

# **MGMP BIOLOGI DAN BAHASA INDONESIA SMA/MA MENGATASI PERMASALAHAN UNTUK PENCARIAN BUKTI EMPIRIS BENTUK SOAL YANG DIUJIKAN DI SEKOLAH**

**Oleh:**

**Bambang Subali dan Pujiati Suyata  
Universitas Negeri Yogyakarta  
email: b\_subali@yahoo.co.id**

## **Abstract**

The ibM community service aims at training the MGMP teachers of biology & Indonesian language subject in Yogyakarta municipality and Sleman regency in order to have ability to develop test which is composed of divergent and convergent test items. Also, the teachers are able to perform qualitative and quantitative analysis to the developed items.

To meet the targets, this community service was conducted using lecturing; questioning and answering; and practicing techniques. With regard to this, the practicing technique include: (1) writing test items, (2) performing deep understanding to the quality of test items qualitatively, and (3) analyzing items using Quest program. Then, the researches guided the teachers through email to develop a test which were used in their schools. The last activity was analysing their data using Quest program.

The findings of this activity, the teachers had successfully produced a blue print and a set of sample items both divergent and convergent model and it was review between the participants. After each participants conducted of the test administration on each school, the result of the Quest analysis showed that some items of the test was conducted by 21 MGMP teachers of biology subject & 16 MGMP teachers of Indonesian language subject didn't fit with the partial credit model. That is because not all teachers are able to develop a good item and also the minimum of samples was tested by politomus scale items. As a result, estimates do not achieve the expected results.

**Key word:** *convergent test items, divergent test items, test item development, testing, and item analysis*

## **A. PENDAHULUAN**

### **1. Analisis Situasi**

Delapan standar penilaian pendidikan di jenjang pendidikan dasar dan menengah telah ditetapkan. Salah satunya adalah berkaitan dengan standar pendidik dan tenaga kependidikan lainnya. Dengan demikian, peningkatan profesionalisme guru dalam rangka menunjang kualitas penyelenggaraan pendidikan pada setiap satuan pendidikan semakin menjadi tuntutan sejalan dengan peningkatan layanan pendidikan. Menurut Standar Penilaian Pendidikan sebagaimana diatur dalam Lampiran Surat Keputusan Menteri Pendidikan Nasional Nomor 20 Tahun 2007, salah satu prinsip penilaian yang diterapkan dalam pendidikan dasar dan menengah adalah penilaian beracuan patokan. Hal ini berarti bahwa sistem penilaian yang diterapkan adalah sistem penilaian berbasis kompetensi. Selanjutnya, dalam Lampiran Surat Keputusan Nomor 20 Tahun 2007 tersebut dinyatakan bahwa dalam melakukan penilaian, instrumen penilaian hasil belajar yang digunakan pendidik harus memenuhi persyaratan seperti berikut: (1) Substansi adalah merepresentasikan kompetensi yang dinilai. (2) Konstruksi adalah memenuhi persyaratan teknis sesuai dengan bentuk instrumen yang digunakan. (3) Bahasa adalah menggunakan bahasa

yang baik dan benar serta komunikatif sesuai dengan taraf perkembangan peserta didik. Selanjutnya, dinyatakan bahwa instrumen penilaian yang digunakan oleh satuan pendidikan dalam bentuk ujian sekolah/madrasah memenuhi persyaratan substansi, konstruksi, dan bahasa, serta memiliki bukti validitas empirik. Oleh karena itu, peningkatan profesionalisme guru baik kedudukannya sebagai pendidik maupun kedudukannya sebagai bagian dari komponen di dalam satuan pendidikan merupakan kebutuhan yang tidak dapat dielakkan. Sementara di lapangan, baik guru maupun satuan pendidikan masih belum sepenuhnya memiliki kemampuan untuk persyaratan tersebut, terlebih dengan adanya kebijakan dilaksanakannya ulangan umum bersama sebagai salah satu bentuk pengukuran pada akhir semester atau kenaikan kelas. Hasil ujian sekolah juga belum disusun sejalan dengan standar penilaian pendidikan pada jenjang pendidikan dasar dan menengah.

Keputusan Menteri Pendidikan Nasional yang menetapkan bahwa mulai tahun 2011 kelulusan peserta didik ditentukan berdasarkan nilai Ujian Nasional dan Nilai Ujian Sekolah juga semakin menuntut satuan pendidikan benar-benar profesional untuk mengembangkan instrumen

yang digunakan dalam ujian sekolah, termasuk dalam hal ini adalah ujian sekolah untuk mata pelajaran Biologi.

Hasil penelitian Hibah Strategis nasional yang dilaksanakan oleh Bambang Subali dan Pujiati Suyata tentang “Standardisasi Penilaian Hasil Belajar Pola Konvergen dan Divergen Berbasis Satuan Pendidikan pada Jenjang SMA” menunjukkan bahwa hasil tahun I (2010) sebanyak 30 guru yang dijadikan penulis soal dan sebagai subjek untuk mengumpulkan dan melaksanakan analisis data secara empiris terhadap kualitas item tes yang disusunnya tidak memahami prinsip pengembangan tes pola divergen dan tidak pernah mengenal analisis item tes untuk tes dengan pola divergen. Hal yang sama ketika dilakukan diseminasi pada tahun II (2011) di tiga provinsi yakni, DIY, Kalbar, dan NTB. Meskipun guru sudah terbiasa menyusun item tes pola konvergen namun tidak pernah memiliki pengalaman untuk menyusun item tes pola konvergen yang digunakan untuk ulangan umum karena ulangan umum dilaksanakan dalam bentuk Ulangan Umum Bersama (UUB) dan belum pernah menyusun item tes pola konvergen untuk kepentingan ujian sekolah (USEK), karena USEK untuk mata pelajaran yang diujikan dalam ujian nasional (UN) baru dilaksanakan mulai tahun

2011. Dalam penelitian tersebut juga terungkap bahwa para guru di tingkat satuan pendidikan juga tidak memahami prinsip pengembangan item tes beracuan kriteria sebagai tuntutan Kurikulum Tingkat Satuan Pendidikan (KTSP) berdasarkan Lampiran Peraturan Menteri Pendidikan Nasional (Permendiknas) Nomor 20 Tahun 2007. Keadaan ini menjadi kelemahan yang mendasar yang akan menjadi ancaman yang serius bagi kualitas pelaksanaan Ujian Akhir Sekolah (UAS) dan UKK yang dilaksanakan dalam bentuk UUB juga pelaksanaan USEK untuk mata pelajaran yang diujikan melalui UN, khususnya bagi guru yang bertanggung jawab menyusun item tes yang diujikan. Dalam hal ini, termasuk mata pelajaran Biologi dan Bahasa Indonesia yang dituntut untuk diujikan melalui USEK, karena nilai USEK dan nilai rapor memperoleh porsi sebanyak 40% sedangkan UN sebanyak 60% untuk menentukan nilai akhir peserta didik.

## **2. Rumusan Masalah**

Permasalahan utama yang dihadapi mitra dalam hal ini yaitu guru sebagai komponen utama pelaksana UAS, Ulangan Kenaikan Kelas (UKK), dan USEK di tingkat satuan pendidikan menengah (SMA dan MA) berupa minimnya pengalaman dalam me-

ngembangkan item soal/tes yang diujikan yang memiliki bukti empiris sehingga dapat memberikan informasi kepada publik bahwa item tes yang diujikan benar-benar *fit* dengan model. Permasalahan kedua, guru kurang memiliki pengetahuan untuk menyusun item tes khususnya pola divergen, seratatidak adanya kemampuan untuk melakukan analisis item tes secara empirik untuk item pola konvergen dan divergen menggunakan prosedur analisis item dengan menerapkan pendekatan modern, yakni yang didasarkan pada *item respons theory*.

### 3. Tujuan Kegiatan Pengabdian

Tujuan kegiatan PPM berupa penerapan Iptek bagi Masyarakat (IbM) ini adalah peningkatan kegiatan pengembangan ilmu dan teknologi di perguruan tinggi yang berkaitan dengan peningkatan profesionalisme guru di sekolah dalam sistem penilaian berbasis kompetensi melalui standarisasi penilaian hasil belajar pola konvergen dan divergen yang digunakan dalam ulangan akhir semester (UAS), ulangan kenaikan kelas (UKK), juga Ujian Sekolah (USEK) khususnya untuk mata pelajaran Biologi dan Bahasa Indonesia di SMA dan MA. Terlebih dengan adanya keputusan Menteri Pendidikan dan Kebudayaan bahwa USEK mulai tahun 2011 harus

dilaksanakan sebagai pendamping ujian nasional (UN) untuk menetapkan nilai akhir kelulusan peserta didik SMA DAN MA. Hasil penerapan IbM ini diharapkan dapat dipublikasikan melalui jurnal nasional terakreditasi agar dapat dijadikan rujukan bagi satuan pendidikan dalam menyiapkan UAS, UKK, dan USEK.

Dengan terlatihnya guru SMA dan MA dalam melaksanakan penilaian hasil belajar pola konvergen dan divergen berbasis satuan pendidikan di SMA dan MA yang terstandarkan, maka sistem penilaian di tingkat satuan pendidikan akan memenuhi tuntutan standar penilaian pendidikan khususnya untuk pendidikan menengah.

### 4. Kajian Pustaka

Kegiatan pengembangan instrumen penilaian merupakan bagian yang tak terpisahkan dari kompetensi Pendidik. Pendidik harus mengakses apa yang dipandang penting, bukan sekedar melakukan apa yang dengan mudah dalam melakukan pengukuran (Bacon, 1995: 85-90). Beragam instrumen penilaian dapat digunakan. Keberagaman instrumen dapat ditilik berdasar fungsi, konten/isi, bentuk item, teknik pengujian, sistem penskoran, dan interpretasi terhadap hasil. Berdasarkan cara menginterpretasikan

hasil pengukuran dapat dibedakan menjadi tes acuan patokan atau *criterion reference test* (CRT) dan tes acuan norma atau *norm reference test* (NRT). CRT digunakan untuk menetapkan status individu berkenaan dengan standar pencapaian kompetensi yang ditargetkan. NRT digunakan untuk memisahkan antara kelompok atas/sukses dan kelompok bawah/gagal sehingga lebih mengarah kepada perbandingan grup. Standar Penilaian Pendidikan sebagaimana diatur dalam Lampiran Surat Keputusan Menteri Pendidikan Nasional Nomor 20 Tahun 2007 menunjukkan bahwa salah satu prinsip penilaian yang diterapkan dalam pendidikan dasar dan menengah adalah penilaian beracuan patokan. Hal ini menunjukkan bahwa bahwa sistem penilaian yang diterapkan dalam implementasi KTSP adalah sistem penilaian berbasis kompetensi.

Pengembangan instrumen penilaian berbentuk tes dilakukan melalui beberapa tahapan, yakni: (1) perancangan tes, (2) uji coba tes, (3) penetapan validitas, (4) penetapan reliabilitas, dan (5) interpretasi skor tes. Perancangan tes tercakup di dalamnya yakni: (1) penetapan tujuan, (2) penyiapan tabel spesifikasi, (3) menyelesaikan format item yang sesuai, (4) menulis item, dan (5) mengedit item. Uji coba instrumen penilaian meliputi

kegiatan: (1) pelaksanaan uji coba, (2) analisis item hasil uji coba, dan (3) penyiapan format perangkat tes yang sudah siap pakai untuk pengujian. Tantangan terbesar dalam mengembangkan instrumen penilaian justru pada lemahnya pemahaman tentang struktur dari substansi pengetahuan yang akan diukur (Ebel & Fresbie, 1986: 32-36).

Pengembangan instrumen penilaian yang mengacu pada acuan kriteria dilakukan melalui tiga langkah, yaitu: (1) mengemukakan maksud/tujuan pembelajaran; (2) menspesifikasikan domain prestasi yang mencerminkan maksud/tujuan pembelajaran; dan (3) mengembangkan item tes. Langkah ketiga mencakup di dalamnya teknologi penulisan item, sedangkan dua langkah sebelumnya mencakup syarat-syarat yang harus dipikirkan dengan seksama-sifat tujuan pembelajaran beserta spesifikasinya, khususnya kaitannya dengan tujuan pembelajaran yang ditetapkan (Roid & Haladyna, 1982: 43-47). Stark et al. (2001) menjelaskan prosedur pengembangan instrumen menggunakan teori respons item (*item response theory* atau IRT) dapat mengatasi keterbatasan-keterbatasan CTT.

Suatu instrumen penilaian berupa tes prestasi ditujukan untuk mencapai representasi yang memadai guna

mewakili domain prestasi yang bersangkutan yang akan diukur. Hal ini merupakan permasalahan validitas isi yang dihadapi pengembang tes sebagaimana dikemukakan oleh Cronbach tahun 1970 juga Helmstadter tahun 1964. Permasalahan validitas juga berkaitan dengan pemilihan jenis format yang paling tepat yang memungkinkan jumlah maksimum item yang mencerminkan sampel pengukuran dalam suatu periode waktu tertentu yang berkaitan dengan prestasi yang diukur (Roid & Haladyna, 1982:43-47).

Pemenuhan validitas konstruk juga menjadi permasalahan dalam melakukan pengukuran. Validitas konstruk adalah sentral untuk menetapkan mutu instrumen penilaian (Embretson & Gorin, 2001: 343-368). Definisi konstruk yang komprehensif akan memfokuskan pengembangan item tes kaitannya dengan *ability* (kemampuan) dan *trait* (diri spesifik) yang akan diukur. Ada tiga jenis model kognitif, yakni model kognitif penguasaan domain, model kognitif spesifikasi test, dan model kognitif performans tugas (Gorin, 2006: 21-33).

Kualitas instrumen tidak terlepas dari pemvalidasian secara empirik. Banyak faktor yang dapat menimbulkan kekeliruan (*error*) ketika dilakukan pengukuran. Meskipun pengembangan tes sudah dilakukan de-

ngan benar namun ada sumber kekeliruan yang pokok khususnya yang berkaitan dengan analisis pengujian, yang dapat dibagi menjadi tiga subelemen, yakni: penskoran, analisis item, dan penyamaan/*equating* (Allalouf, 2007: 36-43).

Pengembangan tes juga tidak dapat terlepas dari permasalahan reliabilitas atau kehandalan perangkat tes yang disusun. Reliabilitas diestimasi dengan koefisien yang menggambarkan derajat relatif kesalahan pengukuran dalam skor-skor tes. Kesalahan pengukuran adalah perbedaan antara skor manapun dengan skor benar (*true score*). Skor benar adalah hasil yang diperoleh jika semua butir soal dalam domain itu dikerjakan dengan benar (Roid & Haladyna, 1982: 43-47).

Perhitungan reliabilitas yang dijelaskan dalam teori tes klasik oleh Stark et al. (2001) yang telah dipaparkan berkaitan dengan pengembangan tes menurut teori tes klasik adalah mengacu kepada NRT yang berbeda bila tes yang digunakan adalah CRT. Subkoviak (1988: 47-55) menyajikan prosedur untuk membantu guru dan praktisi dalam menerapkan metode untuk memperoleh indeks reliabilitas untuk tes penguasaan (*mastery test*) seperti koefisien persetujuan (*agreement coefficient*) dan koefisien kappa dengan

sekali pengadministrasian. Sementara Kupermintz (2004:193-204) menyajikan suatu prosedur untuk menghitung reliabilitas tes acuan norma jika data-nya berupa data kategorik.

Di dalam IRT, varians kekeliruan (*error*) berkebalikan dari fungsi informasi. Artinya semakin besar kekeliruan semakin jelek informasi hasil pengukuran yang diperoleh. Secara lebih terperinci, untuk semua penempuh ujian dengan kemampuan tetap yang sama, besarnya nilai variansi kekeliruan atau ragam galat akan berkebalikan dengan fungsi informasi tetap yang diperoleh. Logika untuk perspektif ini bahwa variansi kekeliruan (ragam galat) tidak ada hubungannya dengan interpretasi semantik perkataan "informasi" atau sebagai suatu yang merupakan kebalikannya (*invers*) jika nilai teta berbeda.

Smith (2003: 26-32) memasalahkan reliabilitas pada pengukuran-pengukuran yang berupa asesmen kelas. Dalam hal ini asesmen kelas dalam konteks penilaian beracuan kriteria karena perubahan dari asesmen tradisional ke asesmen alternatif dalam setting asesmen kelas, terutama dalam kaitannya dari segi reliabilitas. Subkoviak (1988) telah menemukan prosedur sekali pengujian/pengadministrasian untuk melakukan analisis secara empiris menentukan indeks per-

setujuan (*agreement index*) atau indeks Kappa untuk memenuhi persyaratan reliabilitas item beracuan kriteria.

Instrumen penilaian kemampuan berpikir konvergen dapat diuji menggunakan instrumen bentuk pilihan. Sementara kemampuan berpikir divergen hanya dapat diukur dengan menggunakan item-butir tes dalam bentuk uraian terbuka atau uraian non-objektif atau uraian terbuka. Menurut Roid & Haladyna (1982:58-62) jawaban yang diharapkan dalam tes uraian terbuka harus luas dan komprehensif. Tes semacam ini cocok pula untuk mengukur kreativitas karena testi diminta menggabung-gabungkan jawaban yang sebelumnya telah didefinisikan/ditetapkan. Tes yang demikian ini dimaksudkan sesuai dengan pengetahuan komprehensif yang harus diuji, dan jawaban terhadap setiap item tes menjamin bahwa pengetahuan kognitif yang diukur adalah umum dan luas.

Prinsip-prinsip dan langkah-langkah untuk menuliskan item uraian terbuka sama dengan prinsip dan langkah penulisan uraian testruktur/jawaban singkat. Akan tetapi, yang lebih penting lagi ialah penyiapan jawaban model atau rubrik. Selain berfungsi sebagai petunjuk untuk menskor hasilnya, jawaban model berfungsi mendidik siswa mengenai kekurangan-

kekurangan dalam jawaban-jawaban mereka.

Tantangan yang dihadapi dalam setiap pengukuran akan berkaitan dengan panjang instrumen penilaian dalam bentuk tes yang akan diujikan dan dan banyaknya kriteria yang digunakan untuk menskala respons yang diberikan kepada testi. Selain itu, dalam pembakuan item-itemnya ukuran sampel juga ikut menentukan tingkat kestabilan yang dicapai.

Berkaitan dengan penetapan model, apakah model dikotomus atau model politomus, model politomus tiga kategori atau lebih dari tiga kategori, akan menentukan cara analisis secara empiris yang akan dikenakan. Model dikotomus akan berbeda dengan model politomus. Jawaban pada model yang dipilih juga harus mengandung elemen-elemen isi pertanyaan tes uraian. Jika isinya perlu diatur dengan cara tertentu, kaitannya dengan banyaknya kategori yang digunakan dalam model politomus yang akan digunakan, maka penskoran harus mencerminkan pengaturan/kaidah yang dimaksud. Oleh karena itu, poin-poin jawaban atau rubrik dalam penskoran sangat diperlukan keberadaannya.

Menurut Han & Hambleton (2007: 15-20) juga Theissen et al. (2001: 295-325), dalam model-model

respons butir dikotomus, hanya jenis data responsnyalah yang biner (yaitu, 0 atau 1). Namun demikian, dalam beberapa situasi tes, respons-respons dapat jadi lebih dari dua kategori. Sebagai contoh, suatu kuesioner yang menanyakan sikap (*attitude*), dengan menggunakan butir skala Likert, mungkin menghasilkan respons 5 kategori (sangat tidak setuju, tidak setuju, setuju, dan sangat setuju, yang dapat diberi kode dari 0 hingga 4).

Analisis item menggunakan program Quest, yang panduan bagi guru sudah dikembangkan melalui penelitian hibah kompetitif strategis nasional, memberikan layanan untuk menganalisis data hasil pengujian yang menggunakan skala dikotomus, skala politomus, dan kombinasi keduanya. Program ini juga ada yang tidak *under window* dengan langkah-langkah yang praktis sehingga jika dilatihkan kepada para guru yang sudah mengenal komputer tidak akan ada hambatan. Program ini juga tidak menuntut banyaknya replikasi yang besar (Adam & Kho, 1996). Dengan demikian, program Quest akan dapat dimanfaatkan untuk membantu para guru di lapangan untuk menganalisis hasil pengujian dalam konteks pengembangan instrumen penilaian hasil belajar yang terstandarkan menggunakan prinsip teori repons item.

Berdasarkan pertimbangan secara teoretik sebagaimana telah dipaparkan di atas maka permasalahan belum profesionalnya guru di lapangan sehubungan dengan tuntutan standar nasional penilaian pendidikan dalam melaksanakan USEK untuk mata pelajaran yang diujikan melalui UN, baik dalam mengembangkan item tes untuk USEK mata pelajaran yang bersangkutan, dalam program ini adalah untuk mata pelajaran Biologi, maupun tanggung jawab satuan pendidikan, dalam hal ini SMA dan MA, memenuhi bukti empiris untuk mendukung validitas item tes yang digunakan dalam USEK akan dapat diatasi dengan melatih para guru untuk mengatasi dua permasalahan utama tersebut. Hal ini mengingat komponen utama dalam pelaksanaan USEK ada di tangan guru pengampu mata pelajaran yang bersangkutan.

## **B. METODE PENGABDIAN**

Pada kegiatan pengabdian kepada masyarakat berupa kegiatan IBM ini, melibatkan empat MGMP, yakni MGMP Biologi SMA dan MA Kota Yogyakarta, MGMP Bahasa Indonesia SMA dan MA Kota Yogyakarta, MGMP Biologi SMA dan MA Kabupaten Sleman, dan MGMP Bahasa Indonesia SMA dan MA Kabupaten Sleman.

Kerangka pemecahan masalah pada program I<sub>b</sub>M ini akan diawali dengan berkoordinasi dengan MGMP Biologi SMA dan MA Kota Yogyakarta, MGMP Bahasa Indonesia SMA dan MA Kota Yogyakarta, MGMP Biologi SMA dan MA Kabupaten Sleman, dan MGMP Bahasa Indonesia SMA dan MA Kabupaten Sleman untuk menunjuk calon peserta. Masing-masing MGMP sebanyak 10 orang guru yang memiliki keahlian menggunakan komputer. Dengan demikian, jumlah peserta sebagai khlayak sasaran untuk kegiatan PPM penerapan I<sub>b</sub>M sebanyak 40 orang guru. Kemudian mengundang calon peserta dengan tembusan izin kepada Kepala Sekolah tempat guru bekerja untuk mengikuti kegiatan program PPM penerapan I<sub>b</sub>M ini. Peserta ditempatkan di Laboratorium Komputer Jurusan Pendidikan Biologi FMIPA UNY untuk mengikuti seluruh kegiatan yang dirancang tim pengabdi.

Kegiatan yang diikuti oleh peserta program PPM penerapan I<sub>b</sub>M ini meliputi: (1) kegiatan ceramah dan tanya jawab cara mengembangkan item tes pola konvergen dan divergen; (2) kegiatan ceramah dan tanya jawab analisis item tes secara kualitatif; (3) ceramah dan tanya jawab analisis kuantitatif menggunakan program Quest terhadap tes yang disusun guru;

(4) kegiatan praktik mengembangkan item tes pola konvergen dan divergen, (5) kegiatan praktik analisis item tes secara kualitatif, dan (6) kegiatan praktik analisis kuantitatif menggunakan program Quest terhadap item tes dengan data simulasi dan dengan data berdasarkan hasil tes yang disusun guru yang bersangkutan. Pengetahuan penyusunan item tes pola divergen penting untuk peningkatan kualitas USEK di masa depan mengingat tersedia *software* berupa Program Quest untuk melakukan analisis secara empiris dan menggunakan indeks persetujuan atau indeks Kappa untuk memenuhi persyaratan reliabilitas item beracuan kriteria. Dengan demikian, bagi satuan pendidikan yang menargetkan kualitas USEK di atas kemampuan minimal sebagaimana terukur melalui UN akan dapat ditumbuhkembangkan.

Metode yang dipilih dalam kegiatan ini yakni model ceramah, tanya jawab, dan praktik. Dalam hal ini, kegiatan praktik meliputi (1) penulisan item tes; (2) praktik menelaah kualitas item tes secara kualitatif; dan (3) praktik menganalisis data menggunakan program Quest. Kemudian dilanjutkan dengan kegiatan pendampingan untuk menyiapkan tes yang akan diterapkan di sekolahnya masing-masing. Kegiatan terakhir menganalisis

data milik tiap sekolah menggunakan program Quest.

Evaluasi kegiatan didasarkan pada hasil kinerja para peserta berupa: (1) kualitas item tes hasil belajar yang telah disusun; (2) hasil *review* sesama peserta atas item tes yang disusun sebagai hasil analisis kualitatif; (3) hasil analisis item menggunakan program Quest sebagai hasil analisis kuantitatif. Untuk meningkatkan kinerja peserta pelatihan diadakan kegiatan pendampingan dengan memanfaatkan internet, yaitu setiap peserta harus mengirimkan semua produk kerja dan berkomunikasi dengan tim pengabdian melalui email. Dengan demikian, apabila ada kesulitan dan kesalahan dapat segera diatasi sebelum kegiatan berakhir.

### C. HASIL DAN PEMBAHASAN

Kegiatan PPM mono I<sub>b</sub>M tahun ini yang semula akan melibatkan 2 dosen sebagai anggota dan melibatkan 3 mahasiswa, akhirnya hanya melibatkan 1 dosen anggota karena dosen yang bersangkutan berangkat studi lanjut ke Australia. Mahasiswa yang terlibat juga awalnya 3 orang, namun pada kegiatan akhir hanya 2 orang karena salah seorang mahasiswa terpaksa pulang ke Banten untuk merawat orang tuanya yang sakit keras.

Akibat turunnya dana yang sangat terlambat waktunya, kegiatan pelaksanaan PPM berupa kegiatan pendidikan dan latihan dilaksanakan mulai hari Sabtu tanggal 5 Oktober 2010 dengan materi diklat pengembangan soal konvergen dan divergen. Kegiatan tatap muka diklat dilanjutkan dengan materi penggunaan program Quest untuk analisis item secara empiris pada tanggal 6 Oktober 2010. Pada akhir kegiatan, peserta diminta untuk menyusun item tes pola konvergen dan divergen untuk diujikan di sekolah. Selanjutnya guru mengujikan tes yang telah disusunnya dan dilakukan pendampingan dengan email selama penyusunan item tes. Kegiatan tatap muka dilanjutkan pada tanggal 26 Oktober untuk praktik menganalisis data yang telah dimiliki peserta dan tanggal 27 Oktober interpretasi hasil analisis data yang telah dilakukan. Adapun jumlah peserta diklat setelah dilakukan penawaran dan pendaftaran, untuk guru Biologi sebanyak 22 orang dan guru Bahasa Indonesia 16 orang. Hasil akhir kegiatan dikumpulkan tanggal 7 November 2013 dan bagi yang mengalami kesulitan dilakukan bimbingan via email dan jika masih tidak berhasil dilakukan bimbingan langsung di lapangan.

Hal yang sangat mendukung kegiatan PPM IbM ini adalah respons

Ketua MGMP dan Kepala Sekolah, dan para guru yang terlibat pengetahuan guru tentang komputer juga sangat mendukung kelancaran kegiatan diklat. Adapun hal yang dirasakan menghambat adalah terlambatnya pendanaan dan dikurangnya anggaran sehingga selain waktu pelaksanaan kegiatan PPM menjadi sangat pendek juga menjadikan guru mengalami kesulitan untuk melaksanakan pengujian atas perangkat tes yang telah disusun. Dengan waktu pelaksanaan kegiatan yang sangat pendek, maka penjadwalan kegiatan PPM ini bertabrakan dengan kegiatan UTS di masing-masing sekolah, yang pelaksanaannya juga tidak serempak karena jadwal UTS diserahkan kepada masing-masing sekolah.

Jalan keluar untuk mengatasi faktor penghambat yaitu dengan melakukan analisis terhadap hasil tes yang sudah ada, yakni UTS yang item tesnya juga disusun oleh guru yang bersangkutan, kemudian untuk perangkat pola divergen yang disusun diharapkan tetap dapat diujikan dalam bentuk ulangan harian, meskipun ada guru yang belum dapat melaksanakan.

Hasil yang dicapai dalam hal penulisan item, masih ada di antara peserta yang kurang memperhatikan syarat-syarat penulisan item. Hal tersebut dapat dibenahi melalui kegiatan

presentasi dan konsultasi secara langsung antara peserta dengan tim. Penggunaan program Quest untuk analisis item akan menghasilkan item yang masuk dalam kategori *perfect score* jika seluruh testi berhasil mengerjakannya. Sebaliknya, jika tidak ada satu pun testi yang berhasil mengerjakan maka akan diperoleh informasi *zero score*. Analisis item terhadap tes

prestasi memasukkan item yang termasuk ke dalam *perfect score* sebagai item yang *fit* terhadap model karena seluruh siswa akan berhasil mengerjakan suatu item jika semuanya benar-benar berhasil menguasai kompetensi setelah melalui proses belajar. Adapun hasil kerja guru dapat dilihat pada Tabel 1.

**Tabel 1. Hasil Analisis Item Tes untuk Mata Pelajaran Biologi Menggunakan Program Quest**

No	N	L	Jenis Item	Fit Thdp Model	Perfect Score	Unfit	Unfit Zero Score	Keterangan
1.	95	30	25 PG, 5 uraian	24	3	6	0	
2.	30	27	20 PG, 7 uraian	26	0	0	1	
3.	54	40	35 PG, 5 uraian	21	0	1	18	
4.	103	52	45 PG, 7 uraian	51	0	0	1	
5.	24	40	30 PG, 10 uraian	39	0	1	0	
6.	24	50	50 PG		1	7	1	soal divergen tidak sempat diujikan
7.	48	30	25 PG, 5 uraian	27	0	0	3	
8.	9	25	26 PG	6	1	13	6	soal divergen tidak sempat diujikan
9.	23	8	8 uraian	8	0	0	0	
10.	40	25	20 PG, 5 uraian	25	0	0	0	
11.	135	30	25 PG, 5 uraian	30	0	0	0	
12.	41	16	16 BS		2	2	0	soal divergen tidak sempat diujikan
13.	89	50	50 PG	50	0	0	0	soal divergen tidak sempat diujikan
14.	52	35	25 PG, 10 uraian	35	2	0	0	
15.	74	45	45 PG	45	0	0	0	soal divergen tidak sempat diujikan
16.	96	40	40 PG, 5 uraian	43	1	2		soal divergen tidak sempat diujikan
17.	40	20	10 PG, 10 uraian	19	0	1	0	
18.	92	21	21 uraian	20	0	1	0	
19.	39	45	40 PG, 5 uraian	44	0	1	0	soal divergen tidak sempat diujikan
20.	137	50	50 PG	49	0	0	1	soal divergen tidak sempat diujikan
21.	33	15	15 uraian	13	5	1	1	soal divergen tidak sempat diujikan

Tabel 1 menunjukkan bahwa item yang disusun guru untuk ulangan tengah semester dan atau ulangan harian masih ada yang belum semua fit. Hal tersebut selain karena faktor kemampuan guru dalam mengembangkan item juga banyak karena kurangnya sampel/testi yang diuji untuk soal dengan jawaban berskala politomus. Akibatnya estimasi tidak mencapai hasil yang diharapkan. Dengan teknik simulasi telah dicoba, estimasi untuk item berskala politomus tiga kategori

yang stabil dapat tercapai jika jumlah testi paling sedikit sebanyak 170 orang. Sementara hasil pengujian di lapangan ada yang sangat sedikit.

Hasil kerja guru Bahasa Indonesia disajikan pada Tabel 2. Ada 2 orang guru yang tidak dapat mengikuti kegiatan sampai akhir karena ada kepentingan yang sangat mendesak yang tidak dapat dihindari. Oleh karena itu, dari 18 peserta hanya dapat dilaporkan hasil kerja 16 orang.

**Tabel 2. Hasil Analisis Item Tes untuk Mata pelajaran Bahasa Indonesia Menggunakan Program Quest**

No	N	L	Jenis Item	Fit Thdp Model	Perfect Score	Unfit	Unfit Zero Score	Keterangan
1.	159	24	20 PG, 4 uraian	23	1	0	1	
2.	32	25	25 PG	25	1	0	0	Soal divergen dibuat tetapi tidak sempat diujikan
3.	67	50	50 PG	50	0	0	0	Soal divergen dibuat tetapi tidak sempat diujikan
4.	24	40	40 PG	39	0	0	1	Soal divergen dibuat tetapi tidak sempat diujikan
5.	29	44	40 PG, 4 uraian	40	0	0	0	
6.	32	30	30 PG	30	0	0	1	Soal divergen dibuat tetapi tidak sempat diujikan
7.	96	45	40 PG, 5 uraian	30	0	5	10	
8.	31	55	50 PG	53	0		0	Soal divergen dibuat tetapi tidak sempat diujikan
9.	16	50	50 PG	50	0	0	0	Soal divergen dibuat tetapi tidak sempat diujikan
10.	49	45	40 PG	40	0	0	0	Soal divergen dibuat tetapi tidak sempat diujikan
11.	30	14	10 PG, 4 uraian	14	0	0	0	
12.	96	45	40 PG, 5 uraian	32	0	0	13	
13.	14	50	50 PG	49	0	1	0	Soal divergen dibuat tetapi tidak sempat diujikan
14.	11	35	30 PG, 5 uraian	29	5	5	1	
15.	53	10	10 uraian	10	0	0	0	
16.	96	40	40 PG	16	10	14	0	Soal divergen dibuat tetapi belum sempat diujikan

Tabel 2 juga menunjukkan bahwa item yang disusun guru untuk UTS dan atau ulangan harian masih ada yang belum semua *fit*. Hal tersebut juga disebabkan oleh karena faktor kemampuan guru dalam mengembangkan item juga banyak karena kurangnya sampel/testi yang diuji untuk soal dengan jawaban berskala politomus. Akibatnya estimasi tidak mencapai hasil yang diharapkan. Dengan teknik simulasi, estimasi untuk item berskala politomus tiga kategori yang stabil dapat tercapai jika jumlah testi sebanyak 170 orang. Sementara hasil pengujian di lapangan ada yang sangat sedikit.

Hasil kegiatan menunjukkan diklat yang belum optimal. Selain keadaan ini diakibatkan oleh faktor keterbatasan waktu pendidikan dan pelatihan yang diselenggarakan berkaitan dengan turunnya anggaran untuk pembiayaan, juga akibat pelaksanaan PPM sedang memasuki waktu pelaksanaan ulangan tengah semester. Akibatnya guru tidak dapat langsung mengujikan tes yang telah disusunnya. Kedua adalah faktor motivasi. Sebagaimana dikemukakan oleh Higgins and Kruglanski, 2000 (Pintrich, 2003: 669) bahwa motivasi berkaitan dengan apa yang dilakukan oleh seseorang kaitannya dengan apa yang menjadi kebutuhannya. Sampai saat

ini, belum ada tuntutan riil di lapangan bahwa tes yang digunakan dalam ulangan baik UTS, UAS, maupun Ujian Sekolah (US) yang mewajibkan guru dan sekolah melaporkan kualitas item dari tes yang digunakan pada kegiatan tersebut. Akibatnya, analisis item menggunakan program Quest yang memerlukan penguasaan komputer, program *excel*, program *notepad* sebagai pendukungnya menjadi dirasa berat oleh sebagian peserta. Sebaliknya bagi beberapa guru yang membutuhkan kum untuk kenaikan pangkat sangat antusias dengan kegiatan ini. Bahkan ada 2 guru yang sangat intensif berkonsultasi untuk menguasai program Quest.

Penguasaan pemakaian komputer pada umumnya menjadi salah satu syarat bagi orang yang bekerja yang memerlukan informasi tertulis seperti penguasaan pengetikan sampai dengan informasi yang berkaitan dengan multimedia yang harus dicari di jaringan internat ([http://en.wikipedia.org/wiki/Computer\\_literacy](http://en.wikipedia.org/wiki/Computer_literacy)). Penguasaan komputer bagi peserta yang muda usia umumnya sangat baik. Sebaliknya, bagi sebagian peserta yang berusia di atas 50 tahun tampak adanya kendala dalam mengoperasikan komputer sampai dengan pemakaian programnya.

Hasil kegiatan PPM yang menunjukkan bahwa baik guru pemula maupun guru yang sudah puluhan tahun mengajar menunjukkan tidak ada perbedaan kinerja dalam menyusun kisi-kisi dan soal maupun penulisan soal. Kebiasaan menulis soal dari tahun ke tahun menunjukkan tidak ada perbedaan kualitas soal yang mencolok. Hal ini juga tidak terlepas dari motivasi guru serta belum adanya tuntutan pemerintah terhadap guru dan sekolah untuk memberikan informasi terhadap tes yang diselenggarakan, baik UTS, UAS, dan US.

## D. PENUTUP

### 1. Simpulan

Hasil kegiatan pengabdian pada masyarakat melalui program IBM yang dapat dicapai melalui kegiatan ini adalah sebagai berikut.

1. Telah terlaksana kegiatan diklat tatap muka untuk meningkatkan pemahaman guru dalam mengembangkan item tes pola konvergen dan divergen, juga cara menganalisis item secara kualitatif dan secara kuantitatif. Analisis data secara kuantitatif dilakukan menggunakan pendekatan empiris dengan program Quest.
2. Sampai akhir kegiatan, dari guru Biologi sebanyak 22 orang sudah berhasil menyerahkan hasil kerja

berupa kisi-kisi, item, dan hasil analisis item menggunakan program Quest. Adapun untuk untuk guru bahasa Indonesia dari 18 orang ada dua orang yang mengundurkan diri/tidak dapat mengikuti kegiatan sampai akhir.

3. Hasil analisis Quest yang mengindikasikan adanya item tes yang tidak *fit* dengan model sudah dapat dipahami oleh guru sebagai indikasi kelemahan item tes yang disusun. Hal tersebut juga disadari sebagai hal yang berkaitan dengan pentingnya analisis item untuk mendukung akuntabilitas keprofesionalan guru meskipun sampai saat ini pemerintah belum mewajibkan sekolah melaporkan kualitas item tes yang digunakan dalam pengujian yang diselenggarakan guru maupun sekolah.

### 2. Saran

Berdasarkan hasil yang dicapai disarankan bahwa pihak MGMP selaku mitra kegiatan dapat mengintensifkan kegiatan lanjutan agar guru peserta dapat menguasai pengembangan tes juga menguasai kegiatan analisis item menggunakan program Quest dengan harapan dapat meningkatkan kualitas profesi guru. Terlebih saat sekolah harus melaporkan hasil Ujian Sekolah secara terpisah, tentulah se-

kolah harus dapat menunjukkan secara terbuka berkenaan dengan kualitas tes yang digunakan.

### UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada Direktur Ditlitabmas Ditjen Dikti, Kemendikbud, yang telah mebiayai kegiatan PPM ini menggunakan dana anggaran PPM Mono tahun tahun 2013.

### DAFTAR PUSTAKA

- Adams, R.J. & Kho, Seik-Tom. 1996. *Acer Quest Version 2.1*. Camberwell, Victoria: The Australian Council for Educational Research.
- Allalouf, A. 2007. An NCME Instructional Module on Quality Control Procedures in the Scoring, Equating, and Reporting of test scores. *Educational Measurement, Issues and Practice*. Washington: Spring 2007. Vol. 26, Iss. 1; pg. 36, 8 pgs.
- Bacon, A. 1995. The Teachers' Perspective on Accountability (Versi Elektronik). *Canadian Journal of Education*. Toronto: Winter 1995. Vol. 20, Iss. 1; pg. 85.
- Bambang Subali & Pujiati Suyata. 2010. "Standardisasi Penilaian Hasil Belajar Pola Konvergen dan Divergen Berbasis Satuan Pendidikan pada Jenjang SMA". *Laporan Penelitian DP2M*. Yogyakarta: Jurdik Biologi, FMIPA UNY.
- Ebel, R.L. & Frisbie, D.A. 1986. *Essentials of Educational Measurement*. New Jersey: Prentice Hall, Inc.
- Embretson, S. & Gorin, J. 2001. Improving Construct Validity with Cognitive Psychology Principles. *Journal of Educational Measurement*. Washington: Winter 2001. Vol. 38, Iss. 4; pg. 343, 26 pgs
- Gorin, J.S. 2006. "Test Design with Cognition in Mind". *Educational Measurement, Issues and Practice*. Washington: Winter 2006. Vol. 25, Iss. 4; pg. 21, 15 pgs.
- Gronlund, N.E. 1998. *Assessment of Student Achievement*, 9th. Boston: Allyn and Bacon.
- Kupermintz, H. 2004. "On the Reliability of Categorically Scored Examinations (Versi Elektro-

- nik).” *Journal of Educational Measurement*. Washington: Fall 2004. Vol. 41, Iss. 3; pg. 193, 12 pgs.
- Peraturan Menteri Pendidikan Nasional Nomor 20 Tahun 2007 tentang Standar Penilaian Pendidikan Satuan Pendidikan Dasar dan Menengah.
- Pintrich, P.R. 2003. “A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Contexts”. *Journal of Educational Psychology*, 2003, Vol. 95, No. 4, 667–68. Diunduh Tanggal 1 November 2013.
- Roid, G.H. & Haladyna, Th.M. 1982. *A Technology for Test-Item Writing*. Orlando: Academic Press, Inc.
- Smith, J.K. 2003. “Reconsidering Reliability in Classroom Assessment and Grading (Versi Elektronik)”. *Educational Measurement, Issues and Practice*. Washington: Winter 2003. Vol. 22, Iss. 4; pg. 26, 8 pgs.
- Stark, S., Chernyshenko, S., Chuah, D., Wayne Lee, & Wilington, P. 2001. *IRT Modeling Lab: IRT Tutorial* [Versi elektronik]. Urbana: University of Illinois.
- Subkoviak, M.J. 1988. “A Practitioner's Guide to Computation and Interpretation of Reliability Indices for Mastery Tests”. *Journal of Educational Measurement*. Spring 1988. Vol.25. No. 1. pp. 47-55
- Thissen, D., Nelson, L, & Surygert, K.A. 2001. “Item Response Theory Applied to to Combination of Multiple-Choice and constructed Response Items—Approximation Methods for Scale Score”. Dalam D. Thissen & H. Wainer (Eds.), *Test Scoring*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

