



Received June 2021 Accepted December 2021 Published December 2022

## ANALYSIS OF THE ITEMS IN THE SCIENCE SEMESTER EXAM AT SMP NEGERI 30 SOLOK SELATAN ACADEMIC YEAR 2019/2020

Lusi, W.P.<sup>1 a)</sup>, Zahra, F.A.<sup>1</sup>

<sup>1</sup>Department of Science Education, Universitas Negeri Padang

<sup>a)</sup>E-mail : [windapitrusi@gmail.com](mailto:windapitrusi@gmail.com)

### ABSTRACT

The Class VII Science Odd Semester Exam Questions at SMP Negeri 30 Solok Selatan for the 2019/2020 academic year are not analyzed qualitatively and quantitatively and do not have standardized questions. Standardized questions are questions that have been tried several times, have a coefficient of validity, reliability, level of difficulty, distinguishing power, and other information deemed necessary. Therefore, the researcher felt the need to analyze the items. This study aims to determine the quality of the items qualitatively and quantitatively (reliability, validity, distinguishing power, difficulty level, and distractor quality). This research is descriptive research. The research population consisted of all the answer sheets for the Class VII IPA odd semester examination at SMP Negeri 30 Solok Selatan for the academic year 2019/2020, totaling 26 students. The sampling technique used total sampling technique. The research instrument used a question construction quality questionnaire. Data collection techniques with document study. The quantitative results were analyzed using the ANATES program. Based on the research results, it was found that: 1) The construct validity of the questions did not yet fulfill all aspects, both in terms of science content, construction, and language; 2) The validity of the questions was low, dominated by invalid questions; 3) Reliability category is sufficient; 4) Poor differentiation; 5) Medium level of difficulty; 6) Distractor quality is good and functioning. Thus, the questions have not met the criteria for good and standardized questions, so they need to be revised.

©Department of Science Education, Universitas Negeri Padang

**Keywords:** Analysis of Question Points, Science, Junior High School, Anates.

### INTRODUCTION

Evaluasi adalah siklus persiapan, pengambilan, dan pemberian data yang penting untuk menentukan pilihan

elektif. Dalam data sosial harus didapatkan informasi yang sesuai dan mendukung tujuan yang telah ditetapkan. Pemahaman dalam mengambil hasil yang dilihat dari pencapaian

tujuan pembelajaran di kelas dapat diperoleh dari penilaian ini (Arikunto, 2015: 3).

Seperti yang diindikasikan oleh Purwanto (2012: 3), asesmen adalah upaya yang dilakukan secara metodis dan hati-hati untuk memahami kemajuan dan kapasitas siswa selama pembelajaran, sebelum pembelajaran, dan setelah pembelajaran, melalui bermacam-macam informasi dan kontras serta standar atau model tertentu. Dalam latihan penilaian, diperlukan data tentang item yang akan dinilai. Dalam latihan mengukur pembelajaran, informasi yang dimaksud adalah informasi seperti perilaku atau penampilan siswa selama siklus pembelajaran, hasil ulangan hari demi hari, nilai ulangan tengah semester, dan nilai ulangan semester terakhir. Untuk memperoleh informasi, salah satu cara yang ditempuh dalam penilaian adalah dengan menguji hasil belajar.

Tes hasil belajar merupakan tes yang digunakan untuk melihat hasil belajar yang telah diberikan pendidik kepada siswa dalam kurun waktu tertentu. Dalam melaksanakan tes hasil belajar, seorang instruktur dapat memperoleh hasil belajar dengan memanfaatkan dua macam tes, yaitu tes yang sudah dinormalisasi dan tes yang dilakukan oleh pendidik yang sebenarnya (Purwanto, 2012: 33). Hasil belajar yang didapat dapat dimanfaatkan sebagai masukan (kritik) bagi pendidik dan otoritas publik dalam perbaikan dan puncak proyek dan latihan pembelajaran. Para pendidik secara rutin memberikan ulangan setiap hari, ulangan semester akhir, ulangan tersusun, ulangan lisan, ulangan kegiatan, dll, istilah-istilah ini pada dasarnya penting untuk kerangka penilaian itu sendiri (Arifin, 2012: 6).

Kualitas ujian (soal) dapat dilihat dari hasil yang diperoleh mahasiswa. Jika banyak siswa mendapatkan nilai rendah, itu berarti

bahwa ujian yang diselenggarakan terlalu merepotkan, sebaliknya jika semua siswa mendapatkan nilai tinggi, sangat mungkin dapat diartikan bahwa ujiannya terlalu sederhana. Untuk menghindari hal tersebut, maka penting untuk mengarahkan pemeriksaan terhadap hakikat pertanyaan tersebut, salah satunya dengan mengurai segala sesuatunya. Penyelidikan adalah siklus yang efisien, yang akan memberikan data yang cukup pasti tentang hal-hal yang diatur. Keuntungan melakukan penyelidikan adalah: (1) membantu kami membedakan hal-hal yang buruk, (2) memperoleh data yang dapat digunakan untuk menyelesaikan pertanyaan untuk alasan tambahan, dan (3) melihat sekilas kondisi masalah yang kami kumpulkan (Arikunto, 2015: 220-221).

Penyelidikan yang layak adalah pertanyaan yang substansial. Sebuah tes dapat dianggap sah jika tes tersebut dapat mengukur apa yang Anda butuhkan untuk mengukur seperti yang ditunjukkan oleh gagasan. Pertanyaan yang sah adalah pertanyaan yang ketika kita mengatakan kita akan mengukur kapasitas X peserta tes, pada saat itu kita harus mengukur sifat atau kualitas luar biasa yang diidentifikasi dengan peserta tes X untuk diperkirakan (Guoin, 1977: 65).

Soal-soal yang dimanfaatkan untuk ulangan semester ganjil adalah soal-soal yang dibuat oleh rombongan IPA IPA Kabupaten Solok Selatan atas usulan MKKS (Musyawarah Kerja Kepala Sekolah). Inkuiri ini dimanfaatkan oleh semua sekolah menengah di Kabupaten Solok Selatan. Namun demikian, penelitian terhadap hal-hal tersebut belum dilakukan, sehingga kualitasnya tidak diketahui secara sah atau eksperimental.

Setelah (ujian) selesai, tidak ada nilai siswa yang sampai di KKM yang telah ditentukan oleh pendidik, khususnya 75. Penunjuk yang layak untuk menentukan

nilai Kriteria Ketuntasan Minimum (KKM) harus dimungkinkan dalam dua Tahapan, secara spesifik tahapan pokok adalah memperoleh informasi tentang kewajaran permintaan keterangan yang telah dibuat dengan penanda dalam prospektus. Demikian pula penelitian Mardapi (2015) menunjukkan bahwa siswa telah mendominasi perencanaan pendidikan, khususnya kapasitas hasil belajar yang diperkirakan telah mencapai aturan pemenuhan terkecil yang telah ditetapkan (KKM) atau bahkan memiliki pilihan untuk melampaui KKM, dan yang belum tuntas didominasi lewat pengobatan.

Oleh karena itu, sangatlah penting untuk menyelesaikan soal ujian semester ganjil IPA, yang nantinya dapat memperoleh hasil belajar psikologi yang memenuhi standar ujian penyelenggara negara. Dengan cara ini alamat ujian semester ganjil harus disetujui dan diuji ulang untuk mengukur kapasitas siswa dengan benar.

### PURPOSE

Tujuan penelitian ini adalah untuk mengukur kualitas konstruksi soal, reabilitas soal, validitas item soal, tingkat kesukaran soal, daya beda, serta mengukur kualitas *option* pengecoh butir soal ujian semester ganjil IPA kelas VII SMP Negeri 30 Solok Selatan tahun pelajaran 2019/2020.

### METHOD

Penelitian ini menggunakan metode deskriptif kuantitatif. Analisis data secara kuantitatif diperoleh melalui strategi studi dokumentasi, khususnya mendapatkan pertanyaan, kunci jawaban, dan lembar jawaban siswa yang nantinya akan disusun dengan menggunakan program Anates, yang konsekuensinya meliputi ketergantungan inkuiri, legitimasi benda, tingkat kesulitan, pemisahan, dan memperkirakan sifat dari

pengalihan pilihan. Penyelidikan subyektif diselesaikan untuk menentukan legitimasi pengembangan inkuiri yang meliputi 3 sudut pandang, yaitu perspektif materi, perkembangan dan bahasa. Informasi tersebut diperoleh dari instrumen polling legitimasi pembangunan yang ditetapkan oleh dosen validator.

## RESULT AND DISCUSSION

### Hasil Penelitian

Analisis 40 butir soal ujian semester ganjil IPA kelas VII SMP Negeri 30 Solok Selatan Tahun Pelajaran 2019/2020 dilakukan secara kualitatif dan kuantitatif. Analisis atau pengembangan subyektif diselesaikan pada 3 bagian evaluasi, yaitu substansi sains, bangunan, dan bahasa. Dari hasil investigasi, tingkat pertanyaan yang memenuhi bagian dari konten sains adalah 40% dengan kelas tidak berdaya, 65% membangun dengan klasifikasi yang layak, dan 60% bahasa dengan kelas yang benar-benar layak. Hasil ini didukung oleh urutan pertanyaan yang bergantung pada perubahan Purwanto (2012: 139), khususnya 0% - 20% dalam klasifikasi buruk, 21% - 40% di kelas tak berdaya, 41% - 60% di kelas klasifikasi yang memadai, 61% - 80% di kelas hebat, dan 81% - 100% klasifikasi mengagumkan. Konsekuensi dari survei hal-hal tersebut dapat dilihat pada Tabel 1.

**Table 1.** Hasil Analisis Validitas Konstruksi

Aspek Penelaahan Soal	Jumlah soal yang memenuhi semua aspek	Persentase (%)
Konten ilmu	16	40
Konstruksi	26	65
Bahasa	24	60

Analisis secara kuantitatif terdiri dari validitas Empiris, reliabilitas, daya pembeda, tingkat kesukaran, dan kualitas pengecoh. Hasil analisis validitas empiris adalah 57,5% soal tidak valid karena nilai  $r$  hitungannya lebih

rendah dari nilai r tabel. Nilai r tabel dari 40 jumlah soal yang diujikan dengan taraf signifikansi 0,05 (5%) adalah 0,304. Secara keseluruhan dapat dilihat pada Tabel 2.

**Tabel 2.** Hasil Analisis Validitas Empiris

Kategori	Jumlah	Persentase(%)
Valid	17	42,5
Tidak Valid	23	57,5

Berdasarkan hasil pengolahan data diperoleh nilai reliabilitas soal objektif dan essay pada ujian semester ganjil IPA kelas VII SMP Negeri 30 Solok Selatan Tahun Pelajaran 2019/2020 secara keseluruhan adalah 0,49 dan 0,41 dengan kriteria cukup dan soal sudah reliabel. Hasil analisis dapat dilihat pada Tabel 3.

**Tabel 3.** Hasil Analisis Reliabilitas

Jenis soal	Jumlah	Reliabilitas
Objektif	35	0,49
Essay	5	0,41

Selanjutnya hasil analisis daya pembeda, tingkat kesukaran, dan kualitas pengecoh dapat dilihat pada Tabel 4, Tabel 5, dan Tabel 6.

**Tabel 4.** Hasil Analisis Daya Pembeda

Kategori	Jumlah	Persentase (%)
Jelek	17	42,5
Sedang	7	17,5
Baik	12	30
Baik sekali	2	5
Nilai negative	2	5

Tabel 4 menunjukkan hasil analisis daya pembeda soal secara keseluruhan yaitu 42,5% jelek, 17,5% sedang, 30% baik, 5% baik sekali, dan 5% nilai negatif. Jumlah peserta tes kelompok bawah dan kelompok atas adalah 7 orang diambil dari 27% dari jumlah anggota ujian.

**Tabel 5.** Hasil Analisis Tingkat Kesukaran

Kategori	Jumlah	Persentase (%)
Sangat sukar	2	5
Sukar	7	17,5
Sedang	19	47,5
Mudah	7	17,5
Sangat mudah	5	12,5

Tabel 5 menunjukkan hasil analisis tingkat kesukaran soal secara keseluruhan yaitu 5% soal sangat sukar, 17,5% soal sukar, 47,5% soal sedang, 17,5% soal mudah, dan 12,5% soal sangat mudah.

**Tabel 6.** Hasil Analisis Kualitas Pengecoh

Kategori	Jumlah	Persentase (%)
Berfungsi	29	82,86
Tidak berfungsi	6	17,14

Tabel 6 menunjukkan hasil analisis kualitas pengecoh soal secara keseluruhan yaitu 82,86% berfungsi dan 17,14% tidak berfungsi.

## Pembahasan

Validitas butir soal menjadi dua jenis, yaitu kecerdasan dan pengamatan. Legitimasi yang masuk akal juga dibagi menjadi substansi dan mengembangkan legitimasi. Dalam penelitian ini, legitimasi substansi tidak dilihat dan dibedah karena matriks inquiry yang menyertakan inquiry pointers tersebut tidak ada. Pengujian dilakukan terhadap legitimasi pembangunan dan legitimasi eksperimental. Pemeriksaan pengembangan legitimasi penyelidikan diselesaikan menggunakan konfigurasi yang mencakup 3 bagian penelitian termasuk substansi ilmu pengetahuan, perkembangan, dan sudut bahasa.

Mengingat konsekuensi dari pemeriksaan build terhadap inquiry, disadari bahwa mengenai konten sains hanya 40% dari inquiry yang mencakup semua sudut pandang yang tergolong merepotkan. Sedangkan hal-hal yang tidak memenuhi semua standar sudut pandang yang berkaitan dengan sains adalah hal-hal nomor 2, 3, 6, 8, 9, 11, 12, 14, 15, 16, 17, 18, 20, 21, 22, 23, 28, 29, 31, 31, 36, 37, 38, dan 39, dengan alasan bahwa hal-hal tersebut tidak memenuhi 5 sudut yang terkandung dalam ilmu pengetahuan. Sudut-sudut ini menggabungkan hal-hal inkuiri yang tidak sesuai dengan petunjuk, tidak hanya ada satu kunci jawaban, substansi bahan tidak

sesuai dengan tujuan estimasi, substansi bahan tidak sesuai dengan kebutuhan. tingkat, jenis, sekolah dan tingkat kelas, dan ada keputusan jawaban yang tidak berhasil.

Dalam pemeriksaan sudut pandang pengembangan, hanya 65% pertanyaan yang mencakup semua perspektif, dengan kelas yang layak. Sementara hal-hal yang tidak memenuhi semua aturan untuk sudut pandang dalam pengaturan pengembangan adalah hal-hal 1, 2, 11, 14, 16, 18, 28, 29, 31, 36, 37, 38, 39, dan 40, dengan alasan bahwa hal-hal ini tidak memenuhi 8 perspektif yang terkandung dalam pengaturan pengembangan. Sudut-sudut ini menggabungkan topik yang tidak diklarifikasi dengan jelas, rencana penyelidikan dan keputusan jawaban tidak dirinci secara jelas, topik tidak memberikan potongan informasi untuk keputusan jawaban yang benar, keputusan jawaban, ada jawaban yang homogen, jawaban yang cukup mirip keputusan, keputusan jawaban adalah seperti angka, waktu tidak diatur, pembicaraan dan gambar tidak berfungsi.

Dalam pemeriksaan sudut bahasa, hanya 60% pertanyaan yang mencakup semua sudut pandang, dengan klasifikasi yang benar-benar layak. Sedangkan hal-hal yang tidak memenuhi semua ukuran untuk bagian-bagian pengaturan bahasa adalah hal-hal nomor 2, 3, 4, 9, 11, 14, 16, 18, 20, 28, 29, 30, 31, 32, 36, dan 37, mengingat fakta bahwa hal-hal ini tidak memenuhi 6 perspektif yang terkandung dalam pengaturan pengembangan. Perspektif ini menggabungkan perincian kalimat tanya jawab, kalimat yang tidak menggunakan bahasa yang baik dan benar sesuai dengan jenis bahasanya, rencana kalimat menyebabkan banyak pemahaman, tidak menggunakan bahasa biasa, ada pertanyaan yang dapat mengganggu siswa, dan Adadefinisi soal yang mengandung SARA (Suku, Agama, Ras, Antar Kelompok).

Legitimasi yang tepat dari pertanyaan yang didapat kewalahan oleh pertanyaan yang tidak valid. Invalid inquiry adalah pertanyaan

yang memiliki harga  $r$  ditentukan lebih sederhana daripada  $r$  tabel. Nilai  $r$  tabel dari 40 jumlah pertanyaan yang dicoba dengan derajat kritis 0,05 (5%) adalah 0,304, pertanyaan yang memiliki  $r$  tabel di bawah 0,304 mengandung arti bahwa pertanyaan yang diajukan tidak besar atau tidak valid, menyiratkan bahwa hal-hal tersebut belum memiliki Pilihan untuk mengukur kapasitas siswa. Hal ini menunjukkan bahwa pertanyaan yang diajukan tidak dapat diterima, karena permintaan yang diumumkan dapat diterima merupakan keabsahan permintaan dengan kelas koefisien antara cukup sampai dengan sangat baik dan kualitas pertanyaan yang tidak tergoyahkan juga termasuk dalam klasifikasi tinggi. Seperti yang ditunjukkan oleh Anderson (dalam Arikunto 2015: 101) sebuah tes mungkin solid namun tidak substansial. Menariknya, tes substansial biasanya solid. Penyelidikan yang layak adalah penelitian yang didelegasikan secara substansial sehingga dapat mengukur kapasitas mahasiswa.

Kualitas pertanyaan yang tidak tergoyahkan secara keseluruhan adalah 0,49 dan 0,41, yang berarti bahwa pertanyaan tersebut memiliki standar yang memadai. Standar tersebut memadai untuk menunjukkan bahwa kualitas pertanyaan yang tak tergoyahkan telah terpenuhi sebagai pertanyaan yang benar-benar layak atau dapat dipercaya dalam memperkirakan konsistensi kapasitas siswa. Semakin tinggi koefisien ketergantungan suatu tes, semakin tinggi ketajaman atau ketepatannya. Tinggi rendahnya koefisien ketergantungan suatu benda dipengaruhi oleh beberapa unsur, tepatnya menurut Departemen Pendidikan Nasional (2008: 3), unsur yang mempengaruhi nilai mutu yang tidak tergoyahkan adalah lamanya ujian, jangka waktu ujian. waktu berurusan dengan pertanyaan, homogenitas sisi ekuator, dan tingkat masalah penyelidikan.

Butir soalyang bagus juga harus memiliki opsi untuk menunjukkan kekuatan

khasnya. Studi tentang kekuatan terpisah berencana untuk menentukan kapasitas pertanyaan pada siswa tertentu yang didelegasikan berbakat tinggi dari siswa yang disebut berkapasitas rendah. Konsekuensi dari pemeriksaan untuk nilai pemisah pertanyaan secara keseluruhan adalah 42,5% buruk, 17,5% sedang, 30% hebat, 5% mengagumkan, dan 5% skor negatif. Jumlah peserta tes kelompok bawah dan kelompok atas adalah 7 orang yang diambil dari 27% dari jumlah keseluruhan peserta tes. Hal ini sesuai penilaian Sudijono (2012: 398-399) bahwa jika individu yang diuji cukup besar (100 individu atau lebih), daya pisah cukup untuk menghitung 27% kelompok atas dan 27% kelompok bawah. , sedangkan individu uji yang ditempatkan di antara dua pertemuan tidak harus dipertimbangkan. Masalah dengan kontras yang bagus, sedang, dan luar biasa adalah pertanyaan yang dapat mengenali siswa berkapasitas tinggi dari siswa berkapasitas rendah. Akibatnya ada 52,5% pertanyaan yang bagus untuk digunakan, 42,5% dalam klasifikasi buruk, yang berarti bahwa banyak pertanyaan ditangani secara efektif oleh siswa berkapasitas rendah dan banyak yang salah dijawab oleh siswa berkapasitas tinggi, di Saat itu ada 5% pertanyaan yang negatif, yang berarti pertanyaan ini tidak dapat mengenali siswa berkapasitas tinggi dari siswa berkapasitas rendah. Hal ini ditunjukkan dengan nilai biser pada angket nomor 16 adalah - 14.29 dan titik biser angket nomor 17 adalah - 14.29.

Kualitas atau tidaknya tes hasil belajar juga dapat dilihat dari tingkat kesusahan atau tingkat kesusahan yang dimiliki setiap benda. Dari hasil investigasi, tingkat masalah umum dari pertanyaan tersebut adalah 5% pertanyaan merepotkan, 17,5% pertanyaan merepotkan, 47,5% pertanyaan sedang, 17,5% pertanyaan sederhana, dan 12,5% pertanyaan mudah. Oleh karena itu, penyelidikan ini didelegasikan secara moderat karena hal yang memiliki daftar tingkat masalah paling tinggi

adalah 47,5% untuk pertanyaan menengah. Semakin rendah tingkat kesulitan pertanyaan, semakin merepotkan pertanyaan, karena beberapa peserta tes menjawab pertanyaan dengan efektif. Sesuai dengan Sunarti dan Selly (2014: 138), derajat kesusahan pertanyaan disesuaikan, "Untuk kesiapansalinan asli ujian, sebaiknya digunakan benda- benda dengan tingkat gangguan yang adil, lebih spesifik menyusahkan 25 %, setengah sedang, dan sederhana 25% ". Berdasarkan Tabel 4, tingkat masalah berada pada klasifikasi sedang.

Pengecoh dikatakan bekerja dengan baik jika setidaknya 5% dari semua peserta ujian dipilih. Jika tidak persis seperti rate ini, cheat tersebut tidak praktis dan harus diganti. Hasil pemeriksaan sifat umum pemerasan penyelidikan adalah 82,86% berfungsi dan 17,14% tidak berfungsi. Pada titik itu cenderung terlihat bahwa 82,86% dari keputusan tanggapan yang tepat dalam pertanyaan dapat berfungsi. Ini menyiratkan bahwa keputusan respons yang tepat (bukan kunci respons yang tepat) telah bekerja dengan tepat, lebih spesifik sebagai interupsi. Sisanya, yaitu 17,14%, tidak berfungsi. Dalam pertanyaan target dengan tingkat masalah yang sederhana dan sangat sederhana, pengganggu tidak berfungsi, lebih spesifiknya pada pertanyaan 1, 2, 8, 20, 25, dan 31.

Konsekuensi dari investigasi secara keseluruhan, 15% dari pertanyaan yang dapat digunakan dan dapat digunakan untuk tes tambahan dan masuk ke bank penyelidikan, 10% dari pertanyaan harus diperbaiki sebelum digunakan sebagai tes kapasitas siswa, dan 75 % dari pertanyaan yang dibuang, pertanyaan ini tidak boleh digunakan untuk survei kapasitas siswa. Dari hasil tersebut, soal tes semester ganjil IPA kelas 7 di SMP Negeri 30 Solok Selatan tahun ajaran 2019/2020 secara umum tidak memenuhi ukuran soal bagus dan sudah dinormalisasi karena soal bagus dandan terstandar harus valid dan reliabel.

## CONCLUSION

### Kesimpulan

Berdasarkan analisis data pada penelitian ini, sangat mungkin dapat dianggap bahwa pertanyaan tes IPA tingkat 7 SMP Negeri 30 Solok Selatan untuk tahun ajaran 2019/2020 memiliki nilai penyelidikan subjektif dan kuantitatif yang menyertai: 1) Legitimasi penyelidikan berkembang sejauh substansi sains, perkembangan, dan bahasa belum memenuhi semua sudut; 2) Legitimasi hal-hal yang rendah, yang diliputi oleh pertanyaan-pertanyaan yang tidak valid; 3) Kelas reliabilitas memadai; 4) Pemisahan klasifikasi yang buruk; 5) Tingkat kesulitan kelas sedang; 6) Kualitas bagus dan klasifikasi pengalihan perhatian yang berguna. Dengan demikian cenderung diasumsikan bahwa soal tes ganjil semester IPA kelas VII SMP Negeri 30 Solok Selatan tahun ajaran 2019/2020 belum memenuhi ukuran soal butir bagus dan ternormalisasi.

### Saran

Berdasarkan penelitian yang dilakukan, maka peneliti menyarankan beberapa hal, yaitu sebagai berikut: 1) Kami menyarankan agar kelompok dan instruktur melakukan latihan investigasi atas pertanyaan-pertanyaan yang telah dibuat, untuk mengetahui sifat dari pertanyaan-pertanyaan tersebut sehingga kemudian mereka dibuat lebih baik secara mengejutkan; 2) Agar kelompok pembentuk inquiry dapat memasukkan soal yang telah memenuhi aturan ke dalam inquiry bank dan dapat digunakan kembali, sedangkan soalnya yang tidak memenuhi ukuran sebaiknya direvisi atau tidak, pada saat ini digunakan; 3) Meminta pendidik mata pelajaran IPA saling membantu dalam perencanaan jejaring inkuiri, pembuatan soal, dan ulangan soal, serta terus berupaya memperbaiki hakikat dirinya dalam membuat soal dan mendobrak soal; 4) Agar sekolah dapat mengawasi latihan penyelidikan inkuiri yang diselesaikan oleh instruktur, dan membantu pendidik dalam mengarsipkan

pertanyaan besar; 5) Dinas Pendidikan Provinsi diandalkan untuk melakukan

pendekatan yang memuat komitmen instruktur untuk berusaha membedah inquiry sebelum dimanfaatkan, seperti halnya mengadakan persiapan / workshop / berbagai upaya untuk meningkatkan sifat instruktur dalam membuat dan menyelidiki pertanyaan; 6) Hasil akhir dari ujian ini dapat dimanfaatkan sebagai informasi awal untuk penelitian tentang kemajuan soal-soal sains sekolah menengah.

## REFERENCES

- Arifin, Zaenal. 2012. *Evaluasi Pembelajaran*. Bumi Siliwangi: Rosdakarya.
- Arikunto, Suharsimin. 2015. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Departemen Pendidikan Nasional. 2008. *Panduan Analisis Butir Soal*. Jakarta: Depdiknas.
- Guion, R.M. (1977). *Content Validity-The Source of My Discontent, Applied Psychological Measurement*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Mardapi, Djemari, Dkk. 2015. Menentukan Kriteria Ketuntasan Minimal Berbasis Peserta Didik. Universitas Negeri Yogyakarta. *Jurnal Penelitian Dan Evaluasi Pendidikan* Volume 19, No 1, Juni 2015 (38-45).
- Purwanto, Ngalim. 2012. *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Bandung: Remaja Rosdakarya Offset.
- Sudijono, Anas. 2012. *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo.

Sunarti & Selly, R. 2014. *Penilaian dalam Kurikulum 2013*. Yogyakarta: Andi.