

PENGEMBANGAN INSTRUMEN TES FORMATIF FISIKA KELAS XI SEMESTER GASAL PROGRAM AKSELERASI ¹⁾

Winda Fitrifitanofa²⁾ Sutadi Waskito, M.Pd³⁾, Rini Budiharti, M.Pd³⁾

ABSTRACT

This research aims to (1) develop a multiple choice formative test for the Odd Semester XI Grade of Acceleration Program corresponding to the existing Standard Competency and Basic Competency, (2) to develop the outline of multiple choice formative test for the Odd Semester XI Grade in Kinematics material with Vector Analysis, Gravitation And Harmonic Movement in Elastic Object, (3) to develop the multiple choice item of for formative test of the Odd Semester XI Grade of Acceleration Program in Kinematics material with Vector Analysis, Gravitation And Harmonic Movement in Elastic Object, and (4) validating the result of formative test instrument development for the Odd Semester XI Grade of Acceleration Program in Kinematics material with Vector Analysis, Gravitation And Harmonic Movement in Elastic Object either qualitatively and quantitatively.

This study was a developmental research using Research and Development method. Techniques of analyzing data used were a descriptive qualitative technique with item study by the expert and a quantitative technique using Microsoft Excel program. The test development was done by studying the item by the expert, and then revising its validation. The test instrument consisted of 64 items that had been tried out for its validity. In the smaller group tryout, 40 items had good quality, and 24 items should be revised. The revision of the items with poor quality (24 items) was done continuously until the expert considered that these had been good for being tried out along with the 40 valid items to the larger group.

The physics formative test of odd semester XI grade of Acceleration Program that had been developed consisted of 3 material for mid-semester including Kinematics with Vector Analysis, Gravitation And Harmonic Movement in Elastic Object. The instrument development was initiated by making an outline corresponding to the syllabus of National Education Service for the XI grade. The development of outline was also needed to facilitate the administration of student competency's track record (authentic assessment). The physics formative test of odd semester XI grader of Acceleration Program that had been validated indicated the test instrument corresponding to the characteristics of good physic test. The test instrument developed had good content validity, high item reliability, for Kinematics with Vector Analysis material (Package 1) had reliability of 0.810710 belonging to very high category, for Gravitation material item (Package 2) had reliability of 0.6844 belonging to high category, and for Harmonic Movement with Elastic Object material had reliability of 0.824764 belonging to very high category. The variance was measured using discrimination index indicating that all results were supported, $D > 0,3$. The difficulty level of item fell into moderate category $0,3 \leq P \leq 0,7$, and confounding coefficient functions well. From the development of physics test of the Odd Semester XI Grade of Acceleration Program, 64 items were obtained with good quality.

Keywords: development, formative test, Physics for XI grade, Acceleration Program

¹⁾ Penelitian dan Pengembangan

²⁾ Mahasiswa Prodi Pendidikan Fisika Jurusan P.MIPA FKIP UNS

³⁾ Dosen Prodi Pendidikan Fisika Jurusan P.MIPA FKIP UNS

PENDAHULUAN

Fenomena pertumbuhan kehidupan masyarakat maju, semakin lama semakin menunjukkan bahwa kunci perkembangan dan pertumbuhan yang terjadi ternyata bergantung pada sumber daya manusia yang berkualitas. Menghadapi persaingan global akan kebutuhan sumber daya manusia (SDM) yang berkualitas merupakan kebutuhan mendesak bagi suatu negara agar dapat sejajar dengan warga dunia lainnya. Indonesia harus menyiapkan sumber daya manusia yang proaktif, kreatif, inovatif, mandiri dan memiliki keunggulan komparatif dan kompetitif untuk menghadapi persaingan global.

Salah satu upaya untuk mengantisipasi kondisi persaingan tersebut adalah melalui pengembangan pendidikan yang berdimensi keunggulan. Pendidikan yang berdimensi keunggulan dalam hal ini adalah pendidikan bagi anak-anak Cerdas Istimewa Bakat Istimewa (CIBI). Di Indonesia sendiri sebenarnya sudah ada perhatian untuk memberikan layanan pendidikan bagi CIBI hal ini terbukti mulai tahun 2000 yakni pada saat Mendiknas dipimpin oleh Yahya Muhaimin Indonesia meluncurkan Program Percepatan Belajar (PPB) atau lebih dikenal program akselerasi pada SD, SMP, dan SMA (Rusman, 2008 : 929).

Program akselerasi dilatarbelakangi oleh realitas hasil-hasil penelitian yang dilakukan oleh Balitbang Depdiknas(1986). Dari penelitian tersebut diperoleh temuan bahwa pada 20 SMA unggulan di Indonesia terdapat 21,75 % siswa dengan kecerdasan umum prestasinya di bawah rerata, sedangkan para siswa yang tergolong berkemampuan dan berkecerdasan luar biasa sebesar 9,7 %. Pada hasil temuan sebelumnya telah diungkapkan, bahwa masih terdapat siswa yang dikategorikan berbakat istimewa mengalami *underachiever* masih tinggi, yaitu menurut Depdikbud (1997) pada SD dan SMP sebesar 2-5 % dan SMA sebesar 8 %. Kemudian riset – riset independen juga menyebutkan demikian, seperti menurut Yaumil Achir (1990) pada SMA di DKI Jakarta ditemukan 39 % siswa mengalami *underachiever*, Yusuf dan Widyastono (1997) menemukan masih terdapat 13,5 % sampai 20% siswa SMP mengalami *underachiever* (Rusman, 2008 : 928-929).

Negara-negara maju dan berkembang seperti Amerika, Singapura, Cina, dan Korea sudah mulai melihat potensi anak – anak CIBI. Negara tersebut mulai menarik perhatian anak CIBI dengan memberikan beasiswa di perguruan tinggi yang bagus dan bahkan menjamin pekerjaan hingga usia 55 tahun, negara – negara tersebut yakin bahwa anak CIBI mempunyai tingkat kreativitas yang tinggi dan mempunyai komitmen serta kerja keras yang tinggi, ini merupakan keunggulan yang dibutuhkan oleh banyak negara di tengah semakin tingginya persaingan perekonomian. Demikian pentingnya pendidikan yang berdimensi keunggulan dalam menjawab tantangan masa depan, maka sangat beralasan apabila pengembangan CIBI di Indonesia perlu mendapatkan perhatian dari berbagai pihak, baik pemerintah, sekolah dan orang tua siswa, agar CIBI tidak mengalami *underachiever*. Seperti yang kita ketahui *underachiever* mengakibatkan tidak maksimalnya kemampuan yang CIBI miliki harapannya dimasa yang akan datang CIBI dapat mewakili bangsa Indonesia pada era globalisasi sekaligus dapat memenangkan persaingan *global* yang semakin tinggi.

Seperti halnya pada program sekolah reguler, pada sekolah akselerasi pun pembelajaran juga dilaksanakan melalui beberapa tahap, yaitu persiapan (*preparation*), implementasi (*implementation*), dan evaluasi (*evaluation*). Pada tahap persiapan (*preparation*) adalah tahapan dimana seorang guru mempersiapkan Bahan Ajar, Rencana Pelaksanaan Pembelajaran, Silabus dan Media Pembelajaran. Tahap Implementasi adalah tahapan penggunaan segala sesuatu yang sudah dipersiapkan guru pada tahap persiapan.

Sedangkan tahap evaluasi adalah tahapan dimana seorang guru melakukan penilaian terhadap hasil belajar peserta didik.

Menurut Nana Sudjana “dalam suatu proses pembelajaran ada 3 komponen penting di dalamnya, yaitu tujuan instruksional, pengalaman belajar (proses belajar mengajar), dan hasil belajar” (2006 : 2). Ketiga komponen tersebut bagaikan mata rantai yang tak terpisahkan satu sama lain. Suatu kegiatan belajar mengajar dikatakan berhasil dapat dilihat dari kesesuaian antara tujuan instruksional dan hasil belajar siswa, dan tentu saja harus melalui suatu pengalaman belajar (proses belajar mengajar).

Hasil belajar siswa dapat diketahui melalui suatu kegiatan penilaian. Kegiatan penilaian yakni suatu tindakan atau kegiatan untuk melihat sejauh mana tujuan – tujuan instruksional telah dapat dicapai dan dikuasai oleh siswa dalam bentuk hasil belajar yang diperilhatkannya setelah peserta didik menempuh pengalaman belajar (proses belajar mengajar) (Suharsimi, 2009:34). Untuk melakukan suatu kegiatan penilaian dan melihat apakah hasilnya sudah sesuai tujuan instruksional atau belum, maka diperlukan suatu instrumen tes. Instrumen tes diperlukan agar didapatkan suatu hasil penilaian yang memiliki akurasi tinggi dalam mengukur kemampuan siswa, oleh karena itu diperlukan suatu instrumen tes yang baku.

Instrumen tes baku adalah suatu instrumen tes yang telah melalui beberapa percobaan dan telah diuji akurasinya baik secara kualitatif maupun kuantitatif (Suharsimi, 2009 : 35). Pada umumnya penilaian hasil belajar di sekolah menggunakan tes buatan guru pada setiap bidang studinya. Sekolah jarang menggunakan tes baku karena meskipun tes baku lebih baik daripada tes buatan guru, namun jumlahnya di dunia pendidikan masih sangat jarang. Hal ini menyebabkan kurang akuratnya penilaian guru terhadap kemampuan siswa dalam memahami suatu materi. Serupa dengan hal tersebut pada program akselerasi, rata-rata guru juga memberikan soal evaluasi dengan menggunakan suatu instrumen tes yang belum teruji atau belum memenuhi standar baku suatu tes karena belum melalui serangkaian uji tes.

Selama ini kebanyakan guru baru mengira-ira tentang tingkat kesulitan soal yang diberikan, kebanyakan belum mempertimbangkan patokan tingkat kesulitan kognitif tes (C1-C6 Taksonomi Bloom). Pada program akselerasi yang pada pembelajarannya menuntut *high level thinking* menuntut pula guru memberikan soal dengan tingkat kesulitan C4 (analisis), C5 (sintesis) dan C6 (evaluasi) agar sesuai dengan tingkat kemampuan siswa tersebut (tidak mengalami *underachiever*) sekaligus juga dapat mengukur kemampuan siswa terhadap suatu materi dengan akurat (Depdiknas, 2009:55). Bisa jadi suatu tes yang dinilai gurunya akan sulit dikerjakan siswa, ternyata dianggap mudah oleh siswa yang diberikan tes tersebut. Tes yang mudah dan ditujukan kepada sekelompok subjek yang kemampuannya tinggi, tidak akan menghasilkan akurasi karena tidak sesuai dengan levelnya. Begitu juga sebaliknya, tes yang sulit tidak akan cocok untuk sekelompok subjek yang kemampuannya rendah. Jadi, tes yang baik adalah tes yang mampu mengukur tingkat kemampuan subjek sasaran

Pada program akselerasi penilaian yang digunakan dalam pendidikan khusus bagi CIBI adalah penilaian otentik (*Autentic Assement*), yaitu proses pengumpulan data yang bisa memberikan gambaran perkembangan belajar siswa (Depdiknas, 2009:56). Gambaran perkembangan belajar siswa perlu diketahui oleh guru agar bisa memastikan bahwa siswa mengalami proses pembelajaran dengan benar. Salah satu cara untuk melakukan penilaian otentik ini adalah dengan melakukan tes formatif. Apabila tes sudah dipersiapkan dan

dilaksanakan dengan secermat mungkin, maka informasi yang dihasilkan dapat menunjukkan sejauh mana tujuan-tujuan instruktusional yang telah ditetapkan itu tercapai.

Informasi yang dihasilkan dari suatu hasil tes dapat dijadikan balikan untuk meningkatkan dan menyempurnakan proses pembelajaran. Oleh karena itu, perlu adanya suatu pengembangan instrumen tes formatif untuk program akselerasi agar didapatkan suatu tes baku yang cocok untuk mengukur kemampuan siswa program akselerasi dengan karakter khas siswa CIBI dan siap pakai sehingga guru bisa menggunakan instrumen tes tersebut untuk mengevaluasi kemampuan siswa apabila guru belum membuat/memiliki instrumen tes yang baku, atau tes baku tersebut bisa juga dijadikan patokan (acuan) guru dalam membuat instrumen tes.

Dari uraian di atas didapatkan identifikasi masalah, yaitu : 1) Dibutuhkannya profesionalisme seorang guru dalam penyelenggaraan pendidikan dan evaluasi hasil belajar siswa. Evaluasi sangatlah penting karena merupakan bagian integral dari proses pembelajaran. 2) Perlunya instrumen tes yang baku dan siap pakai, untuk mengevaluasi hasil belajar siswa di sekolah Program Akselerasi. 3) Instrumen tes yang sesuai dengan *autentic assessment* dibutuhkan guna mendukung evaluasi pembelajaran di Program Akselerasi.

Dari identifikasi masalah yang muncul, maka dalam penelitian ini penulis membatasi masalah agar tujuan dalam penelitian ini dapat tercapai secara optimal. Adapun pembatasan masalah tersebut adalah sebagai berikut : 1) Penyusunan tes formatif Program Akselerasi sesuai dengan Standar Kompetensi, dan Kompetensi Dasar yang ada. 2) Penyusunan instrumen tes kelas XI tengah semester gasal pada materi Kinematika dengan Analisis Vektor, Gravitasi, dan Gerak Harmonik pada Benda Elastik. 3) Bentuk penulisan item tes adalah pilihan ganda. 4) Analisis hasil penulisan item tes secara kualitatif dan kuantitatif.

Masalah yang akan diteliti dalam penelitian ini adalah : 1) Bagaimana merancang tes formatif untuk kelas XI tengah semester gasal Sekolah Program Akselerasi untuk mendukung *autentic assesment*? 2) Bagaimana menyusun instrumen tes formatif pilihan ganda kelas XI tengah semester gasal? 3) Apakah instrumen tes formatif yang dirancang sesuai dengan kriteria kualitatif suatu instrumen tes yang baku? 4) Apakah instrumen tes formatif yang dirancang sesuai dengan kriteria kuantitatif suatu instrumen tes yang baku ?

Penelitian ini yang telah dilakukan bertujuan untuk : 1) Menyusun tes formatif pilihan ganda untuk kelas XI semester gasal Program Akselerasi yang sesuai dengan Standar Kompetensi, dan Kompetensi Dasar yang ada. 2) Menyusun kisi-kisi instrumen tes formatif pilihan ganda untuk kelas XI semester gasal materi Kinematika dengan Analisis Vektor, Gravitasi dan Gerak Harmonik pada Benda Elastik. 3) Menyusun item soal pilihan ganda untuk tes formatif kelas XI semester gasal Program Akselerasi untuk materi Kinematika dengan Analisis Vektor, Gravitasi dan Gerak Harmonik pada Benda Elastik. 4) Memvalidasi hasil penyusunan instrumen tes formatif kelas XI semester gasal Program Akselerasi untuk materi Kinematika dengan Analisis Vektor, Gravitasi dan Gerak Harmonik pada Benda Elastik secara kualitatif dan kuantitatif.

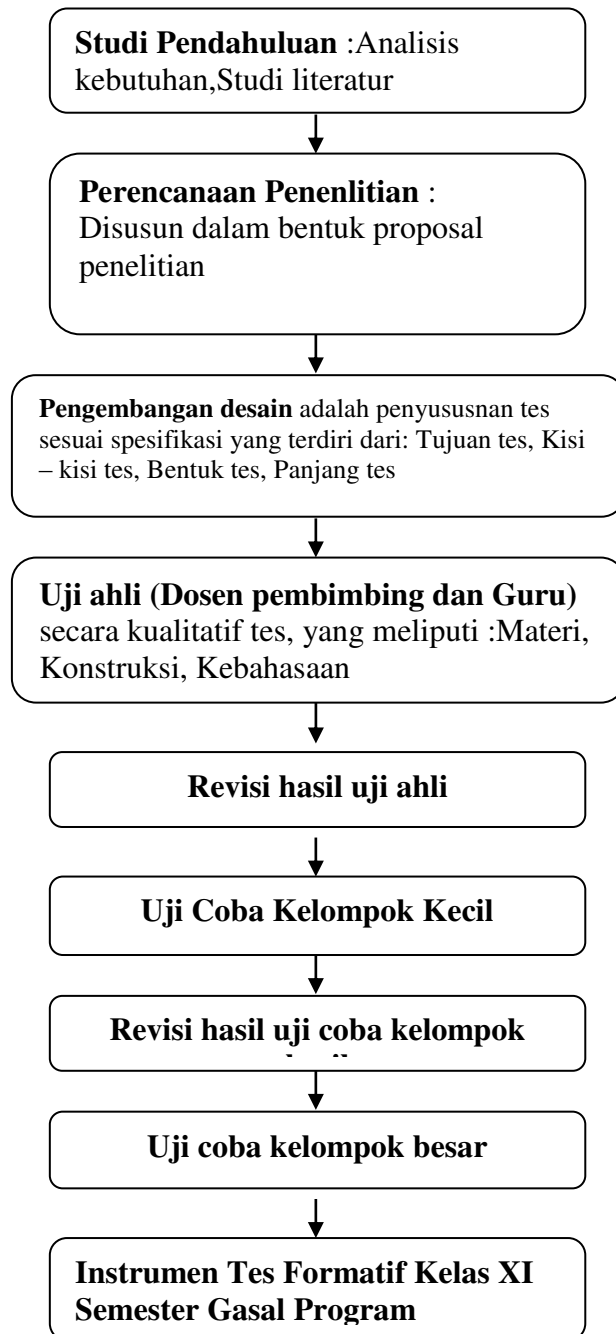
Dalam penelitian ini akan dikembangkan seperangkat instrumen tes untuk SMA Kelas XI semester gasal Program Akselerasi, dengan ketentuan sebagai berikut: 1) pengembangan instrumen tes tengah semester gasal berdasarkan pada Standar Kompetensi dan Kompetensi Dasar, 2) pengembangan tes berdasarkan indikator yang mengacu pada Standar Kompetensi dan Kompetensi Dasar, 3) tes formatif yang dapat mengukur

kemampuan kognitif siswa, 4) bentuk tes berupa tes pilihan ganda yang terdiri dari stem, key, dan distarktor, 5) tiap soal terdapat 5 options jawaban, 6) Kompetensi Dasar yang ada dijabarkan dalam materi Kinematika dengan Analisis Vektor, Gravitasi dan Gerak Harmonik pada Benda Elastik

Hasil penelitian ini diharapkan dapat memberikan manfaat baik bagi peneliti, bagi guru, bagi siswa, dan bagi sekolah. Bagi Peneliti manfaat yang bisa diambil adalah terpecahkannya masalah yang diteliti. Bagi Guru dapat menyajikan sebuah pilihan untuk mengatasi kesulitan dalam penyusunan instrumen tes terutama tes pilihan ganda kelas XI tengah semester gasal Program Akselerasi, dapat memberikan masukan tentang alternatif model pengembangan tes pembelajaran yang layak dan mampu mengukur ketercapaian indikator dari suatu pembelajaran Program Akselerasi, dapat membangkitkan kinerja guru dalam meningkatkan kualitas kegiatan evaluasi pembelajaran Program Akselerasi. Bagi Siswa dapat mempersempit lingkup belajar siswa karena tes yang diujikan mengacu pada indikator sesuai tahapan proses pembelajaran, dan dapat mengetahui hasil prestasi belajarnya secara bertahap. Selain itu penelitian ini juga bermanfaat bagi sekolah yaitu memberi masukan yang dapat digunakan untuk alternatif perbaikan kualitas pelaksanaan evaluasi pembelajaran pada masa yang akan datang.

METODE PENELITIAN

Penelitian ini menggunakan metode penelitian pengembangan *research and development* versi Borg and Gall (1989) yang dikutip oleh Nana Syaodih (2007 : 169) yang meliputi sepuluh langkah yaitu : (1) Studi Pendahuluan, (2) Perencanaan (*planning*), (3) Pengembangan draf produk (*develop preliminary form of product*), (4) Uji Produk Terbatas (*Preliminery Field Test*), (5) Revisi Hasil Uji Terbatas, (6) Uji Kelompok Kecil, (7) Revisi Hasil Uji Kelompok Kecil, (8) Uji Kelompok Besar, (9) Penyempurnaan produk akhir (*final product revision*) (10) Desiminasi dan Implementasi Produk Akhir. Pada penelitian ini dilakukan langkah 1-8 karena adanya keterbatasan peneliti. Delapan langkah penelitian tersebut dapat dijelaskan pada skema prosedur penelitian.



Gambar 1. Skema Prosedur Penelitian

INSTRUMEN PENGAMBILAN DATA

- Pengambilan data dalam penelitian dilakukan dengan 2 instrumen sebagai berikut:
- Instrumen non tes berupa daftar cek. Daftar cek ini digunakan dalam penelaahan butir soal. Daftar cek disini berisi pernyataan mengenai ranah materi, konstruksi, dan bahasa dari butir soal. Penelaah hanya memberi tanda (\checkmark) pada nomor soal yang sesuai dengan pernyataan yang diberikan.
 - Instrumen tes berupa soal tes formatif pilihan ganda, yang terdiri dari 3 paket tes. Paket 1 merupakan tes dengan materi Kinematika dengan Analisis Vektor yang terdiri dari 24 soal, Paket 2 merupakan tes dengan materi Gravitasi yang terdiri dari 15 soal,

Paket 3 merupakan tes dengan materi Gerak Harmonik pada Benda Elastik yang terdiri dari 25 soal.

UJI COBA INSTRUMEN TES

Uji Tes secara Kualitatif

Uji coba tes secara kualitatif dilaksanakan sebelum tes diujicobakan secara kuantitatif. Uji coba tes secara kualitatif dilakukan oleh ahli, yang menilai instrumen soal dari tiga aspek, yaitu aspek materi, konstruksi, bahasa dan validitas isi dari soal tes yang telah dibuat.

Pada penelitian ini, uji awal terhadap desain produk dilakukan oleh ahli, yakni dosen pembimbing dan guru bidang studi Fisika SMA Negeri 1 Karangnyar. Uji awal ini dilakukan dengan menganalisis secara kualitatif desain instrumen tes yang dibuat. Analisis ini dilakukan untuk memperbaiki soal tes yang telah disusun sehingga soal tes yang dibuat memiliki kualitas yang baik dilihat dari materi, konstruksi, dan bahasa. Dalam penelaahan butir soal ini digunakan lembar penelaahan berupa daftar cek.

Pada telaah dari para ahli ini poin pertama yang dikaji adalah materi. Soal yang telah dibuat sudah sesuai dengan materi yang terdapat dalam kisi-kisi atau belum. Pada telaah materi ini peran yang paling utama adalah guru bidang studi, karena guru bidang studi yang memberikan materi secara langsung. Dari telaah secara materi ini semua soal sudah sesuai dengan materi, karena sebelum dilakukan penyusunan soal dilakukan terlebih dahulu telaah materi berdasarkan silabus yang berasal dari diknas dan dari sekolah.

Pada telaah secara konstruksi dan bahasa dari soal yang dibuat. Pada konstruksi soal yang direvisi bervariasi penyebabnya. Secara konstruksi ada soal yang kurang lengkap dalam penulisannya, misalnya dalam menuliskan satuan atau ada pokok soal yang belum tercantum. Selain itu mengenai pilihan jawaban juga menjadi bahan revisi, pilihan jawaban pada awalnya belum merupakan hasil dari penggunaan rumus yang salah. Dengan konstruksi yang demikian pengecoh akan tidak berfungsi dan jawaban mudah ditebak. Oleh sebab itu, untuk pilihan jawaban dibuat dengan memperkirakan pilihan jawaban apabila terjadi kombinasi rumus yang salah, sehingga walaupun rumus yang digunakan salah, tetap terdapat opsi jawabannya.

Dalam hal bahasa yang ditelaah adalah tentang pemilihan kata, dan penulisan soalnya. Kebanyakan soal yang telah dibuat banyak menimbulkan penafsiran yang berbeda dengan maksud soal sebenarnya dan agak membingungkan siswa oleh karena itu perlu diperbaiki.

Uji Tes secara Kuantitatif

Uji coba tes secara kuantitatif dilakukan untuk mendapatkan data empiris kualitas soal dari hasil pekerjaan siswa. Uji coba tes secara kuantitatif dilakukan dua kali, yaitu uji coba kelompok kecil dengan subyek coba 24 siswa Program Akselerasi SMA N 1 Karangnyar dan uji coba kelompok besar dengan subyek coba 56 siswa Program Akselerasi SMA N 3 Surakarta. Data hasil pekerjaan siswa kemudian diolah dengan Program Microsoft Excel untuk mengetahui kualitas soal dilihat dari beberapa aspek yaitu :

1). Taraf kesukaran

Tingkat kesukaran adalah angka yang menunjukkan proporsi siswa yang menjawab betul suatu soal (Slameto, 2001:215). Makin besar tingkat kesukaran berarti soal itu makin mudah demikian juga sebaliknya yaitu makin rendah tingkat kesukaran berarti soal itu makin sukar. Menurut Suharsimi Arikunto (2001 : 207), "Soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sulit". Oleh karena itu, dapat dinyatakan bahwa soal yang baik adalah soal dengan taraf kesukaran yang sedang. Untuk

menentukan taraf kesukaran dari tiap-tiap item soal digunakan rumus $P = B/J_s$, dimana P merupakan indeks kesukaran, B merupakan banyaknya siswa yang menjawab soal betul, dan J_s merupakan jumlah seluruh siswa peserta tes

Taraf kesukaran soal dapat ditentukan berdasarkan hasil perhitungan indeks kesukaran dengan ketentuan soal sukar adalah soal yang $0,00 \leq P < 0,30$, soal dikatakan sedang jika $0,30 \leq P \leq 0,70$, soal dikatakan mudah jika $0,70 < P \leq 1,00$ (Allen & Yen, 1979 : 121)

2). Daya Beda

Daya pembeda item adalah kemampuan suatu butir item tes hasil belajar untuk dapat membedakan antara testee yang berkemampuan tinggi dengan testee yang kemampuannya rendah demikian rupa sehingga sebagian besar testee yang memiliki kemampuan yang tinggi untuk menjawab butir item tersebut lebih banyak menjawab butir item tersebut lebih banyak yang menjawab betul, sementara testee yang kemampuannya rendah untuk menjawab butir item tersebut sebagian besar tidak dapat menjawab item dengan betul (Anas sudijono, 2001).

Dalam menghitung daya pembeda setiap soal, dapat digunakan rumus sebagai berikut : $D = B_A/J_A - B_B/J_B = P_A - P_B$ dimana J merupakan jumlah peserta tes, J_A merupakan banyaknya siswa kelompok atas, J_B merupakan banyaknya siswa kelompok bawah, B_A merupakan banyaknya siswa kelompok atas yang menjawab benar, B_B merupakan banyaknya siswa kelompok bawah yang menjawab benar, P_A merupakan proporsi siswa kelompok atas yang menjawab benar dan P_B merupakan proporsi siswa kelompok bawah yang menjawab benar. Daya pembeda soal (nilai D) diklasifikasikan sebagai berikut menurut Djemari(2005 :5) yaitu soal dinyatakan ditolak jika $D < 0,1$, soal direvisi jika $0,1 \leq D \leq 0,3$, soal diterima jika $D > 0,3$.

3). Penyebaran (Distribusi) Jawaban

Penyebaran pilihan jawaban dijadikan dasar dalam penelaahan soal. Hal ini dimaksudkan untuk mengetahui berfungsi tidaknya jawaban yang tersedia. Soal pilihan jawaban (pengecoh) dapat dikatakan berfungsi apabila pengecoh paling tidak dipilih oleh 5 % peserta tes atau siswa dan lebih banyak dipilih oleh kelompok siswa yang yang belum paham materi (Depdiknas, 2009: 14)

4). Reliabilitas

Reliabilitas suatu tes adalah kemampuan suatu tes untuk memberikan hasil yang relatif ajeg atau tetap bila digunakan pada waktu atau tempat yang berlainan. Untuk menghitung reliabilitas digunakan rumus yang dikemukakan oleh Kuder dan Richardson (rumus KR-20) sebagai berikut :

$$r_{11} = \left[\frac{n}{n-1} \right] \left[\frac{S^2 - \sum pq}{S^2} \right]$$

Dimana r_{11} merupakan reliabilitas tes secara keseluruhan, p merupakan proporsi subyek yang menjawab item dengan benar, q merupakan proporsi subyek yang menjawab item dengan salah ($q = 1-p$), $\sum pq$ merupakan jumlah hasil perkalian antara p dan q , n merupakan banyaknya item, dan S merupakan standar deviasi dari tes (Suharsimi Arikunto, 2001 : 100-101)

Hasil perhitungan tingkat reliabilitas tersebut kemudian dikonsultasikan dengan tabel *r product moment*. Apabila harga $r_{hitung} > r_{tabel}$, maka dapat ditarik kesimpulan bahwa instrumen tes reliabel. Selain itu, terdapat beberapa kriteria nilai reliabilitas sebagai berikut, yaitu soal dikatakan memiliki reliabilitas sangat tinggi apabila $0,800 \leq r_{11} < 1,00$, soal dikatakan memiliki reliabilitas tinggi apabila $0,600 \leq r_{11} < 0,800$, soal

memiliki realibilitas cukup apabila $0,400 \leq r_{11} < 0,600$, soal memiliki realibilitas rendah apabila $0,200 \leq r_{11} < 0,400$, soal memiliki realibilitas sangat rendah apabila $0,000 \leq r_{11} < 0,200$ (Slameto, 2001:215).

HASIL PENELITIAN DAN PEMBAHASAN

Model pengembangan yang dipakai dalam penelitian ini adalah model prosedural yaitu model yang bersifat deskriptif, menunjukkan langkah-langkah yang harus diikuti untuk menghasilkan produk berupa instrument tes untuk evaluasi pembelajaran. Langkah langkah yang dilakukan dalam pengembangan instrumen tes ini adalah sebagai berikut :

1. Studi Pendahuluan

Langkah ini adalah langkah awal yang harus dilakukan dalam pengembangan tes. Langkah ini meliputi:

a. Analisis Kebutuhan

Permasalahan pada penelitian dan pengembangan ini adalah masih sedikitnya guru yang menggunakan instrumen evaluasi yang berkualitas baik. Instrumen tes yang valid diperlukan agar mampu mengukur kemampuan penguasaan materi siswa dengan akurat. Masih minimnya guru yang menggunakan instrumen tes akurat, disebabkan karena untuk mendapatkan suatu instrumen yang akurat diperlukan banyak sekali pengujian baik secara kualitatif maupun secara kuantitatif yang cukup memakan waktu.

Penelitian menemukan masalah pada saat melakukan observasi di beberapa Sekolah Menengah Atas Program Akselerasi, yaitu SMA Negeri 1 Surakarta, SMA Negeri 3 Surakarta, dan SMA Negeri 1 Karanganyar. Wawancara dilakukan dengan bapak Sriyanto, S.Pd selaku guru Fisika Program Akselerasi SMA Negeri 1 Surakarta, Bapak Drs. Subandrio selaku guru Fisika Program Akselerasi SMA N 3 Surakarta dan dengan Bapak Brata, M.Pd selaku guru Fisika Program Akselerasi SMA N 1 Karanganyar. Dokumen hasil observasi dapat dilihat pada Lampiran 3. Rata-rata guru memberikan soal evaluasi dengan menggunakan suatu instrumen tes buatan guru sendiri yang belum teruji atau belum memenuhi standar baku suatu tes karena belum melalui serangkaian uji tes. Selama ini kebanyakan guru menggunakan soal yang dinilai sulit oleh guru tersebut tanpa melihat tingkat kesulitan kognitif tes (C1-C6 Taksonomi Bloom), padahal sesuai pedoman dari Depdiknas (2009 : 55) untuk Program Akselerasi yang pada pembelajarannya menuntut *high level thinking* menuntut pula guru memberikan soal dengan tingkat kesulitan C4 (analisis), C5 (sintesis) dan C6 (evaluasi). Hal ini diperlukan agar terjadi keseimbangan antara kecerdasan siswa, proses pembelajaran, dan evaluasinya.

b. Studi Literatur

Studi literatur dilakukan untuk pengenalan sementara terhadap produk yang akan dikembangkan. Studi literatur pada penelitian ini dilakukan dengan studi teoritik yakni mengkaji teori-teori yang relevan sehingga dapat digunakan sebagai dasar pengembangan. Studi literatur juga diperlukan untuk mengetahui langkah-langkah yang paling tepat dalam pengembangan suatu produk. Suatu produk pendidikan kemungkinan bukan hal yang sama sekali baru. Produk sejenis atau produk yang mirip telah dikembangkan oleh pengembang lain di tempat lain. Hal-hal tersebut dikaji melalui studi literatur berbentuk dokumen-dokumen hasil penelitian atau hasil evaluasi tentang pengembangan instrumen tes formatif, tentang asesmen, dan juga tentang sekolah program Akselerasi.

2. Merencanakan Penelitian

Setelah melakukan studi pendahuluan, peneliti dapat melanjutkan langkah kedua, yaitu merencanakan penelitian. Perencanaan penelitian ini disusun dalam bentuk proposal penelitian yang meliputi kemampuan-kemampuan yang diperlukan dalam pelaksanaan penelitian, rumusan yang hendak dicapai dengan penelitian tersebut, desain atau langkah-langkah penelitian. Perencanaan penelitian ini dilakukan dengan arahan dari pembimbing dan beberapa ahli pada saat dilaksanakannya seminar proposal.

3. Pengembangan Desain

Langkah pengembangan desain produk pendidikan dalam penelitian ini menggunakan alur pengembangan tes menurut Djemari Mardapi (2004:88) dalam menyusun instrumen tes yaitu :

a. Menyusun spesifikasi tes

Langkah awal dalam mengembangkan tes adalah menetapkan spesifikasi tes yang berisi tentang uraian yang menunjukkan keseluruhan karakteristik yang harus dimiliki suatu tes. Penyusunan spesifikasi tes mencakup kegiatan berikut :

1) Menentukan tujuan tes

Ditinjau dari tujuannya ada empat macam tes yang banyak digunakan di lembaga pendidikan yaitu tes penempatan, tes diagnostik, tes formatif, dan tes sumatif. Pada penelitian ini , tes yang dikembangkan adalah tes formatif. Tes formatif pada umumnya dilakukan pada akhir setiap Satuan Acara Pelajaran (SAP) dan terutama diarahkan pada ranah kognitif siswa. Dari hasil penilaian menggunakan tes formatif guru akan melihat sejauh mana siswa telah memahami suatu materi pelajaran yang diajarkan guru, dalam penelitian ini akan dilakukan penyusunan (pengembangan) instrumen tes formatif Fisika untuk kelas XI semester gasal Program Akselerasi dengan spesifikasi materi Kinematika dengan Analisis Vektor, Gravitasi, dan Gerak Harmonik pada Benda Elastik. Tes formatif dipilih karena dengan tes formatif dapat mendukung terlaksananya *autentic assesment* , selanjutnya mampu memberikan informasi rekam jejak siswa dalam penguasaan materi.

2) Penulisan kisi – kisi soal

Kisi – kisi merupakan tabel matrik yang berisi spesifikasi soal-soal yang akan dibuat. Kisi – kisi ini merupakan acuan bagi penulis soal, sehingga siapapun yang menulis soal akan menghasilkan soal yang memiliki isi dan tingkat kesulitan relatif sama. Berikut ini adalah langkah – langkah dalam mengembangkan kisi – kisi tes, yaitu (a) Menentukan Standar Kompetensi (SK), Kompetensi Dasar (KD), materi pokok yang akan diujikan (b) Menentukan indikator (c) Menentukan jumlah soal tiap pokok bahasan dan sub pokok bahasan . Dalam pembuatan kisi-kisi dilakukan konsultasi dengan dosen pembimbing dan guru sekolah Program Akselerasi kelas XI tempat dilakukannya penelitian.

3) Menentukan bentuk tes

Pemilihan bentuk tes yang tepat ditentukan sesuai tujuan tes, jumlah peserta tes, waktu yang tersedia untuk memeriksa lembar jawaban tes, cakupan materi tes, dan karakteristik mata pelajaran yang diujikan. Penelitian ini menggunakan bentuk tes objektif pilihan ganda karena jumlah peserta tes banyak, sehingga waktu koreksi lebih singkat dan cakupan materinya lebih menyeluruh, selain itu dari pihak sekolah hanya memberikan sedikit waktu untuk melakukan uji tes pada siswa mengingat waktu Program Akselerasi yang terbatas namun dengan materi yang banyak, yaitu materi 1 semester hanya dilaksanakan 4 bulan.

4) Menentukan panjang tes

Penentuan panjang tes berdasarkan pada cakup materi ujian dan kelelahan peserta tes. Pada umumnya tes tertulis menggunakan waktu 90-150 menit. Pada umumnya waktu yang dibutuhkan untuk mengerjakan tes bentuk pilihan ganda adalah 2-3 menit untuk tiap butir soal. Pada penelitian ini dibuat tes sebanyak 25 butir pada materi Kinematika dengan waktu 75 menit, 15 butir pada materi Gravitasi dengan waktu 45 menit, dan 25 butir pada materi Gerak Harmonis pada Benda Elastik dengan waktu 75 menit.

b. Menulis soal tes

Penulisan tes dilakukan setelah langkah pertama, yaitu menyusun spesifikasi tes, dilakukan. Penulisan soal merupakan langkah menjabarkan indikator menjadi pertanyaan – pertanyaan yang karakteristiknya sesuai dengan perincian pada kisi-kisi yang telah dibuat.

4. Melakukan Uji Kualitatif

Pada penelitian ini, uji awal terhadap desain produk dilakukan oleh ahli, yakni dosen pembimbing dan guru bidang studi Fisika SMA Negeri 1 Karangnyar. Uji awal ini dilakukan dengan menganalisis secara kualitatif desain instrumen tes yang dibuat. Analisis ini dilakukan untuk memperbaiki soal tes yang telah disusun sehingga soal tes yang dibuat memiliki kualitas yang baik dilihat dari materi, konstruksi, dan bahasa. Dalam penelaahan butir soal ini digunakan lembar penelaahan berupa daftar cek.

Pada telaah dari para ahli ini poin pertama yang dikaji adalah materi. Soal yang telah dibuat sudah sesuai dengan materi yang terdapat dalam kisi-kisi atau belum. Pada telaah materi ini peran yang paling utama adalah guru bidang studi, karena guru bidang studi yang memberikan materi secara langsung. Dari telaah secara materi ini semua soal sudah sesuai dengan materi, karena sebelum dilakukan penyusunan soal dilakukan terlebih dahulu telaah materi berdasarkan silabus yang berasal dari diknas dan dari sekolah.

Pada telaah secara konstruksi dan bahasa dari soal yang dibuat. Pada konstruksi soal yang direvisi bervariasi penyebabnya. Secara konstruksi ada soal yang kurang lengkap dalam penulisannya, misalnya dalam menuliskan satuan atau ada pokok soal yang belum tercantum. Selain itu mengenai pilihan jawaban juga menjadi bahan revisi, pilihan jawaban pada awalnya belum merupakan hasil dari penggunaan rumus yang salah. Dengan konstruksi yang demikian pengecoh akan tidak berfungsi dan jawaban mudah ditebak. Oleh sebab itu, untuk pilihan jawaban dibuat dengan memperkirakan pilihan jawaban apabila terjadi kombinasi rumus yang salah, sehingga walaupun rumus yang digunakan salah, tetap terdapat opsi jawabannya.

Dalam hal bahasa yang ditelaah adalah tentang pemilihan kata, dan penulisan soalnya. Kebanyakan soal yang telah dibuat banyak menimbulkan penafsiran yang berbeda dengan maksud soal sebenarnya dan agak membingungkan siswa oleh karena itu perlu diperbaiki.

Dari hasil telaah kualitatif di dapatkan hasil sesuai Tabel 4.2 sebagai berikut :

Tabel 1. Hasil Telaah Kualitatif Desain Soal

Instrumen tes	Soal yang harus direvisi
Kinematika dengan Analisis Vektor (Paket 1)	1,7,10,13,16,17,24
Gravitasi (Paket 2)	8,11,14,15

5. Melakukan Revisi Soal Hasil Telaah Kualitatif

Revisi instrumen tes hasil telaah kualitatif dilakukan dengan berdasarkan koreksi dari ahli yang mencakup aspek materi, konstruksi, dan bahasa. Dalam pelaksanaannya revisi instrumen tes ini dilakukan dengan saran dari para ahli. Revisi akan terus dilakukan sampai para ahli menilai instrumen tes tersebut layak untuk diujicobakan. Instrumen tes hasil revisi dan merupakan soal uji kelompok kecil dapat dilihat di Lampiran 6.

6. Melaksanakan Uji Coba Kelompok Kecil

Uji coba kelompok kecil dilakukan setelah instrumen tes yang telah dibuat mendapat persetujuan untuk diujicobakan dari para ahli. Ujicoba ini dilakukan untuk mendapatkan data empiris awal tentang kualitas tes. Soal yang dibuat ini menggunakan teori tes klasik yaitu tes yang hasil tesnya tergantung pada kemampuan siswanya. Instrumen tes pada saat diujikan pada siswa yang kemampuannya rendah maka taraf kesukaran soal seharusnya bernilai rendah/soal tergolong sulit dan sebaliknya apabila soal diujikan pada siswa dengan kemampuan yang tinggi maka taraf kesukaran soal seharusnya menjadi tinggi/soal tergolong mudah. Dalam pembuatan soal ini dispesifikkan untuk mengevaluasi pemahaman materi siswa program akselerasi, sehingga kemampuan siswanya homogen yang tergolong siswa CIBI. Untuk tes formatif digunakan tes dengan tingkat kesukaran sedang. Soal yang telah dibuat, dirancang dengan jenjang tingkat kemampuan kognitif siswa C4-C5-C6 (*high level thinking*). Ujicoba kelompok kecil dilakukan di SMA Negeri 1 Karanganyar dengan subyek coba 24 siswa kelas XI Program Akselerasi.

Setelah ujicoba tes kelompok kecil dilakukan, hasil tes kemudian dianalisis secara kuantitatif menggunakan program *microsoft excel* untuk mengetahui reliabilitas tes, tingkat kesukaran item soal, daya beda soal, dan efektifitas distraktor tiap item soal. Dari analisis kuantitatif ujicoba kelompok kecil di SMA Negeri 1 Karanganyar didapatkan hasil sebagai berikut :

a. Reliabilitas Instrumen tes

Reliabilitas dihitung dengan rumus Kuder Richardson 20 atau KR-20. Kriteria realibilitas yang digunakan dalam menganalisis instrumen tes yang dibuat berdasarkan kriteria reliabilitas dari Slameto (2001:215) yaitu kriteria sangat tinggi untuk reliabilitas $0,800 \leq r_{11} < 1,00$, tinggi untuk reliabilitas $0,600 \leq r_{11} < 0,800$, cukup tinggi untuk reliabilitas $0,400 \leq r_{11} < 0,600$, rendah untuk reliabilitas $0,200 \leq r_{11} < 0,400$, dan sangat rendah untuk reliabilitas $0,000 \leq r_{11} < 0,200$. Semakin tinggi reliabilitas suatu tes, menunjukkan bahwa tes tersebut semakin *ajeg* dalam mengukur kemampuan siswa, artinya tes tersebut memberikan hasil yang relatife tidak berbeda apabila di lakukan tes pada subyek yang sama meskipun dalam waktu yang berbeda. Hasil analisis reliabilitas uji kelompok kecil dapat dilihat pada Tabel 2.

Tabel 2. Hasil Analisis Reliabilitas Instrumen Tes Uji Kelompok Kecil.

Instrumen Tes	Reliabilitas	Kriteria Reliabilitas
Kinematika dengan Analisis Vektor (Paket 1)	0.69398	Tinggi
Gravitasi (Paket 2)	0.62755102	Tinggi
Gerak Harmonik pada Benda Elastik (Paket 3)	0.769453	Tinggi

Dari hasil analisis reliabilitas tes didapatkan hasil instrumen tes paket 1 memiliki nilai reliabilitas 0.69398 atau kriteria tinggi, paket 2 memiliki nilai reliabilitas 0.62755102 atau kriteria tinggi, dan paket 3 memiliki nilai 0.769453 atau kriteria tinggi. Ketiga paket instrumen tes yang dibuat memiliki kriteria tinggi yang berarti memiliki konsistensi yang tinggi dalam mengukur kemampuan siswa yang sama, meskipun diujikan dalam waktu yang berbeda.

b. Tingkat Kesukaran

Tingkat kesukaran adalah angka yang menunjukkan proporsi siswa yang menjawab betul suatu soal (Slameto, 2001:215). Hasil analisis kuantitatif instrumen tes dari hasil uji coba kelompok kecil dapat dilihat pada Tabel 3.

Tabel 3.. Hasil Analisis Tingkat Kesukaran Instrumen Tes.

No	Instrumen Tes		Kriteria Tingkat Kesukaran	Nomer Soal	Jumlah
1	Kinematika dengan Analisis Vektor (Paket 1)		Sukar	19,20	2
			Sedang	2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,21,22,23,24	21
			Mudah	1	1
2.	Gravitasi (Paket 2)		Sukar	-	0
			Sedang	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15	15
			Mudah	-	0
3.	Gerak Harmonik pada Benda Elastik (Paket 3)		Sukar	18	1
			Sedang	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,19,20,21,22,23,24,25	24
			Mudah	-	0

Dari hasil analisis instrumen tes secara kuantitatif didapatkan hasil bahwa untuk materi Paket 1 terdapat 2 soal tergolong mudah, 21 soal tergolong sedang, dan 1 soal tergolong mudah. Pada Paket 2 semua soal tergolong sedang. Pada Paket 3 terdapat 1 soal tergolong sukar, 24 soal tergolong sedang, dan tidak terdapat soal yang tergolong mudah. Hasil analisis tingkat kesukaran tes ini kemudian akan dipadukan dengan hasil analisis daya pembeda item tes, dan efektivitas distraktor untuk melihat kualitas tiap-tiap soal, setelah itu akan diambil keputusan mana soal yang diterima, direvisi, dan ditolak.

Dalam pembuatan instrumen tes ini, soal yang akan dipertahankan adalah soal dengan tingkat kesukaran sedang. Hal ini disebabkan karena tujuan dari pembuatan instrumen tes ini adalah untuk keperluan pengukuran hasil belajar (kompetensi) sehingga soal yang dianggap baik adalah yang termasuk kategori sedang. Keputusan ini diambil berdasarkan teori Allen & Yen (1979 :121) dalam Elvin & Surantoro (2010:185) yang menyatakan "...soal yang terlalu atau cenderung mudah lebih tepat digunakan untuk diagnostik, sedangkan soal yang terlalu sulit atau cenderung sulit lebih tepat digunakan untuk tes

seleksi. Oleh karena itu untuk keperluan tes yang mengukur hasil belajar (kompetensi) siswa tertentu akan dianggap baik bila termasuk dalam interval soal yang sedang “.

c. Daya Beda

Daya beda soal adalah kemampuan suatu soal untuk membedakan antara siswa yang mempunyai kemampuan tinggi dengan siswa yang mempunyai kemampuan rendah.

Hasil analisis daya beda instrumen tes ini dapat dilihat pada Tabel 4.

Tabel 4. Hasil Analisis Daya Beda Instrumen Tes

No	Instrumen Tes	Kriteria	Nomer Soal	Jumlah
Daya Beda				
1	Kinematika dengan Analisis Vektor (Paket 1)	Diterima	2,3,4,5,6,7,9,11,14,16, 21	11
		Direvisi	1,8,12,17,18,20,22,23,24	9
		Ditolak	10,13,15,19	4
2.	Gravitasi (Paket 2)	Diterima	1,2,3,4,5,6,7,11,12,13,14, 15	12
		Direvisi	8,9,10	3
		Ditolak	-	-
3.	Gerak Harmonik pada Benda Elastik (Paket 3)	Diterima	1,3,4,5,6,8,9,11,12,13,14, 15,16,17,18,19,20,21,22, 23,24,25	22
		Direvisi	2,10,	2
		Ditolak	7	1

Dari hasil analisis kuantitatif terdapat 11 soal diterima, 9 soal direvisi, dan 4 soal ditolak pada paket 1. Pada paket 2 terdapat 12 soal diterima dan 3 soal direvisi, sedangkan pada paket 3 terdapat 22 soal diterima, 2 soal direvisi, dan 1 soal ditolak. Hasil analisis daya beda ini kemudian akan dianalisis dengan tingkat kesukaran, efektifitas distraktor, dan reliabilitasnya untuk mengetahui keterterimaannya.

d. Efektifitas Distraktor

Efektifitas distraktor merupakan seberapa baik pilihan jawaban salah yang tersedia dapat mengecoh peserta tes yang tidak mengetahui kunci jawaban yang tersedia. Semakin banyak suatu jawaban salah dipilih maka semakin baik pula pengecoh menjalankan tugasnya.

Pada uji kelompok kecil ini peserta tes ada 24 orang siswa, jadi 5 % dari 24 siswa adalah 1.2, berarti dalam tes ini suatu distraktor dapat dikatakan efektif apabila dipilih minimal dua orang. Distraktor yang baik terutama dipilih dari kelompok bawah, akan tetapi distraktor kurang maksimal jika hanya dipilih oleh kelompok bawah saja, sebaiknya dipilih pula oleh kelompok atas. Distraktor sebaiknya dipilih lebih banyak oleh kelompok

bawah, jika dipilih lebih banyak oleh kelompok atas maka berarti distraktor tersebut menyesatkan dan sebaiknya diganti dengan distraktor lain. Distraktor dikatakan berfungsi apabila semua distraktor pada tiap soal berfungsi baik, apabila dalam satu soal ada distraktor yang belum berfungsi, atau menyesatkan maka distraktor pada soal tersebut dikatakan belum maksimal berfungsi maka distraktor tersebut sebaiknya direvisi. Tabel 5 menunjukkan hasil rekapitulasi efektifitas distraktor.

Tabel 5. Rekapitulasi Hasil Analisis Efektifitas Distraktor

No	Instrumen Tes	Kriteria Efektifitas Distraktor	Nomer Soal	Jumlah
1	Kinematika dengan Analisis Vektor (Paket 1)	Berfungsi	2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24	23
		Belum maksimal	1	1
2.	Gravitasi (Paket 2)	Berfungsi	1,2,3,5,6,7,8,10,11,12,13,14,15	13
		Belum maksimal	4,9	2
3.	Gerak Harmonik pada Benda Elastik (Paket 3)	Berfungsi	1,3,4,6,10,11,12,14,15,16,17,18,19,20,21,22,23,24,25	19
		Belum maksimal	2,5,7,8,9,13	6

Dari 64 soal yang telah dibuat terdapat 55 soal yang sudah memiliki distraktor yang telah berfungsi baik, dan 9 soal yang distraktornya belum maksimal. Soal yang distraktornya belum maksimal sebaiknya direvisi sebelum digunakan untuk ujicoba selanjutnya.

Setelah didapatkan hasil analisis reliabilitas, tingkat kesukaran, daya beda, dan efektifitas distraktor, kemudian dari masing-masing soal diputuskanlah soal yang sudah baik, dan soal yang perlu direvisi, atau soal mana yang ditolak sehingga perlu diganti dengan soal yang baru. Menurut Elvin & Surantoro (2010 :187) suatu soal dapat dimasukkan dalam kriteria soal diterima, direvisi, atau ditolak jika memenuhi kriteria keputusan untuk penilaian item soal sebagai berikut : (1) item soal diterima, apabila karakteristik item soal memenuhi semua kriteria. Item soal yang terlalu sukar atau terlalu mudah, tetapi memiliki daya beda dan distribusi pengecoh item yang memenuhi kriteria, butir soal tersebut dapat diterima atau dipilih; (2) item soal direvisi, apabila salah satu atau lebih dari ketiga kriteria karakteristik item soal tidak memenuhi kriteria ; (3) item soal ditolak, jika item soal memiliki karakteristik yang tidak memenuhi semua kriteria.

Dari 64 soal yang telah dibuat didapatkan 40 soal dikatakan berkualitas baik atau diterima dan 24 soal direvisi. Perinciannya dapat dilihat pada Tabel 4.6 .

Tabel 6. Rincian Hasil Analisis Kuantitatif Uji Kelompok Kecil

Instrumen tes	Kriteria	Nomer Soal	Jumlah
Kinematika dengan Analisis Vektor (Paket 1)	Soal yang diterima	2,5,6,8,9,11,14,16,22,23	10
	Soal yang direvisi	1,3,4,7,10,12,13,15,17,18,19,20,21,24	14
	Soal yang ditolak	-	-
Gravitasi (Paket 2)	Soal yang	1,2,3,5,6,7,11,12,13,14,	11

	diterima	15	
	Soal yang direvisi	4,8,9,10	4
	Soal yang ditolak	-	-
Gerak Harmonik pada Benda Elastik (Paket 3)	Soal yang diterima	1,3,4,6,10,11,12,14,15,16,17,18,19,20,21,22,23,24,25	19
	Soal yang direvisi	2,5,7,8,9,13	6
	Soal yang ditolak	-	-

7. Melaksanakan Revisi Hasil Uji Kelompok Kecil

Setelah dilakukan uji kelompok kecil dan dilakukan analisis secara kuantitatif mengenai reliabilitas, tingkat kesukaran, daya pembeda, dan efektifitas distraktor kemudian dilakukanlah revisi terhadap soal yang belum baik kualitasnya. Soal yang harus direvisi berjumlah 24 item. Revisi soal dilakukan dengan terlebih dahulu melakukan analisis kemungkinan faktor yang menyebabkan soal harus direvisi. Revisi soal dilakukan dengan masukan dari ahli evaluasi, ahli materi, dan guru Fisika SMA Negeri 3 Surakarta. Dalam pembuatan instrumen ini tidak menggunakan ahli bahasa secara khusus karena aspek kebahasaan soal sudah sekaligus masuk dalam penelaahan ketiga ahli tersebut. Revisi soal selesai setelah para ahli menilai soal yang direvisi layak untuk digunakan untuk ujicoba selanjutnya.

8. Melaksanakan Uji Coba Kelompok Besar

Uji coba kelompok besar dilakukan di kelas XI Program Akselerasi SMA Negeri 3 Surakarta dengan subyek coba 56 siswa. Uji coba kelompok besar ini hanya dilakukan di SMA N 3 Surakarta, karena di SMA N 3 Surakarta masih digunakan susunan materi yang sama sesuai dengan pedoman untuk kelas XI dari Depdiknas. Tidak semua sekolah Program Akselerasi masih mengikuti pedoman susunan materi dari Depdiknas, ada materi yang seharusnya diajarkan di kelas XI tetapi diajarkan di kelas X. Hal ini menyebabkan kesulitan menentukan waktu ujicoba jika melihat jadwal siswa akselerasi sendiri yang sudah sangat padat selain itu juga kurang maksimal karena materi tersebut sudah sangat lama dipelajari oleh siswa, kemungkinan besar banyak siswa yang sudah lupa dengan materi tersebut. Seandainya tetap dilakukan ujicoba bisa dilakukan di siang hari setelah pulang sekolah, dan siswa terlebih dahulu diberi tahu bahwa akan dilakukan tes, tetapi kemungkinan besar *mood* siswa untuk mengerjakan akan berkurang karena harus belajar lagi materi yang bukan menjadi beban belajar di kelas XI.

Uji coba kelompok besar dilakukan untuk mendapatkan data empiris tentang kualitas soal, setelah melalui beberapa tahap. Berikut ini adalah hasil analisis uji kelompok besar secara kuantitatif soal :

a. Reliabilitas Instrumen Tes

Sama halnya dengan analisis kuantitatif pada uji coba kelompok kecil reliabilitas dihitung dengan rumus Kuder Richardson 20 atau KR-20. Hasil analisis reliabilitas uji kelompok besar dapat dilihat pada Tabel 7.

Tabel 7 Hasil Analisis Reliabilitas Instrumen Tes Uji Kelompok Kecil.

Instrumen Tes	Reliabilitas	Kriteria Reliabilitas
Kinematika dengan Analisis Vektor (Paket 1)	0.810701	Sangat tinggi
Gravitasi (Paket 2)	0.6844	Tinggi
Gerak Harmonik pada Benda Elastik (Paket 3)	0.824764	Sangat tinggi

Dari hasil analisis reliabilitas tes didapatkan hasil instrumen tes paket 1 memiliki nilai reliabilitas 0.810701 atau kriteria sangat tinggi, paket 2 memiliki nilai reliabilitas 0.6844 atau kriteria tinggi, dan paket 3 memiliki nilai 0.824764 atau kriteria sangat tinggi. Paket instrumen tes yang dibuat memiliki kriteria reliabilitas yang berbeda paket 1, dan 3 dengan kriteria sangat tinggi, sedangkan paket 2 memiliki kriteria tinggi dari kriteria-kriteria ini dapat disimpulkan bahwa instrument tes yang dibuat memiliki konsistensi yang tinggi dalam mengukur kemampuan siswa yang sama, meskipun diujikan dalam waktu yang berbeda.

b. Tingkat Kesukaran

Tingkat kesukaran item tes dihitung dengan menggunakan rumus Du Bois, yaitu $P = \frac{B}{J_s}$. Hasil analisis kuantitatif tingkat kesukaran instrumen tes dari hasil uji coba kelompok besar dapat dilihat pada Tabel 8.

Tabel 8. Hasil Analisis Tingkat Kesukaran Instrumen Tes.

No	Instrumen Tes	Kriteria Tingkat Kesukaran	Nomer Soal	Jumlah
1	Kinematika dengan Analisis Vektor (Paket 1)	Sukar	-	0
		Sedang	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24	24
		Mudah	-	0
2.	Gravitasi (Paket 2)	Sukar	-	0
		Sedang	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15	15
		Mudah	-	0
3.	Gerak Harmonik pada Benda Elastik (Paket 3)	Sukar	-	0
		Sedang	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25	25
		Mudah	-	0

Dari hasil analisis instrumen tes secara kuantitatif didapatkan hasil bahwa tuntut materi Paket 1 24 soal tergolong sedang dan tidak ada soal yang tergolong sukar atau mudah. Pada Paket 2 tidak ada soal yang tergolong mudah atau sukar. Pada Paket 3 semua soal tergolong sedang. Dari tingkat kesukaran item tes ini semua soal tergolong kriteria sedang, hal ini menunjukkan bahwa item tes yang dibuat sudah tepat jika ditujukan untuk mengukur hasil belajar siswa.

c. Daya Beda

Daya beda instrumen tes ini dilakukan dengan rumus indeks diskriminasi yaitu dengan menghitung perbedaan proporsi kelas atas yang menjawab benar dengan proporsi kelas bawah yang menjawab benar. Hasil analisis daya beda instrumen tes ini dapat dilihat pada Tabel 8

Tabel 8. Hasil Analisis Daya Beda Instrumen Tes

No	Instrumen Tes	Kriteria Daya Beda	Nomer Soal	Jumlah
1	Kinematika dengan Analisis Vektor (Paket 1)	Diterima	1,2,3,4,5,6,7,8,9,10,11,12, 13,14,15,16,17,18,19,20,2 1,22, 23, 24	24
		Direvisi	-	0
		Ditolak	-	0
2.	Gravitasi (Paket 2)	Diterima	1,2,3,4,5,6,7,8,9,10,11,12, 13,14,15	15
		Direvisi	-	0
		Ditolak	-	0
3.	Gerak Harmonik pada Benda Elastik (Paket 3)	Diterima	1,2,3,4,5,6,7,8,9,10,11,12, 13,14,15,16,17,18,19,20,2 1,22, 23,24,25	25
		Direvisi	-	0
		Ditolak	-	0

Dari hasil analisis kuantitatif dapat diketahui bahwa semua soal yang diujikan pada uji kelompok besar ini memiliki kriteria daya beda diterima. Kriteria diterima menunjukkan soal tersebut sudah mampu membedakan kemampuan siswa kelompok atas dan siswa kelompok bawah.

d. Efektifitas Distraktor

Efektifitas distraktor merupakan kriteria seberapa baik pilihan jawaban salah yang tersedia dapat mengecoh peserta tes yang tidak mengetahui kunci jawaban yang tersedia. Semakin banyak suatu jawaban salah dipilih maka semakin baik pula pengecoh menjalankan tugasnya.

Pada uji kelompok besar ini peserta tes ada 56 orang siswa, jadi 5 % dari 56 siswa adalah 2.8, berarti dalam tes ini suatu distraktor dapat dikatakan efektif apabila dipilih minimal 3 orang, terutama dari kelompok bawah, akan tetapi distraktor kurang maksimal jika hanya dipilih oleh kelompok bawah saja, sebaiknya dipilih pula oleh kelompok atas. Distraktor sebaiknya dipilih lebih banyak oleh kelompok bawah, jika dipilih lebih banyak oleh kelompok atas maka distraktor tersebut menyesatkan dan sebaiknya diganti dengan distraktor lain. Distraktor dikatakan berfungsi apabila semua distraktor pada tiap soal berfungsi baik, apabila dalam satu soal ada distraktor yang belum berfungsi, atau menyesatkan maka distraktor pada soal tersebut dikatakan belum maksimal berfungsi

sebaiknya distraktor tersebut direvisi. Hasil rekapitulasi analisis distraktor dapat dilihat Tabel 9.

Tabel 9 Rekapitulasi Hasil Analisis Efektifitas Distraktor

No	Instrumen Tes	Kriteria Efektifitas Distraktor	Nomer Soal	Jumlah
1	Kinematika dengan Analisis Vektor (Paket 1)	Berfungsi	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25	24
		Belum maksimal	-	0
2.	Gravitasi (Paket 2)	Berfungsi	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15	15
		Belum maksimal	-	0
3.	Gerak Harmonik pada Benda Elastik (Paket 3)	Berfungsi	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25	25
		Belum maksimal	-	0

Dari hasil analisis efektifitas distraktor dapat diketahui bahwa semua distraktor sudah berfungsi dengan baik.

Setelah didapatkan hasil analisis reliabilitas, tingkat kesukaran, daya beda, dan efektifitas distraktor, kemudian dari masing-masing soal diputuskanlah soal yang sudah baik, dan soal yang perlu direvisi, atau soal mana yang ditolak sehingga perlu diganti dengan soal yang baru. Menurut Elvin & Surantoro (2010 :187) suatu soal dapat dimasukkan dalam kriteria soal diterima, direvisi, atau ditolak jika memenuhi kriteria keputusan untuk penilaian item soal sebagai berikut : (1) item soal diterima, apabila karakteristik item soal memenuhi semua kriteria. Item soal yang terlalu sukar atau terlalu mudah, tetapi memiliki daya beda dan distribusi pengecoh item yang memenuhi kriteria, butir soal tersebut dapat diterima atau dipilih; (2) item soal direvisi, apabila salah satu atau lebih dari ketiga kriteria karakteristik item soal tidak memenuhi kriteria ; (3) item soal ditolak, jika item soal memiliki karakteristik yang tidak memenuhi semua kriteria.

Dari 64 soal yang telah dibuat didapatkan bahwa semua soal dikatakan berkualitas baik atau diterima. Perinciannya dapat dilihat pada Tabel 10.

Tabel 10. Rincian Hasil Uji Kelompok Besar

Instrumen tes	Kriteria	Nomer Soal	Jumlah
Kinematika dengan Analisis Vektor (Paket 1)	Diterima	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25	24
	Direvisi	-	0
	Ditolak	-	0

Gravitasi (Paket 2)	Diterima	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15	15
	Direvisi	-	0
	Ditolak	-	0
Gerak Harmonik pada Benda Elastik (Paket 3)	Diterima	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25	25
	Direvisi	-	0
	Ditolak	-	0

Berdasarkan prinsip pelaksanaan penilaian otentik instrumen tes formatif yang disusun sudah mampu untuk mendukung pelaksanaan penilaian otentik karena dengan adanya tes formatif yang dilaksanakan sebanyak 3 kali dapat dilihat kemajuan perkembangan belajar siswa, apakah naik atau turun (prinsip *Keeping track*). Selain itu hasil dari tes formatif yang dilaksanakan mampu menunjukkan ketercapaian kemampuan peserta didik karena dari skor yang dihasilkan dapat dilihat apakah nilai siswa mencapai kriteria ketuntasan minimum (KKM) atau tidak (prinsip *Checking Up*). Tes formatif ini digunakan diakhir suatu pokok materi, dari hasil tes tersebut dapat dilihat rata-rata nilai siswa apabila masih banyak yang mendapat nilai di bawah KKM dapat sebagai bahan evaluasi terhadap pelaksanaan pembelajaran (prinsip *Finding Out*). Tes formatif yang disusun merupakan penjabaran dari suatu Standar Kompetensi, dari 3 kali tes tersebut dapat diambil kesimpulan sementara (melalui rata-rata skor tes formatif) apakah siswa tersebut sudah mencapai KKM atau belum (prinsip *Summing Up*).

Penilaian otentik memiliki beberapa karakteristik, yaitu : 1) penilaian merupakan bagian dari proses pembelajaran. 2) penilaian mencerminkan hasil proses belajar pada kehidupan nyata. 3) menggunakan bermacam-macam instrumen, pengukuran, dan metode yang sesuai dengan karakteristik dan esensi pengalaman belajar. 4) penilaian harus bersifat komprehensif dan holistik yang mencakup semua aspek dari tujuan pembelajaran. Dari karakteristik tersebut dapat diketahui bahwa instrumen tes yang disusun mampu mendukung pelaksanaan penilaian otentik meskipun tidak semua karakteristik tetapi hanya pada karakteristik nomor satu dan empat, karena pada penilaian otentik yang menilai proses pembelajaran memerlukan banyak metode yang tidak hanya *paper and pencil* (bentuk tes) tetapi juga hasil karya (*product*), penugasan (*project*), unjuk kerja (*performance*), dan kumpulan hasil kerja (*portfolio*).

KESIMPULAN DAN SARAN

Berdasarkan hasil analisis data secara kualitatif dan secara kuantitatif hasil penelitian mengenai pengembangan instrumen tes formatif kelas XI Program Akselerasi adalah :

1. Tes formatif pilihan ganda untuk kelas XI semester gasal Program Akselerasi disusun berdasarkan Standar Kompetensi “Mendeskripsikan gejala alam dalam cakupan mekanika klasik system diskret (partikel) sehingga mampu mendukung pelaksanaan *authentic assessment* jenis *paper and pencil* karena ketersinambungan materinya .
2. Tes formatif pilihan ganda untuk kelas XI semester gasal Program Akselerasi telah memenuhi kriteria kualitatif suatu instrumen yang baku karena telah ditelaah oleh ahli evaluasi, ahli bahasa, dan ahli materi dan memenuhi kriteria baik pada aspek konstruksi soal, materi dan bahasa.
3. Tes Fisika yang dikembangkan memenuhi kriteria baik secara kualitatif karena telah ditelaah oleh tim ahli. Secara kuantitatif memiliki reliabilitas soal yaitu 0,810701 untuk Paket 1 yang tergolong sangat tinggi, Paket 2 0,6844 yang tergolong tinggi , dan paket

3 0,824764 yang tergolong sangat tinggi. Daya beda soal diterima, yaitu $D > 0,3$, taraf kesukaran sedang yaitu $0,3 \leq P \leq 0,7$, dan pengecoh yang berfungsi dengan baik.

Berdasarkan hasil analisis secara kualitatif dan kuantitatif penelitian pengembangan tes formatif kelas XI pilihan ganda untuk Program Akselerasi ini memiliki saran pemanfaatan dan pengembangan produk lebih lanjut. Dalam pemanfaatannya produk soal yang dibuat disarankan untuk digunakan sebagai bank soal, sedangkan dalam pengembangan produk lebih lanjut dapat dilakukan pengembangan soal tes formatif untuk materi lain, dan jenjang kelas yang lain, yaitu kelas X, dan kelas XII, selain itu juga dapat dilakukan pengembangan instrumen penilaian otentik tidak hanya jenis *paper and pencil* tetapi juga jenis yang lain, untuk penelitian yang sejenis disarankan untuk mengujicobakan tes dengan subyek coba yang lebih banyak agar didapatkan hasil yang lebih baik lagi.

DAFTAR PUSTAKA

1. Anas, S. (2008). *Pengantar Evaluasi Pendidikan*. Jakarta: PT Raja Grafindo Persada.
2. Caldwell, J.S.(2008). *Comprehension Assesment*. New York :The Guildford Press.
3. Depdiknas.(2009). *Analisis Butir Soal*.Jakarta: Depdiknas
4. Depdiknas. (2009). *Pedoman Penyelenggaraan Pendidikan Untuk Peserta Didik Cerdas Istimewa*. Jakarta : Depdiknas
5. Djemari, M. (2009). *Pengembangan Instrumen Penelitian Pendidikan*. Yogyakarta: UNY Press
6. Echols, J., Hasan, S. (1983). *Kamus Inggris-Indonesia*. Jakarta-Indonesia. Jakarta : Erlangga.
7. Elvin,Y.,Surantoro.(2004). *Pemanfaatan Telaah Kualitatif dan Kuantitatif sebagai Prasyarat Penyelesaian Item Tes Pilihan Ganda dalam Penyelesaian Tugas Akhir di LPTK*. Jurnal Akademika. Vol II no 2.
8. Milne,J.H. (2008). *Technological Support for Asignment Assesment : A New Zealand Higher Education Survey*. Australian Journal of Educational Technology. 24(5). 487-504
9. Mohammad, K. (1984). *Teknik Analisa Item*. Surabaya: Usaha Nasional.
10. Nana, S.(2006). *Penilaian Hasil Proses Belajar Mengajar*. Bandung : PT. Remaja Rosdyakarya.
11. Nana, S.(2007). *Metode Penelitian Pendidikan*. Bandung : PT. Remaja Rosdyakarya.
12. Rusman, N.(2008). *Program Akselerasi Bagi Siswa Berbakat Akademik*.Jurnal pendidikan dan Kebudayaan.No 073.Tahun ke-14
13. Siregar, H. (2003). *Peranan Fisika Pada Disiplin Ilmu Teknik Kimia*. [online]. <http://library.usu.ac.id/download/ft/tkimia-harrys2.pdf>. Diakses tanggal 10 Desember 2009.
14. Slameto.(2001). *Evaluasi Pendidikan*.Jakarta : Bumi Aksara.
15. Subratha. (2007). *Meningkatkan Motivasi Belajar Siswa dalam Pembelajaran Fisika*. [online]. <http://www.pdf-search-engine.com/meningkatkan->
16. Sugiyono.(2010). *Metode Penelitian Kuantitatif, Kualitatif Dan R&D*.Bandung : Alfabeta
17. Suharsimi, A. (2006). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta : Bumi Aksara
18. Suharno, dkk. (2000). *Belajar Dan Pembelajaran II*. Surakarta: UNS Press
19. Sumadi,S. (1987). *Pengembangan Tes Hasil Belajar*. Jakarta : CV Rajawali
20. Sumarna,S.(2004). *Panduan Penulisan Tes Tertulis*. Bandung : Remaja: Erlangga