dikuatkan oleh Muhson (2017) bahwa kriteria pengecoh alternatif butir soal yang baik adalah jika alternatif tersebut dijawab oleh minimal 5% peserta tes, sehingga alternatif itu dianggap efektif. Berdasarkan uji efektivitas pengecoh (*distractor*) dengan berbantuan AnBuso terlihat hasilnya didominasi oleh soal yang perlu direvisi sebesar 30 % dan sisanya 17,5 % dengan kriteria baik, 17,5 %, (cukup baik) dan 35 % (tidak baik), mengetahui hal tersebut maka sebaiknya soal sebanyak 14 kriteria baik dan cukup baik perlu dipertahankan sedangkan kriteria revisi pengecoh dan tidak baik harus segera diperbaiki untuk melengkapi jumlah soal agar lebih banyak digunakan dalam mengukur kompetensi siswa.

SIMPULAN

Berdasarkan analisis dan pembahasan maka dapat disimpulkan bahwa: 1) aspek materi/substansi terdapat revisi pada indikator pertama dan kedua; 2) aspek kontruksi terdapat revisi pada indikator pertama, kedua, keempat, dan kelima; 3) aspek bahasa terdapat revisi pada indikator pertama, kedua, ketiga, dan keempat; 4) Uji validitas terdapat 18 soal dapat dipertimbangkan untuk digunakan; 5) Uji reliabilitas butir mempunyai tingkat kepercayaan sangat tinggi; 6) Uji taraf kesukaran terdapat 23 soal dapat dipertimbangkan untuk digunakan; 7) Uji daya beda soal terdapat 37 soal dapat dipertimbangkan untuk digunakan; dan 8) Uji efektivitas pengecoh/*distractor* terdapat 14 soal dapat dipertimbangkan untuk digunakan. Berdasarkan uji secara aspek kualitatif dan aspek kuantitatif dapat disimpulkan bahwa instrumen tes yang dapat digunakan adalah soal nomor 1, 3, 4, 5, 6, 9, 11, 13, 14, 17, 20, 21, 22, 24, 26, 27, 28 dan 37 karena lebih berkualitas.

DAFTAR PUSTAKA

Aliman, et al (2020). Pengembangan Instrumen Tes Kemampuan Berpikir Spasial Bagi Siswa SMA. *Geodika: Jurnal Kajian Ilmu dan Pendidikan Geografi*, 4(1), 1-10.

- Arifin, Z. (2013). *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Arikunto, S. (2018). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: PT Bumi Aksara.
- Ayudia, A., Suryanto, E., & Waluyo, B. (2017). Analisis Kesalahan Penggunaan Bahasa Indonesia dalam Laporan Hasil Observasi pada Siswa SMP. *Basastra*, 4(1), 34-49.
- Guilford, J. P. (1956). *Fundamental Statistics in Psychology and Education*. New York: Mc Graw-Hill Book Co. Inc.
- Kunandar. (2014). *Penilaian Autentik*. Jakarta: PT. Raja Grafindo Persada.
- Muhson, A. (2017). *Panduan Penggunaan AnBuso*. Yogyakarta: Universitas Negeri Yogyakarta.
- Nur, A. S., & Palobo, M. (2018). Pelatihan Analisis Butir Soal Berbasis Komputerisasi pada Guru SD. *MATAPPA: Jurnal Pengabdian Kepada Masyarakat*. 1(1) 5–11.
- Nurjanah, & Marlianingsih, N. (2015). Analisis Butir Soal Pilihan Ganda dari Aspek Kebahasaan. *Faktor Jurnal Ilmu Kependidikan*, 2(1), 69–78.
- Rahayu, T. D., Purnomo, B. H., & Sukidin, S. (2014). Analisis Tingkat Kesukaran dan Daya Beda Pada Soal Ujian Tengah Semester Ganjil Bentuk Pilihan Ganda Mata Pelajaran Ekonomi Kelas X di SMA Negeri 5 Jember Tahun Ajaran 2012-2013. *Jurnal Edukasi*, 1(1), 39-43.
- Solichin, M. (2017). Analisis Daya Beda Soal. Taraf Kesukaran, Butir Tes, Validitas Butir Tes, Interpretasi Hasil Tes Validitas Ramalan dalam Evaluasi Pendidikan. *Journal Unipdu*, 2(2), 192–213
- Subali. B. (2014). Analisis Soal Baik Kualitatif maupun Kuantitatif. *Disajikan pada Kegiatan Workshop Item Development Dosen Poltekes Kebidanan Politeknik Kesehatan Surakarta tanggal 18-19 Agustus 2014 di Griya Persada Conventional Hotel & Resort, Jl Boyong Kalinrang Barat*. 1-36
- Sudijono, A. (2015). *Pengantar Evaluasi Pendidikan*. Jakarta: PT. Raja Grafindo Persada.
- Widoyoko, S.E.P. (2016). *Penilaian Hasil Pembelajaran di Sekolah*. Yogyakarta: Pustaka Pelajar.
- Wijayanto, P.A., (2017). An Evaluation of the Quality of Geography Learning Program at SMA Negeri 1 Mayong Using Ekop Model. *Jurnal Pendidikan Indonesia*. 6(2), 172-180.
- Wijayanto, P. A., Allifah, A., & Amirrudin, A. (2016). Evaluasi Kualitas Instrumen Tes dalam Pembelajaran Geografi di MAN 2 Kota Batu. *Jurnal Geografi: Media Informasi Pengembangan dan Profesi Kegeografian*, 13(2), 101-113.