

# **Pengembangan Software AnBuso Sebagai Alat Analisis Butir Soal yang Praktis dan Aplikatif <sup>1</sup>**

Oleh: Ali Muhson, Barkah Lestari, Supriyanto & Kiromim Baroroh  
(Fakultas Ekonomi, Universitas Negeri Yogyakarta)  
Email: alimuchson@yahoo.com

## **Abstrak**

Penilaian merupakan salah satu faktor penting dalam pembelajaran. Guna meningkatkan kualitas penilaian pembelajaran, penelitian ini bertujuan untuk mengembangkan software AnBuso yang dapat dimanfaatkan guru dalam melakukan analisis butir soal secara praktis dan aplikatif; menguji kelayakan dari software AnBuso; dan mengidentifikasi kendala yang dihadapi dalam menggunakan software AnBuso. Penelitian dan pengembangan ini menggunakan dokumentasi, kuesioner dan wawancara untuk mengumpulkan data. Data yang terkumpul dianalisis secara deskriptif. Penelitian ini menghasilkan software AnBuso yang dapat dimanfaatkan guru dalam melakukan analisis butir soal secara praktis dan aplikatif. Software tersebut dinilai sangat layak oleh guru dilihat dari aspek kepraktisan dan kemudahan, aspek kebermanfaatan dan aspek substansi isi serta aspek tampilan. Kendala yang dihadapi guru dalam menggunakan software ini terkait dengan lemahnya penguasaan guru terhadap program Microsoft Excel, kurang terbiasanya melakukan analisis butir soal, pemahaman konsep analisis butir soal yang terbatas, dan kendala teknis yang terdapat dalam software.

Kata Kunci: AnBuso, kelayakan software, analisis butir soal

## **A. Pendahuluan**

Pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya dan masyarakat. Fungsi dan tujuan pendidikan nasional tertuang dalam Undang-Undang Nomor 20 Tahun 2003. Dalam UU tersebut dinyatakan bahwa pendidikan nasional berfungsi mengembangkan kemampuan dan membentuk watak serta peradaban bangsa yang bermartabat, bertujuan untuk berkembangnya potensi peserta didik agar menjadi manusia yang beriman dan bertaqwa kepada Tuhan Yang Maha Esa, berakhlak mulia, sehat, berilmu, sehat, kreatif, mandiri, dan menjadi warga negara yang demokratis serta bertanggung jawab.

Peningkatan kualitas pendidikan mutlak perlu terus dilakukan guna memenuhi tuntutan di atas. Banyak faktor yang mempengaruhi kualitas pendidikan di antaranya adalah guru, kondisi siswa, pengelola sekolah, lingkungan, dan kurikulum. Hasil penelitian Mardapi, dkk (2010) juga menemukan bahwa faktor determinan yang menentukan kualitas pembelajaran yakni faktor pendidik, kepala sekolah, dan manajemen. Hal ini mengindikasikan bahwa pendidik memiliki peran yang sangat sentral dalam meningkatkan kualitas pendidikan.

---

<sup>1</sup> Disampaikan dalam Seminar Nasional Evaluasi Pendidikan (SNEP) dengan tema "Pengembangan Pendidik: Implementasi Asesmen Otentik Pendidikan dalam Rangka Meningkatkan Kompetensi dan Kinerja Profesional Berkelanjutan" pada tanggal 22 November 2014 di Pascasarjana UNNES Semarang

Menurut Irawan sebagaimana dikutip Setiaji (2010) dalam kegiatan pembelajaran, perlu ditempuh strategi-strategi pokok yaitu 1) menciptakan iklim pembelajaran kondusif, 2) diagnosis kebutuhan belajar, 3) perencanaan, 4) formulasi tujuan, 5) mengembangkan model umum, 6) menetapkan materi dan teknik pembelajaran, dan 7) peranan evaluasi. Pendapat tersebut menunjukkan bahwa proses pembelajaran akan berjalan dengan baik jika pendidik juga kompetensi yang memadai baik dalam merancang, melaksanakan dan mengevaluasi hasil belajar. Hal senada juga disampaikan oleh Mardapi bahwa usaha peningkatan kualitas pendidikan dapat ditempuh melalui peningkatan kualitas pembelajaran dan kualitas sistem penilaian. Keduanya saling terkait, sistem pembelajaran yang baik akan menghasilkan kualitas belajar yang baik. Selanjutnya sistem penilaian yang baik akan mendorong guru untuk menentukan strategi mengajar yang baik dan memotivasi siswa untuk belajar yang lebih baik (Widoyoko, 2009). Dengan demikian salah satu faktor penting untuk efektivitas pembelajaran adalah faktor evaluasi baik terhadap proses maupun hasil pembelajaran.

Ebel & Frisbie (1991: 23) menyatakan bahwa *the purpose of evaluation is to make a judgment about the quality or worth of something-an educational program, worker performance or proficiency, or student attainment*. Hal ini menunjukkan bahwa dalam evaluasi terkandung sebuah upaya untuk membuat keputusan tentang kualitas pencapaian program serta kinerja dan kemampuan peserta didik. Untuk itu dalam melakukan evaluasi perlu memperhatikan keseluruhan aspek yang akan diukur agar mampu menggambarkan dengan tepat sasaran yang dituju.

Evaluasi dapat meliputi evaluasi hasil belajar peserta didik yang dilakukan oleh pendidik untuk memantau proses, kemajuan, dan perbaikan hasil belajar peserta didik secara berkesinambungan; dan evaluasi peserta didik, satuan pendidikan, dan program pendidikan yang dilakukan oleh lembaga mandiri secara berkala, menyeluruh, transparan, dan sistematis, untuk menilai pencapaian standar nasional pendidikan.

Evaluasi dapat mendorong siswa untuk lebih giat belajar secara terus menerus dan juga mendorong guru untuk lebih meningkatkan kualitas proses pembelajaran serta mendorong sekolah untuk lebih meningkatkan fasilitas dan kualitas manajemen sekolah. Sehubungan dengan hal tersebut, maka di dalam pembelajaran dibutuhkan guru yang tidak hanya mampu mengajar dengan baik tetapi juga mampu melakukan evaluasi dengan baik. Kegiatan evaluasi sebagai bagian dari program pembelajaran perlu lebih dioptimalkan. Evaluasi tidak hanya bertumpu pada penilaian hasil belajar, tetapi juga perlu penilaian terhadap input, output maupun kualitas proses pembelajaran itu sendiri.

Menurut Sudijono (2011) secara administratif, evaluasi dapat berfungsi untuk memberikan laporan, bahan keterangan (data) dan gambaran mengenai hasil yang telah dicapai. Dengan melakukan evaluasi, akan disusun dan disajikan laporan mengenai kemajuan dan perkembangan peserta didik setelah mereka mengikuti proses pembelajaran. Nilai-nilai hasil belajar peserta didik merupakan data yang sangat penting untuk pengambilan keputusan. Karena itu pendidik harus mampu membuat laporan administrasi hasil evaluasi secara baik dan mudah untuk digunakan dalam pengambilan keputusan.

Namun demikian laporan administrasi guru, terutama yang terkait dengan evaluasi pembelajaran sangatlah bervariasi. Hasil penelitian Niron, dkk (2005) menemukan bahwa ada perbedaan format dalam perangkat administrasi/manajemen pelaksanaan KBK pada dimensi perencanaan, pelaksanaan dan evaluasi yang digunakan guru SMA pada lima kabupaten/Kota di lingkungan Dinas Pendidikan Propinsi Daerah Istimewa Yogyakarta. Hal ini mengindikasikan bahwa kemampuan guru dalam membuat laporan hasil evaluasi belajar kurang memadai. Selama ini guru juga merasa kebingungan dalam membuat laporan tersebut, terutama dalam hal melakukan

analisis butir soal. Bahkan yang terjadi hampir sebagian besar guru tidak melakukan analisis butir soal secara cermat. Karena itu perlu ada upaya untuk membantu guru dalam melakukan analisis tes hasil belajar yang praktis dan aplikatif sehingga mudah untuk dilaksanakan dan diterapkan.

Selama ini memang sudah banyak software yang dikembangkan untuk keperluan analisis butir soal seperti ITEMAN, RASCAL, ASCAL, BILOG, FACETS dan CONQUEST, namun pada kenyataannya kemampuan guru untuk menguasai software tersebut sangat kurang memadai. Hal itu terjadi karena di samping software tersebut berbahasa asing sehingga sulit untuk memahami cara penggunaannya, software tersebut juga cukup rumit untuk digunakan dan kurang praktis dan aplikatif. Informasi yang diberikan dalam software tersebut juga sangat beragam sehingga mempersulit guru untuk menguasainya. Guna mengatasi kelemahan terutama dalam penguasaan bahasa, ada juga yang telah mengembangkan software yang berbahasa Indonesia, seperti SPS dan ANATES, namun software tersebut juga tidak banyak digunakan guru karena terkendala teknis dan kurang aplikatif.

Untuk memenuhi kebutuhan software yang aplikatif dalam melakukan analisis butir soal, perlu dikembangkan software baru yang dalam penggunaannya mudah untuk dipahami guru. Software tersebut adalah **AnBuso (Analisis Butir Soal)** yang dikembangkan dengan program aplikasi Microsoft Excel. Program aplikasi Excel tentu sangat familiar di kalangan guru karena sudah banyak dipakai di berbagai kalangan. Dengan memanfaatkan kepopuleran program aplikasi tersebut akan mempermudah penggunaannya. Dalam software AnBuso tersebut akan diketahui baik tidaknya soal yang dibuat guru, baik dari sisi daya beda, tingkat kesulitan, maupun efektivitas distraktornya. Di samping itu dalam software tersebut juga memberikan informasi tentang kemampuan seluruh siswa dan tingkat ketercapaian KKM. Software ini juga dirancang untuk mampu mengidentifikasi dan mengelompokkan siswa yang masuk dalam program remedial berdasarkan materi yang belum dikuasai sehingga akan mempermudah guru dalam pelaksanaan program remedial.

Hakikat mengajar adalah memfasilitasi siswa agar mendapat kemudahan dalam belajar. Menurut Joyce, Weil & Showers (Depdiknas, 2004:7) mengajar (teaching) adalah membantu siswa memperoleh informasi, ide, keterampilan, nilai, cara berpikir, sarana untuk mengekspresikan diri dan belajar bagaimana belajar. Hasil jangka panjang dari proses mengajar adalah kemampuan siswa untuk dapat belajar dengan mudah dan efektif.

Kemajuan teknologi informasi dan kemajuan bisnis dewasa ini, di masyarakat telah menimbulkan tuntutan baru. Tuntutan yang didasarkan kepada kebutuhan untuk mengetahui atau menentukan potensi seorang manusia memerlukan adanya assesmen atau evaluasi kompetensi. Di dunia usaha, perusahaan menginginkan potensi para buruh yang direkrutnya. Badan usaha milik swasta maupun milik Negara ingin mengetahui potensi para pegawainya. Di dunia pendidikan, perguruan tinggi ingin mengetahui para lulusan di jenjang di bawahnya (SMA/MA/SMK) yang melanjutkan di universitas yang diinginkannya. Para penyelenggara sekolah, menginginkan agar para lulusan jenjang sebelumnya agar memilih melanjutkan studi di sekolahnya; Ini semua bisa dilakukan dengan baik jika mereka melakukan atau memperoleh data dengan melakukan asesmen.

Media asesmen ini bisa bervariasi namanya termasuk di antaranya tes, ulangan, ujian, dan evaluasi. Jika tindakan untuk mengetahui potensi para siswa dilakukan secara nasional sama, atau dengan standar yang sama, maka kegiatan tersebut bisa disebut ujian nasional; sebaliknya jika kegiatan untuk mengetahui potensi seseorang tersebut dilakukan oleh masing-masing daerah misal propinsi/kabupaten, maka disebut kegiatan asesmen daerah. Demikian juga jika kegiatan asesmen dilakukan oleh unit sekolah, maka bisa disebut tes/ ujian /ulangan sekolah.

Evaluasi mempunyai arti yang berbeda untuk guru yang berbeda. Berikut merupakan beberapa arti yang telah secara luas dapat diterima oleh para guru di lapangan. Gronlund (1985) berpendapat evaluasi adalah suatu proses yang sistematis untuk menentukan atau membuat keputusan, sampai sejauh mana tujuan program telah tercapai. Pendapat yang sama juga dikemukakan oleh Wrightstone, dkk (1956) yang mengemukakan bahwa evaluasi pendidikan adalah penaksiran terhadap pertumbuhan dan kemajuan siswa ke arah tujuan atau nilai-nilai yang telah ditetapkan dalam kurikulum (Djaali & Pudji Muljono, 2007). Sedangkan Purwanti (2008: 6) berpendapat bahwa evaluasi adalah proses pemberian makna atau penetapan kualitas hasil pengukuran dengan cara membandingkan angka hasil pengukuran tersebut dengan kriteria tertentu.

Pengertian tersebut menunjukkan bahwa dalam evaluasi harus sudah bisa menilai kemajuan siswa dalam mencapai tujuan pembelajaran. Berdasarkan informasi tersebut dapat digunakan sebagai dasar untuk membuat suatu keputusan tentang status siswa dalam kelompoknya, bahkan hasil evaluasi tersebut juga dapat digunakan untuk menentukan langkah-langkah selanjutnya.

Menurut Sukardi (2011) beberapa fungsi penting evaluasi di antaranya adalah seperti berikut: 1) sebagai alat guna mengetahui apakah peserta didik telah menguasai pengetahuan, nilai-nilai dan keterampilan yang telah diberikan oleh seorang guru. 2) untuk mengetahui aspek-aspek kelemahan peserta didik dalam melakukan kegiatan belajar. 3) mengetahui tingkat ketercapaian siswa dalam kegiatan belajar, 4) sebagai sarana umpan balik bagi seorang guru, yang bersumber dari siswa. 5) sebagai alat untuk mengetahui perkembangan belajar siswa, 6) sebagai materi utama laporan hasil belajar kepada para orang tua siswa.

Oleh karena demikian bervariasinya fungsi evaluasi, maka sangat penting bagi para guru, agar ketika mereka merencanakan kegiatan evaluasi, sebaiknya perlu mempertimbangkan lebih dahulu fungsi dan karakteristik yang manakah evaluasi yang hendak dibuat untuk para siswa.

Dilihat dari fungsinya, jenis evaluasi atau penilaian terbagi ke dalam beberapa macam. Menurut Sudjana (2011) membagi penilaian menjadi penilaian formatif, penilaian sumatif, penilaian diagnostik, penilaian selektif, dan penilaian penempatan. Penilaian formatif adalah penilaian yang dilaksanakan pada akhir program belajar mengajar untuk melihat tingkat keberhasilan proses belajar mengajar itu sendiri. Penilaian formatif berorientasi pada proses, yang akan memberikan informasi kepada guru apakah program atau proses belajar mengajar masih perlu diperbaiki.

Penilaian sumatif adalah penilaian yang dilaksanakan pada akhir unit program misalnya penilaian yang dilaksanakan pada akhir catur wulan, akhir semester atau akhir tahun. Tujuan penilaian ini adalah untuk mengetahui hasil yang dicapai oleh para siswa, yakni seberapa jauh siswa telah mencapai kompetensi yang ditetapkan dalam kurikulum. Penilaian ini berorientasi pada produk/hasil.

Penilaian diagnostik adalah penilaian yang bertujuan untuk mengetahui kelemahan-kelemahan siswa serta faktor-faktor penyebabnya. Pelaksanaan penilaian semacam ini biasanya bertujuan untuk keperluan bimbingan belajar, pengajaran remedial, menemukan kasus-kasus dan lain-lain.

Penilaian selektif adalah penilaian yang dilaksanakan dalam rangka menyeleksi atau menyaring. Memilih siswa untuk mewakili sekolah dalam lomba-lomba tertentu termasuk jenis penilaian selektif. Untuk kepentingan yang lebih luas penilaian selektif misalnya seleksi penerimaan mahasiswa baru atau seleksi yang dilakukan dalam rekrutmen tenaga kerja.

Adapun penilaian penempatan adalah penilaian yang bertujuan untuk mengetahui keterampilan prasyarat yang diperlukan bagi suatu program belajar dan penguasaan belajar seperti yang diprogramkan sebelum memulai kegiatan belajar untuk program itu. Dengan kata lain penilaian ini berorientasi pada kesiapan siswa untuk menghadapi program baru dan kecocokan program belajar dengan kemampuan yang telah dimiliki siswa.

Evaluasi merupakan bagian dari proses belajar-mengajar yang secara keseluruhan tidak dapat dipisahkan dari kegiatan mengajar. Pada sebagian guru masih ada asumsi yang kurang tepat. Asumsi yang tidak pada tempatnya itu termasuk misalnya, adalah merupakan hal biasa dan kegiatan evaluasi tidak mempunyai tujuan tertentu, kecuali bahwa evaluasi adalah kegiatan yang diharuskan oleh peraturan atau undang-undang. Aturan yang mengikat tersebut termasuk pasal UU RI tahun 2003 tentang Sisdiknas 58, yang menyatakan “evaluasi hasil belajar peserta didik dilakukan oleh pendidik untuk memantau proses, kemajuan dan perbaikan hasil belajar peserta didik secara berkesinambungan”. Untuk mencapai tujuan tersebut, uraian berikut mendiskusikan cara evaluasi yang dilakukan guru untuk menghasilkan kegiatan belajar-mengajar yang lebih baik. Ada empat pertimbangan yang perlu diperhatikan oleh seorang guru dalam melakukan evaluasi belajar. Keempat macam pertimbangan tersebut ialah,

- a. Mengidentifikasi tujuan yang dapat dijabarkan dari (1) prosedur evaluasi dan hubungannya dengan mengajar, (2) pengembangan interes kebutuhan individu, (3) kebutuhan individu siswa, (4) kebutuhan yang dikembangkan dari komunitas/masyarakat, (5) dikembangkan evaluasi hasil belajar pendahulunya, (6) dikembangkan dari analisis pekerjaan dan (7) pertimbangan dari para ahli evaluasi.
- b. Menentukan pengalaman belajar yang biasanya direalisasi dengan *pretest* sebagai awal, pertengahan dan akhir belajar (*post test*)
- c. Menentukan standar yang bisa dicapai dan “menantang” siswa belajar lebih giat. Pembuatan standar yang dapat diajarkan melalui penilaian materi, penggunaan alat bantu visual. Di samping itu standar juga dapat dibuat melalui pengembangan dan pemakaian alat observasi yang sering dilakukan oleh seorang guru untuk memenuhi kepentingan mereka.
- d. Mengembangkan ketrampilan dan mengambil keputusan guna a) memilih tujuan, b) menganalisis pertanyaan problem solving, c) menentukan nilai seorang siswa.

Dalam setiap proses belajar mengajar ada beberapa komponen penting yaitu, pembukaan, inti belajar mengajar, penutup dan evaluasi. Sebagai komponen belajar-mengajar, evaluasi merupakan akhir dari setiap proses yang selalu ada dalam rencana proses belajar mengajar mempunyai tujuan yaitu untuk menentukan apakah materi yang disampaikan oleh para guru bisa diterima dengan baik oleh para peserta didiknya. Kegiatan evaluasi juga penting bagi para guru, di samping untuk mengukur tingkat penguasaan materi siswa, juga dapat digunakan untuk *feedback* perbaikan dalam menggunakan metode mengajar, pemilihan strategi penyampaian dan penentuan pencapaian hasil belajar. Evaluasi dapat diartikan sebagai proses pengambilan informasi tentang proses belajar-mengajar guna menentukan tingkat pencapaian hasil belajar seorang siswa. Untuk dapat melakukan evaluasi guru perlu melakukan pengukuran (Sukardi, 2011).

Untuk mendapatkan instrumen berkualitas tinggi, selain dilakukan analisis secara teori (telaah butir berdasarkan aspek isi, konstruksi, dan bahasa) perlu juga dilakukan analisis butir secara empirik. Secara garis besar, analisis butir secara empirik ini dapat dibedakan menjadi dua, yaitu dengan pendekatan teori tes klasik (*Classical Test Theory*) dan teori respons butir (*Item Response Theory, IRT*). Teori tes klasik atau disebut teori skor murni klasik (Allen & Yen,

1979:57) didasarkan pada suatu model aditif, yakni skor amatan merupakan penjumlahan dari skor sebenarnya dan skor kesalahan pengukuran.

Validitas dan reliabilitas pada perangkat tes digunakan untuk menentukan kualitas tes. Kriteria lain yang dapat digunakan untuk menentukan kualitas tes adalah indeks kesukaran, daya pembeda dan efektivitas distraktor. Validitas suatu perangkat tes dapat diartikan merupakan kemampuan suatu tes untuk mengukur apa yang seharusnya diukur (Gorin, 2007; Azwar, 2012: 51). Menurut Cohen-Swrdlik (2009) *validity, as applied to a test, is a judgment or estimate of how well a test measures what it purports to measure in a particular context*. Pengertian ini menunjukkan bahwa validitas dalam sebuah tes menjadi hal yang sangat penting karena akan mengukur kemampuan peserta didik secara tepat.

Validitas juga merujuk pada ketepatan (*appropriateness*), kebermaknaan (*meaningfulness*) dan kemanfaatan (*usefulness*) kesimpulan yang didapatkan dari interpretasi skor tes (Kusaeri dan Suprananto, 2012). Ada tiga tipe validitas, yaitu validitas isi, validitas konstruk dan validitas kriteria (Moss, et.al., 2006; Cohen-Swrdlik 2009: 185).

Ada dua macam validitas isi, yaitu validitas kenampakan dan validitas logika (Azwar, 2012). Validitas isi berarti sejauh mana suatu perangkat tes mencerminkan keseluruhan kemampuan yang hendak diukur (Azwar, 2012), yang berupa analisis rasional terhadap domain yang hendak diukur. Validitas kenampakan didasarkan pada pertanyaan apakah suatu butir-butir dalam perangkat tes mengukur aspek yang relevan dengan domainnya. Validitas logika berkaitan dengan keseksamaan batasan pada domain yang hendak diukur, dan merupakan jawaban apakah keseluruhan butir merupakan sampel representatif dari keseluruhan butir yang mungkin dibuat.

Tingkat kesukaran suatu butir soal, yang disimbolkan dengan  $p_i$ , merupakan salah satu parameter butir soal yang sangat berguna dalam penganalisisan suatu tes. Hal ini disebabkan karena dengan melihat parameter butir ini, akan diketahui seberapa baiknya kualitas suatu butir soal. Jika  $p_i$  mendekati 0, maka soal tersebut terlalu sukar, sedangkan jika  $p_i$  mendekati 1, maka soal tersebut terlalu mudah, sehingga perlu dibuang. Hal ini disebabkan karena butir tersebut tidak dapat membedakan kemampuan seorang siswa dengan siswa lainnya.

Sudjana (2011: 137) menyatakan bahwa secara umum indeks kesukaran suatu butir sebaiknya terletak dalam kategori sedang yakni pada interval 0,31 – 0,70. Pada interval ini, informasi tentang kemampuan siswa akan diperoleh secara maksimal. Dalam merancang indeks kesukaran suatu perangkat tes, perlu dipertimbangkan tujuan penyusunan perangkat tes tersebut.

Untuk menentukan daya pembeda, dapat digunakan indeks diskriminasi, indeks korelasi biserial, indeks korelasi *point biserial*, dan indeks keselarasan. Pada analisis butir dalam penelitian ini, hanya digunakan indeks korelasi *point biserial*.

Pada suatu butir soal, indeks daya beda dikatakan baik jika lebih besar atau sama dengan 0,3 (Nunnally & Bernstein, 2009: 304; Ebel & Frisbie, 1991; 232). Indeks daya pembeda suatu butir yang kecil nilainya akan menyebabkan butir tersebut tidak dapat membedakan siswa yang kemampuannya tinggi dan siswa yang kemampuannya rendah. Pada analisis tes dengan *Content-Referenced Measures*, indeks daya pembeda butir tidak terlalu perlu menjadi perhatian, asalkan tidak negatif (Ebel & Frisbie, 1991). Jika nilainya kecil, menunjukkan bahwa kemencengan distribusi skor dari populasi, yang juga mengakibatkan validitas tes menjadi rendah.

Menurut Sudijono (2011) pada saat membicarakan tes objektif bentuk *multiple choice item* tersebut untuk setiap butir *item* yang dikeluarkan dalam tes hasil belajar telah dilengkapi dengan beberapa kemungkinan jawab, atau yang sering dikenal dengan istilah *option* atau alternatif.

*Option* atau alternatif itu jumlahnya berkisar antara 3 sampai dengan 5 buah, dan dari kemungkinan-kemungkinan jawaban yang terpasang pada setiap butir *item* itu, salah satu di

antaranya adalah merupakan jawaban betul (kunci jawaban), sedangkan sisanya adalah merupakan jawaban salah. Jawaban-jawaban salah itulah yang biasa dikenal dengan istilah *distractor* (pengecoh).

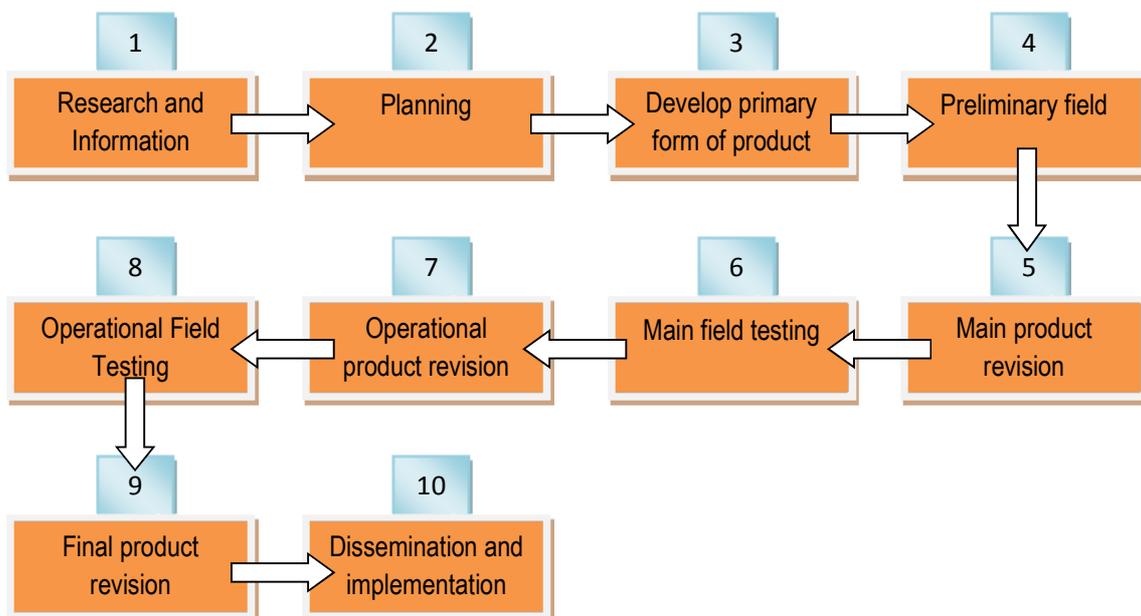
Menganalisis fungsi distraktor sering dikenal dengan istilah lain, yaitu : menganalisis pola penyebaran jawaban *item*. Adapun yang dimaksud dengan pola penyebaran jawaban *item* adalah suatu pola yang dapat menggambarkan bagaimana *testee* menentukan pilihan jawabannya terhadap kemungkinan-kemungkinan jawab yang telah dipasangkan pada setiap butir *item*.

Suatu kemungkinan dapat terjadi, yaitu bahwa dari keseluruhan alternatif yang dipasang pada butir *item* tertentu, sama sekali tidak dipilih oleh *testee*. Dengan kata lain, *testee* menyatakan “blangko”. Pernyataan blangko ini sering dikenal dengan istilah omiet dan biasa diberi lambang dengan huruf O. Distraktor dinyatakan telah dapat menjalankan fungsinya dengan baik apabila distraktor tersebut sekurang-kurangnya sudah dipilih oleh 5 % dari seluruh peserta tes.

Sebagai tindak lanjut atas hasil penganalisaan terhadap fungsi distraktor tersebut maka distraktor yang sudah dapat menjalankan fungsinya dengan baik dapat dipakai lagi pada tes-tes yang akan datang, sedangkan distraktor yang belum dapat berfungsi dengan baik sebaiknya diperbaiki atau diganti dengan distraktor yang lain.

## B. Metode Penelitian

Dalam model pengembangan, Gall, et.al. memuat panduan sistematika langkah-langkah yang dilakukan oleh peneliti agar produk yang dirancangnya mempunyai standar kelayakan. Dengan demikian, yang diperlukan dalam pengembangan ini adalah rujukan tentang prosedur produk yang akan dikembangkan. Menurut Gall, et.al. (2003:569) *Educational research and development (R & D) is the use of research findings to design new products and procedures, followed by the application of research methods to field-test, evaluate, and refine the products and procedures until they meet specified criteria of effectiveness, quality, or similar standards.*



Gambar 1: Skema Prosedur Pengembangan Hasil Adaptasi dari Prosedur Pengembangan

Dalam teknologi pembelajaran, deskripsi tentang prosedur dan langkah-langkah penelitian pengembangan sudah banyak dikembangkan. Gall, et.al. (2003) menyatakan bahwa prosedur penelitian pengembangan pada dasarnya terdiri dari dua tujuan utama, yaitu: (1) mengembangkan produk, dan (2) menguji keefektifan produk dalam mencapai tujuan.

Tujuan pertama disebut sebagai fungsi pengemban sedangkan tujuan kedua disebut sebagai validasi. Dengan demikian, konsep penelitian pengembangan lebih tepat diartikan sebagai upaya pengembangan yang sekaligus disertai dengan upaya validasinya. Pendekatan penelitian dan pengembangan mencakup 10 langkah umum seperti model yang terlihat di Gambar 1.

Penelitian ini melibatkan guru-guru di provinsi DIY, pengawas sekolah, pejabat dinas, dan para pakar dari perguruan tinggi. Pada tahap ujicoba terbatas dilakukan hanya kepada 6-12 guru sedangkan pada tahap ujicoba luas dilakukan kepada 50 guru yang berasal dari lima kabupaten/kota di provinsi DIY. Penentuan responden dilakukan secara *purposive sampling* dengan mempertimbangkan kemampuan guru dalam penguasaan komputer, khususnya program aplikasi Microsoft Excel.

Teknik pengumpulan data yang digunakan dalam penelitian ini meliputi:

1. Observasi

Observasi digunakan untuk memperoleh data atau informasi tentang kemampuan guru dalam menggunakan program aplikasi yang telah dikembangkan. Hal ini diperlukan untuk diperoleh data tentang kemampuan guru dalam penggunaan software yang telah dikembangkan.

2. Dokumentasi

Teknik dokumentasi dalam penelitian ini digunakan untuk memperoleh data tentang dokumen tes yang digunakan guru dalam mengukur kemampuan peserta didik, baik tes formatif maupun tes sumatif. Dokumen tersebut dapat berupa soal-soal ujian dan ulangan harian, program remedial, dan hasil analisis butir soal yang selama ini digunakan guru.

3. Angket

Angket digunakan untuk mengungkap masukan-masukan yang diperlukan dari guru, pengawas, pejabat dinas pendidikan dan para pakar. Angket ini juga sekaligus digunakan untuk mengukur tingkat kelayakan dari software yang telah dikembangkan. Angket yang dikembangkan meliputi angket kelayakan software baik yang terkait dari sisi tampilan, substansi materi/isi, dan aspek kepraktisan dan kemudahan.

4. Wawancara

Wawancara dilakukan kepada guru, pengawas, pejabat Dinas Pendidikan dan para pakar. Teknik digunakan untuk mengungkap berbagai kelebihan dan kelemahan dari software yang dikembangkan agar dapat dijadikan sebagai masukan untuk penyempurnaan.

Analisis data yang digunakan meliputi analisis deskriptif, baik secara kualitatif maupun kuantitatif. Analisis kuantitatif dilakukan terhadap data kuantitatif yang diperoleh dari angket. Analisis yang dilakukan meliputi analisis kelayakan software baik dilihat dari sisi tampilan, substansi materi/isi, maupun kepraktisan dan kemudahan. Dalam melakukan analisis ini digunakan tiga kriteria, yakni tinggi, cukup dan rendah (lihat Tabel 2).

Tabel 2. Kategorisasi Penilaian Kelayakan Software

No	Skor	Kategori
1	Lebih dari $M + 1,8 SD$	Sangat layak
2	$M + 0,6 SD$ s.d. $M + 1,8 SD$	Layak
3	$M - 0,6 SD$ s.d. $M + 0,6 SD$	Cukup
4	$M - 1,8 SD$ s.d. $M - 0,6 SD$	Tidak layak
5	Kurang dari $M - 1,8 SD$	Sangat tidak layak

Adapun data yang diperoleh dari dokumen dan wawancara dianalisis secara kualitatif dengan mengikuti prosedur interaktif yang dikembangkan oleh Miles dan Huberman (1994: 10-12). Aktivitas dalam analisis data kualitatif dilakukan secara interaktif dan berlangsung secara terus menerus sampai tuntas, sehingga datanya jenuh. Ukuran kejenuhan data ditandai dengan tidak diperolehnya lagi data atau informasi baru. Aktivitas dalam analisis meliputi reduksi data (*data reduction*), penyajian data (*data display*) serta Penarikan kesimpulan dan verifikasi (*conclusion drawing/verification*).

### C. Hasil dan Pembahasan

Pada penelitian tahun pertama sudah dihasilkan sebuah draft software AnBuso yang sudah tervalidasi dan sudah diujicoba secara terbatas. Pada penelitian ini hasil software tersebut kemudian dilakukan ujicoba secara luas dengan melibatkan user lebih banyak dan lebih variatif. Hasil pengembangan software yang dihasilkan pada tahun kedua ini melakukan beberapa perubahan sesuai masukan peserta. Beberapa perubahan yang penting yang dihasilkan adalah penyesuaian software ini dengan diberlakukannya kurikulum 2013 terutama yang terkait dengan masalah penilaian. Karena itu pada penelitian ini dilakukan revisi perbaikan yang meliputi perubahan tampilan, sheet **Input01**, sheet **Laporan Peserta**, sheet **Peserta Remedial**, dan perubahan formula.

#### 1. Perubahan tampilan

Perubahan tampilan perlu dilakukan karena dianggap terlalu banyak variasi warna sehingga terlihat kurang menarik. Oleh karena itu dilakukan perubahan-perubahan sesuai masukan. Perubahan tampilan tidak hanya dilakukan pada sheet input tetapi juga dilakukan perubahan tampilan pada sheet laporan. Berikut ini contoh perubahannya:



Gambar 1. Tampilan Sheet Input01 Sebelum dan Sesudah Perubahan

HASIL ANALISIS SOAL PILIHAN GANDA						
Satuan Pendidikan : SMA Negeri 80 Yogyakarta Nama Tes : Sumatif Mata Pelajaran : Ekonomi Kelas/Program : XI/IPS Tanggal Tes : 12 Agustus 2012 SK/KD :						
No Butir	Daya Beda		Tingkat Kesukaran		Alternatif Jawaban Tidak Efektif	Keterangan
	Koefisien	Keterangan	Koefisien	Keterangan		
1	0.088	Tidak Baik	0.700	Sedang	B	Tidak Baik
2	0.000	Tidak Baik	1.000	Mudah	ABC	Tidak Baik
3	0.204	Cukup Baik	0.700	Sedang	B	Revisi Pengecoh
4	0.437	Baik	0.600	Sedang	-	Baik
5	0.244	Cukup Baik	0.850	Mudah	C	Cukup Baik
6	-0.029	Tidak Baik	0.700	Sedang	AB	Tidak Baik
7	0.417	Baik	0.550	Sedang	D	Revisi Pengecoh
8	-0.219	Tidak Baik	0.600	Sedang	B	Tidak Baik
9	0.067	Tidak Baik	0.450	Sedang	A	Tidak Baik
10	0.312	Baik	0.900	Mudah	BC	Cukup Baik
11	0.321	Baik	0.700	Sedang	-	Baik

Sebelum Perubahan

HASIL ANALISIS SOAL PILIHAN GANDA						
Satuan Pendidikan : SMA Negeri 80 Yogyakarta Nama Tes : Sumatif Mata Pelajaran : Ekonomi Kelas/Program : XI/IPS Tanggal Tes : 30 Agustus 2014 Pokok Bahasan/Sub :						
No Butir	Daya Beda		Tingkat Kesukaran		Alternatif Jawaban Tidak Efektif	Keterangan
	Koefisien	Keterangan	Koefisien	Keterangan		
1	0.410	Baik	0.800	Mudah	B	Revisi Pengecoh
2	0.000	Tidak Baik	1.000	Mudah	ABC	Tidak Baik
3	0.261	Cukup Baik	0.750	Mudah	-	Cukup Baik
4	0.601	Baik	0.600	Sedang	-	Baik
5	0.033	Tidak Baik	0.800	Mudah	-	Tidak Baik
6	0.303	Baik	0.650	Sedang	AB	Revisi Pengecoh
7	0.569	Baik	0.750	Mudah	D	Revisi Pengecoh
8	0.183	Tidak Baik	0.750	Mudah	AB	Tidak Baik
9	0.464	Baik	0.600	Sedang	-	Baik
10	0.303	Baik	0.950	Mudah	BC	Revisi Pengecoh
11	0.227	Cukup Baik	0.850	Mudah	-	Cukup Baik

Sesudah Perubahan

Gambar 2. Tampilan Sheet Laporan Butir Sebelum dan Sesudah Perubahan

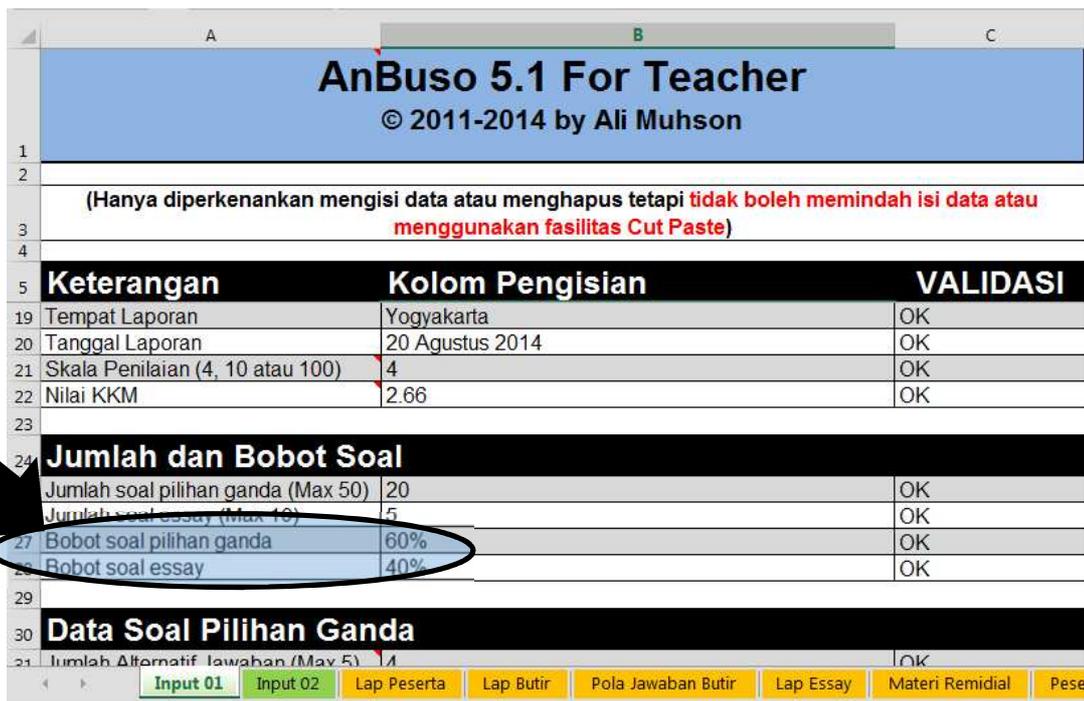
## 2. Perubahan Sheet Input01

Akibat diberlakukannya kurikulum 2013, software AnBuso juga dilakukan penyesuaian agar software ini mampu mengakomodasikan kepentingan guru dalam membuat penilaian sesuai dengan kurikulum 2013. Beberapa perubahan yang dilakukan mencakup dalam hal penentuan skala penilaian. Pada software sebelumnya skala penilaian yang disediakan hanya 1-10 dan 1-100, sementara itu kurikulum 2013 menggunakan skala penilaian 1-4, karena itu dalam software ini dilakukan penyesuaian dengan menyediakan skala penilaian 1-4. Perubahannya dapat dilihat pada Gambar 3.

AnBuso 5.1 For Teacher		
© 2011-2014 by Ali Muhson		
(Hanya diperkenankan mengisi data atau menghapus tetapi tidak boleh memindah isi data atau menggunakan fasilitas Cut Paste)		
Keterangan	Kolom Pengisian	VALIDASI
Tempat Laporan	Yogyakarta	OK
Tanggal Laporan	20 Agustus 2014	OK
Skala Penilaian (4, 10 atau 100)	4	OK
Nilai KKM	2.66	OK
Jumlah dan Bobot Soal		
Jumlah soal pilihan ganda (Max 50)	20	OK
Jumlah soal essay (Max 10)	5	OK
Bobot soal pilihan ganda	60%	OK
Bobot soal essay	40%	OK
Data Soal Pilihan Ganda		
Jumlah Alternatif Jawaban (Max 5)	4	OK

Gambar 3. Perubahan Skala Penilaian

Berdasarkan hasil ujicoba luas juga ditemukan beberapa kendala dalam menentukan bobot penilaian antara soal objektif dan soal essay. Karena itu software ini juga dilakukan perubahan dalam penentuan bobot tersebut dengan menyediakan kolom tersendiri untuk bobot soal objektif dan soal essay (perubahannya dapat dilihat pada Gambar 4). Dengan cara tersebut diharapkan guru atau user akan semakin praktis dalam menentukan bobot penilaiannya. Bahkan software ini juga memungkinkan untuk digunakan hanya untuk soal objektif saja atau untuk soal essay saja.



AnBuso 5.1 For Teacher		
© 2011-2014 by Ali Muhson		
(Hanya diperkenankan mengisi data atau menghapus tetapi tidak boleh memindah isi data atau menggunakan fasilitas Cut Paste)		
Keterangan	Kolom Pengisian	VALIDASI
Tempat Laporan	Yogyakarta	OK
Tanggal Laporan	20 Agustus 2014	OK
Skala Penilaian (4, 10 atau 100)	4	OK
Nilai KKM	2.66	OK
Jumlah dan Bobot Soal		
Jumlah soal pilihan ganda (Max 50)	20	OK
Jumlah soal essay (Max 10)	5	OK
<b>Bobot soal pilihan ganda</b>	<b>60%</b>	OK
<b>Bobot soal essay</b>	<b>40%</b>	OK
Data Soal Pilihan Ganda		
Jumlah Alternatif Jawaban (Max 5)	4	OK

Gambar 4. Penambahan Bobot Penilaian

### 3. Perubahan Laporan Peserta

Penyesuaian dengan kurikulum 2013 juga berdampak pada perubahan pada Sheet **Laporan Peserta**. Pada bagian ini dimunculkan hasil penilaian peserta tes menurut kurikulum 2013 lengkap dengan predikatnya (hasil perbaikannya dapat dilihat pada Gambar 5). Predikat penilaian dilakukan penyesuaian berdasarkan Permendikbud Nomor 52. Sementara itu pada bagian yang lain tidak mengalami perubahan karena sudah sesuai dengan yang diharapkan guru.

DAFTAR NILAI SISWA									
Satuan Pendidikan		: SMA Negeri 80 Yogyakarta							
Nama Tes		: Sumatif							
Mata Pelajaran		: Ekonomi							
Kelas/Program		: XI/IPS							
Tanggal Tes		: 30 Agustus 2014							
Pokok Bahasan/Sub		:							
								KKM	
								2.66	
No	NAMA PESERTA	L/P	HASIL TES OBJEKTIF			NILAI TES ESSAY	NILAI AKHIR	PREDIKAT	KET
			BENAR	SALAH	NILAI				
1	RIANDITA PRATIWI	P	19	1	3.80	3.51	3.68	A	Tuntas
2	N VIAN BURHANUDIN L	L	13	7	2.60	2.77	2.67	B	Tuntas
3	DWI CAHYO ADHI	L	12	8	2.40	3.1	2.70	B	Tuntas
4	ISHMATUL FATHIYAH	P	12	8	2.40	3.3	2.77	B	Tuntas
5	BAGUS PRATOMO ARIF AL HAKIM	L	11	9	2.20	2.8	2.48	B-	Belum tuntas
6	R BRATA KUSUMASMARA	L	15	5	3.00	3.1	3.06	B+	Tuntas
7	LILIANA ROMANTIKA SARI	P	12	8	2.40	3.0	2.65	B-	Belum tuntas
8	RIYAN ABI SAYEKTI	P	18	2	3.60	2.52	3.17	B+	Tuntas
9	DWI TITIK PUSPITOSARI	P	12	8	2.40	2.65	2.50	B-	Belum tuntas
10	PIPIT HARYADI	L	11	9	2.20	2.58	2.35	B-	Belum tuntas
11	AWIG JUWANTO	L	17	3	3.40	2.65	3.10	B+	Tuntas
12	ELIN DAHRIAH	P	15	5	3.00	2.71	2.6	B	Tuntas

Gambar 5. Perbaikan Sheet Laporan Peserta

#### 4. Perubahan Sheet Peserta Remedial

Pada dasarnya tujuan guru melakukan analisis butir soal di samping untuk mengetahui kualitas butir soal yang telah dibuat juga informasi hasil penilaian pesertanya dapat dimanfaatkan untuk melakukan rencana tindak lanjut baik untuk keperluan remedial maupun pengayaan. Oleh karena itu hasil laporan peserta haruslah mampu memberikan gambaran siapa saja peserta yang masuk pada kelompok pengayaan dan remedial. Pada pengembangan software ini sudah mampu mengantisipasi hal tersebut namun demikian pada versi sebelumnya hanya sebatas pengelompokan peserta remedial saja dan belum disediakan kolom untuk melakukan tindak lanjut.

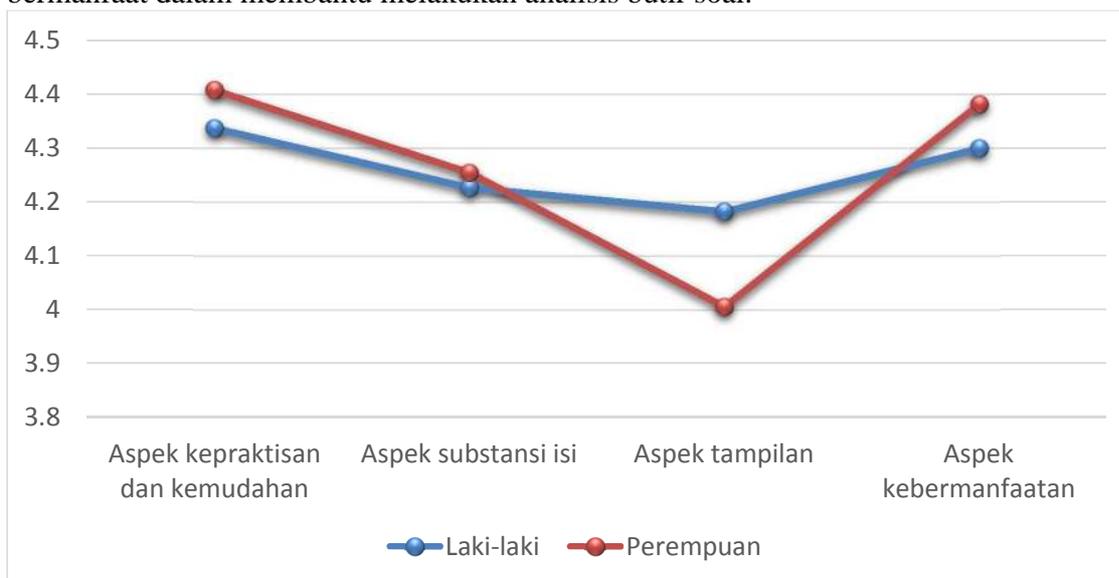
Guna memenuhi hal tersebut pada sheet **Peserta Remedial** dilakukan perbaikan yakni tidak hanya menemukan kelompok peserta remedial menurut kemampuan yang diukur melainkan disediakan kolom untuk pengisian jadwal kegiatan remedialnya. Hal itu diperlukan agar mampu mengakomodasikan kebutuhan guru dalam membuat jadwal remedial yang lebih praktis. Perbaikan pada sheet **Laporan Peserta Remedial** tersebut dapat dilihat pada Gambar 6.

PENGELOMPOKAN PESERTA REMIDIAL							
1							
2							
3	<b>Satuan Pendidikan</b>	: SMA Negeri 80 Yogyakarta					
4	<b>Nama Tes</b>	: Sumatif					
5	<b>Mata Pelajaran</b>	: Ekonomi					
6	<b>Kelas/Program</b>	: XI/IPS					
7	<b>Tanggal Tes</b>	: 30 Agustus 2014					
8	<b>Pokok Bahasan/Sub</b>	:					
9	No	Kompetensi Dasar	Peserta Remedial	Hari	Tgl	Jam	Tempat
10	<b>Soal Objektif</b>						
11	1	Masalah Pokok Ekonomi	DWI CAHYO ADHI ; LILIANA ROMANTIKA SARI ; ELIN DAHRIAH; FITRIYANTI SYAH;				
12	2	Kelangkaan	Tidak Ada				
13	3	Sistem ekonomi	N VIAN BURHANUDIN L ; DWI CAHYO ADHI ; FITRIYANTI SYAH; TINA RETNANINGTYAS; NUR SABAR SETIANI;				
14	4	Perilaku Konsumen	N VIAN BURHANUDIN L ; DWI CAHYO ADHI ; ISHMATUL FATHIYAH; BAGUS PRATOMO ARIF AL HAKIM ; DWI TITIK PUSPITOSARI; FITRIYANTI SYAH; TINA RETNANINGTYAS; SITI SAONAH;				
15	5	Kebijakan Moneter	DWI CAHYO ADHI ; LILIANA ROMANTIKA SARI ; NING TYAS TRIASTUTI; NURLITA GADISA R;				
16	6	Elastisitas Permintaan	N VIAN BURHANUDIN L ; DWI CAHYO ADHI ; R BRATA				
		Input 01	Input 02	Lap Peserta	Lap Butir	Pola Jawaban Butir	Lap Essay
				Materi Remedial	Peserta Remedial		

Gambar 6. Perbaikan Sheet Laporan Peserta

Hasil uji kelayakan memperlihatkan bahwa software AnBuso yang dikembangkan dalam penelitian ini terbukti dinilai sangat layak. 51% responden menyatakan layak dan 46% menyatakan sangat layak sementara yang lainnya menyatakan cukup layak. Hal ini menunjukkan bahwa software yang dikembangkan ini memiliki kebermanfaatan yang tinggi dalam membantu guru untuk melakukan analisis butir soal.

Aspek kelayakan yang dinilai paling tinggi adalah aspek kepraktisan dan kemudahan, dan aspek kebermanfaatan. Sementara aspek yang dinilai paling rendah adalah aspek tampilan. Hal ini mengindikasikan bahwa software AnBuso dianggap praktis dan mudah untuk digunakan serta bermanfaat dalam membantu melakukan analisis butir soal.



Gambar 7. Hasil Uji Kelayakan Software

Walaupun software AnBuso ini dinilai layak namun dalam kenyataannya ada beberapa kendala yang dihadapi guru dalam menggunakan software ini. Dari sisi manfaat yang dihasilkan dari software ini memang sangat baik namun masih ada beberapa guru yang kurang mahir dalam penguasaan komputer, khususnya program Microsoft Excel. Umumnya pengetahuan guru terhadap program ini sangatlah terbatas. Masih banyak menu dan fasilitas yang disediakan Microsoft Excel namun belum dimanfaatkan secara optimal. Oleh karena software ini terintegrasi dengan program Excel maka akibatnya guru kurang lancar dalam memanfaatkan software ini.

Kesediaan dan kebiasaan guru dalam melakukan analisis butir soal juga masih dalam kategori jarang. Hanya sedikit guru yang selalu melakukan analisis butir soal baik terhadap soal harian yang dibuatnya, soal semeseteran maupun soal ujian akhir. Karena kebiasaan mereka tersebut akibatnya guru kurang lancar dalam melakukan analisis butir soal karena memang belum terbiasa.

Pengetahuan dan penguasaan guru tentang konsep analisis butir soal juga masih terbatas. Sementara pengembangan software ini juga didasarkan pada konsep analisis butir soal terutama analisis klasik, akibatnya pemahaman guru terhadap angka-angka yang dihasilkan dari software masih kurang. Walaupun hasil analisis sudah dikemas dalam bentuk laporan yang siap ditandatangani, namun masih perlu dijelaskan tentang arti dan makna dari hasil analisis tersebut.

Masih ada beberapa tuntutan guru yang belum mampu diakomodasikan dalam software ini. Walaupun software ini sudah memadai namun masih ada beberapa keterbatasan yang dihadapi dalam software ini, yaitu:

1. Software ini belum merupakan program aplikasi yang berdiri sendiri yang cara menjalankannya harus melalui instalasi, namun software ini masih menjadi bagian integral dalam program Microsoft Excel. Akibatnya software ini tidak akan dapat dijalankan jika tidak menggunakan program Excel.
2. Keterbatasan dalam hal jumlah peserta tes/siswa. Jumlah maksimal yang bisa dianalisis hanya sebanyak 50 siswa sehingga tidak dapat digunakan untuk analisis yang peserta tesnya melebihi 50.
3. Keterbatasan jumlah soal. Jumlah soal pilihan ganda hanya terbatas maksimal 50 soal dan soal essay maksimal 10 soal. Hal ini tentu membatasi pengguna jika soal yang dimiliki melebihi batas yang ditentukan.
4. Tidak tersedia soal dengan kode yang berbeda (soal paralel) sehingga sulit diaplikasikan untuk keperluan soal yang memiliki lebih dari satu jenis soal paralel dengan seri yang berbeda.

#### **D. Kesimpulan**

Penelitian ini berhasil mengembangkan software AnBuso yang dapat dimanfaatkan guru dalam melakukan analisis butir soal secara praktis dan aplikatif. Software ini dibuat dengan program Microsoft Excel yang di dalamnya terdapat sheet untuk input data, sheet data processing, sheet laporan hasil analisis dalam bentuk tabel dan gambar.

Software yang dihasilkan terbukti sangat layak oleh guru dilihat dari aspek kepraktisan dan kemudahan, aspek kebermanfaatan dan aspek substansi isi serta aspek tampilan. Walaupun demikian, ada beberapa kendala yang dihadapi guru dalam menggunakan software ini seperti lemahnya penguasaan guru terhadap program Microsoft Excel, kurang terbiasanya melakukan analisis butir soal, pemahaman konsep analisis butir soal yang terbatas, dan kendala teknis yang terdapat dalam software.

## Daftar Pustaka

- Allen, M.J. & Yen, W.M. (1979). *Introduction to measurement theory*. California: Wadsworth, Inc.
- Azwar, Syaifuddin (2012) *Dasar-dasar Psikometri*. Yogyakarta: Pustaka Pelajar
- Cohen–Swerdlik (2009) *Psychological Testing and Assessment: An Introduction to Tests and Measurement 7<sup>th</sup> Edition*. New York: McGraw–Hill
- Depdiknas (2004). *Pedoman Umum Pengembangan Penilaian Kurikulum Berbasis Kompetensi SMA*. Jakarta: Direktorat Pendidikan Dasar dan Menengah, Direktorat Pendidikan Lanjutan Pertama.
- Djaali & Mulyono, Pudji. (2007). *Pengukuran dalam Bidang Pendidikan*. Jakarta: Grasindo
- Ebel, Robert L. & David A. Frisbie (1991) *Essential of Educational Measurement (5<sup>th</sup> Edition)*. New Delhi: Prentice-Hall, Inc.
- Gall, M.D., Gall, J.P & Borg, W.R. (2003) *Educational Research: An Introduction*. Boston: Pearson Education. Inc.
- Gorin, J.S. (2007) “Reconsidering Issues in Validity Theory”. *Educational Researcher*, 36(8), 456-462.
- Gronlund, Norman E. (1985). *Measurement and Evaluation in Teaching - Fifth. Edition*. New York: Macmillan Publishing Company
- Kusaeri dan Suprananto (2012) *Pengukuran dan Penilaian Pendidikan*. Yogyakarta: Graha Ilmu
- Mardapi, Djemari, dkk (2010). “Pengembangan Model Penjaminan Mutu Perbaikan Hasil Ujian Nasional SMP”. *Laporan Penelitian*. Jakarta: Balitbang Diknas.
- Miles, M.B. & Huberman, A.M. (1994). *Qualitative data analysis: an expanded sourcebook (2<sup>nd</sup> edition)*. California: Sage Publication Inc.
- Moss, P.A. Girard, B.J. & Haniford, L.C. (2006) “Validity in Educational Assessment”. *Review of Research in Education*, 30, 109-162.
- Niron, Maria D., dkk (2005). “Pendampingan Manajemen Pelaksanaan Kurikulum Berbasis Kompetensi Di Sekolah Se Propinsi DIY” *Laporan Penelitian tidak diterbitkan*. Yogyakarta: FIP UNY
- Nunnally, Jum C. & Ira H. Bernstein (1994) *Psychometric Theory (3<sup>rd</sup> Edition)*. New York: McGraw-Hill, Inc.
- Purwanti, Endang. (2008). *Asesmen Pembelajaran SD*. Jakarta: Direktorat Jenderal Pendidikan Tinggi Departemen Pendidikan Nasional.
- Setiaji, Khasan (2010) “Model Implementasi Kuliah Kerja Nyata Tematik Penuntasan Butaakasa Universitas Negeri Semarang Tahun 2008 di Kecamatan Bandar Kabupaten Batang” *Lembaran Ilmu Kependidikan*. (39) 1. Hal 10-14
- Sudijono, Anas (2011) *Pengantar Evaluasi Pendidikan*. Jakarta: RajaGrafindo Persada
- Sudjana, Nana (2011) *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya

- Sukardi (2011) *Evaluasi Pendidikan: Prinsip & Operasionalnya*. Jakarta: Bumi Aksara
- Widoyoko, Eko Putro (2009) *Evaluasi Program Pembelajaran*. Diambil dari <http://www.umpwr.ac.id/web/download/publikasi-ilmiah> pada tanggal 22 Maret 2012
- Wrightstone, J.W. (1956). *Evaluation in modern education – First edition*. New York: American Book Company.