

STANDARDISASI PENILAIAN BERBASIS SEKOLAH

Bambang Subali, Pujiati Suyata
Universitas Negeri Yogyakarta
Jl. Colombo, Karang Malang, Yogyakarta 55281
b_subali@yahoo.co.id, pujiati_suyata@uny.ac.id

Abstrak

Penelitian ini bertujuan mengembangkan panduan acuan standardisasi penilaian hasil belajar berbasis sekolah, kasus pada mata pelajaran Biologi dan Bahasa Indonesia di SMA. Hasil yang dicapai yakni tersusunnya panduan standardisasi penilaian yang terdiri dari (a) *learning continuum* Biologi dan Bahasa Indonesia SMA yang telah ditelaah pakar terkait, (b) Panduan Penyusunan Butir Tes Pola Konvergen dan Divergen, dan (3) Panduan Analisis Data Pengukuran Pendidikan menggunakan Program Quest dengan menggunakan bahasa Indonesia sebagai bahasa pengantar. Ketiganya sudah divalidasi melalui kegiatan diseminasi kepada para guru Biologi dan Bahasa Indonesia di Provinsi Daerah Istimewa Yogyakarta, Kalimantan Barat, dan Nusa Tenggara Barat pada tahun kedua. Kemudian divalidasi lagi melalui kegiatan diseminasi kepada para widyaiswara di LPMP DI Yogyakarta, Jawa tengah, Bangka Belitung, Kalimantan Barat, Kalimantan Selatan, Sulawesi Selatan, dan Nusa Tenggara Barat pada tahun ketiga. Oleh karena itu, ketiga perangkat tersebut sudah dapat dijadikan pegangan bagi para praktisi di lapangan.

Kata kunci: *standardisasi penilaian, learning continuum hasil belajar, CTT dan IRT*

THE STANDARDIZATION OF SCHOOL BASED ASSESSMENT

Bambang Subali, Pujiati Suyata
Universitas Negeri Yogyakarta
Jl. Colombo, Karang Malang, Yogyakarta 55281
b_subali@yahoo.co.id, pujiati_suyat@uny.ac.id

Abstract

This study aims to develop the guidance of the standardization on the achievement assessment based on school, cases on Biology and Indonesian Language subjects in senior high schools. The study resulted standardization guidance consisting of (a) learning continuum of achievement on Biology and Indonesia Language at Senior High Schools; (b) the guidance of writing the instrument to measure convergent and divergent pattern of learning achievement on Biology and Indonesia Language at Senior High Schools; and (c) the guidance to analyse data of educational measurement by using Quest Program. Those have been validated through dissemination to Bilolgi and Indonesian teachers in DIY, Kalbar, and NTB Provinces in the second year, and again by the instructors of LPMP in DIY, Jateng, Ba-Bel, Kalbar, Kalsel, Sulsel, and NTB Provinces at the third year. Thereby all of that guidance can be used by the practitioners in the fields.

Keywords: *standardization of assessment, learning continuum result of learning, CTT and IRT*

Pendahuluan

Meskipun penilaian alternatif dituntut dalam Peraturan Menteri Pendidikan Nasional (Permendiknas) Nomor 20 Tahun 2007 (Depdiknas, 2007), tidak berarti bahwa ujian tulis tidak diperlukan. Ujian tulis masih diterapkan dalam ujian nasional dan ujian distrik di Amerika Serikat (Hargreaves & Schmidt, 2002: 69-95). Dalam hal penggunaan tes tertulis, pengenalan penulisan butir dengan pola konvergen dalam bentuk pilihan dan isian terbatas sudah banyak diperoleh para guru, baik melalui bimbingan teknis (Bimtek) tingkat nasional, regional, maupun *inhouse training*. Demikian pula analisis butir menggunakan pendekatan teori klasik atau *classical test theory* (CTT) serta dengan pendekatan teori respons butir atau *item response theory* (IRT). Sebagai contoh pada CD kumpulan materi Bimtek profesional bulan April 2009 oleh Ditjen Dikdasmen juga dikenal pemakaian analisis butir dengan pendekatan CTT dan IRT. Namun demikian, pengenalan butir pola divergen beserta teknik analisisnya, baik menggunakan CTT maupun IRT hampir tidak pernah dilakukan. Hal tersebut merupakan salah satu efek pemakaian soal bentuk pilihan ganda (PG) yang digunakan dalam UN. Hasil wawancara pendahuluan dengan sebagian besar guru Biologi SMA di Kota Yogyakarta dan guru SMA unggulan di Kabupaten Sleman menunjukkan bahwa memang sebagian guru hampir tidak pernah menggunakan soal uraian.

Menyongsong perkembangan asesmen nasional di negara lain yang juga mengujikan soal bentuk uraian (Gorin, 2006: 21-36), dan bagaimana upaya untuk mengontrol kualitas hasil ujian (Allalouf, 2007: 36-43) beserta permasalahan yang dihadapinya (Brennan, 2001: 6-18), juga rendahnya prestasi siswa dalam mengerjakan soal TIMSS (IEA, 2003), Hasil penelitian Subali (2009: 203) menunjukkan penguasaan siswa Biologi SMA terhadap keterampilan proses sains pola divergen dengan *partial credit model* (PCM) satu parameter logistik (1-PL) tiga kategori tergolong rendah. Bahkan bila diskala dengan skala dikotomis ada yang tidak memiliki kemampuan berpikir divergen menurut model Rasch. Oleh karena itu, perlu dilakukan upaya untuk meningkatkan kemampuan guru dalam melakukan penilaian.

Kelemahan guru dalam mengembangkan penilaian antara lain pada *learning continuum* hasil belajar sesuai dengan konsep pengukuran. *Learning continuum* hasil tersebut merupakan *abstract continuum* yang terbentang dari $-\infty$ sampai dengan $+\infty$. Bentangan tingkat kesukaran dan tingkat kemampuan siswa tersebut berada pada satu garis sehingga akan dapat diketahui posisi setiap subjek terhadap tingkat kesulitan butir yang dikerjakan. Tingkat kemampuan tes maupun tingkat kesukaran butir dalam *Rasch Model* (RM) diekspresikan pada satu garis berupa absis pada grafik dengan satuan berupa logit (*logg-odd unit*) (Keeves & Alagumalai, 1999: 27). Proporsi varians total dari estimasi skala untuk *person* sebesar β_n yang berasosiasi dengan varians parameter, ditentukan oleh besarnya indeks separasi *person* sebesar S. Indeks separasi *person* itulah yang dianggap sama dengan koefisien reliabilitas tes (Wright & Masters: 1982: 106). Namun, tetap harus diperhatikan bahwa perhitungan besarnya *error* pengukuran pada indeks separasi *person* berbeda dengan perhitungan *error* varians pada CTT (Keeves & Masters, 1999: 275-276).

Butir tes dalam bentuk uraian nonobjektif menghasilkan jawaban yang diberikan peserta tes secara luas dan komprehensif (Roid & Haladyna, 1982: 58-62). Hal ini sangat cocok dengan pengukuran keterampilan berpikir divergen yang menyebar dari satu titik (Atherton, 2005: 1). Dengan demikian, akan ada variasi jawaban benar dari testi. Namun demikian banyak kelemahan guru dalam mengonstruksi butir tes (Jehlen, 2007, (29-34), terlebih jika dikaitkan dengan pemenuhan validitas konstruk suatu pengukuran (Edward & Bagozzi, 2000: 155-173; Gorin, 2006: 21-35).

Penskalaan atau penskoran politomus diberikan kepada respons tes uraian karena respons yang muncul dapat diberi poin nilai dengan kisaran performansi yang terendah (nol) hingga lebih dari satu level di atasnya (misalnya 2, 3, atau n). Model yang dapat dipakai untuk mencari karakteristik informasi butir yang terkait dengan penskalaan respons yang muncul cukup banyak. Keragaman model tersebut terdapat baik pada penskalaan politomus maupun dikotomus (Han & Hambleton, 2007: 15-20; Thissen *et al.*, 2001: 295-325).

Ukuran sampel bervariasi sesuai dengan banyak parameter dan model yang digunakan. Ukuran sampel untuk data politomus menggunakan

Graded Response Model (GRM) yang merupakan model dengan dua parameter logistik (2-PL) sekitar 250 dapat diterima untuk aplikasi dalam penelitian, sedangkan 500 sampai 1000 untuk penggunaan operasional (Muraki & Bock, 1998: 35). Untuk keperluan penelitian dapat menggunakan sampel yang kecil (Crocker & Algina, 1986: 322). Ukuran sampel untuk model 1-PL berupa *Rasch Model* (RM) antara 30 sampai 300 dengan batas INFIT t sebesar -2 sampai +2 (Bond & Fox, 2007: 43).

Dalam penelitian ini digunakan model satu parameter dari RM dan PCM dari program Quest. Elemen sentral program Quest adalah IRT dengan *Rasch Model* (RM). Program Quest dalam melakukan estimasi parameter, baik untuk butir maupun testi (*case/person*) menggunakan *unconditional* (UCON) atau *joint maximum likelihood* (Adams & Kho, 1996: 89). Permasalahannya adalah tidak tersedia petunjuk manual pemakaian program Quest dengan pengantar bahasa Indonesia. Panduan Program Quest yang berbahasa Inggris pun sulit dan tidak mudah dipahami bagi pengguna awam. Oleh karena itu, penelitian ini bertujuan untuk memperoleh panduan penilaian yang komprehensif, mulai dari pengembangan *learning continuum*, pembuatan kisi-kisi, penulisan butir soal pola konvergen dan divergen, telaah antarguru sejawat, sampai dengan pemakaian program Quest untuk mencari bukti empiris bahwa butir yang disusun *fit* dengan model 1-PL baik dengan penskalaan politomus, dikotomus, ataupun kombinasinya.

Metode Penelitian

Penelitian ini menggunakan model *research and development*. Riset ditujukan untuk memperoleh model buku panduan komprehensif tentang standarisasi penilaian hasil belajar pola konvergen dan divergen, yang tersusun atas cara pengembangan *learning continuum*, penyusunan kisi-kisi, penulisan butir pola konvergen, dan divergen, telaah guru sejawat, dan analisis empiris menggunakan program Quest untuk mengetahui fit setiap butir dengan model 1-PL. Tahun pertama menggunakan model uji coba terpakai, dalam arti panduan digunakan oleh peserta dan disempurnakan selama pemakaian berlangsung dengan memperhatikan masukan dari pe-

serta. Model penskoran juga menggunakan model uji coba terpakai sehingga butir yang tidak *fit* dengan model dikeluarkan (tidak diperhitungkan) saat menentukan skor siswa. Pada tahun kedua dilakukan sosialisasi dan diseminasi di tiga provinsi yakni Provinsi DIY, Kalimantan Barat, dan Nusa Tenggara Barat.

Item characteristic curve (ICC) akan mendatar (*flat*) bila besarnya INFIT MNSQ untuk butir atau e lebih besar dari satuan logit $>1,30$ akan berakibat membentuk *platokurtic curve*, jika satuan logit $<0,77$ akan terlalu runcing membentuk *leptokurtic curve* (Keeves & Alagumalai 1999: 36). Oleh karena itu, dalam program Quest ditetapkan bahwa suatu butir atau *person* dinyatakan *fit* dengan model dengan batas kisaran INFIT MNSQ dari 0,77 sampai 1,30 (Adams & Kho, 1996:30 & 90). Ada pula peneliti yang menggunakan batas yang lebih ketat, yakni dengan kisaran 0,83 sampai dengan 1,20 dan ada yang menggunakan pengujian berdasarkan besarnya nilai INFIT t , yakni menggunakan kisaran nilai t adalah ± 2 (pembulatan $\pm 1,96$) jika taraf kesalahan/ α sebesar 5% (Keeves & Alagumalai 1999: 34-36; Bond & Fox, 2007: 43).

Hasil Penelitian dan Pembahasan

Pada tahun pertama, kegiatan penelitian ini diawali dengan pengembangan *learning continuum* hasil belajar untuk mata pelajaran Biologi dan bahasa Indonesia SMA oleh tim peneliti. *Learning continuum* dalam hal ini berkedudukan sebagai *abstract continuum* pengukuran yang memuat aspek kompetensi mata pelajaran yang bersangkutan dan merupakan acuan dalam pengembangan instrumen penilaian hasil belajar kedua mata pelajaran tersebut.

Sebelum menyusun *learning continuum* peneliti mengadakan wawancara tentang kedalaman atau tinjauan SK dan KD oleh para guru di lapangan sebagai implementasi dari Permendiknas Nomor 24 Tahun 2006. Dari hasil wawancara dengan sejumlah guru SMAN yang tergolong favorit di Kota Yogyakarta, dan guru SMA Unggulan di Kabupaten Sleman umumnya menggunakan SK dan KD dalam standar isi yang dirumuskan dalam Permendiknas Nomor 22 tahun 2006. Namun, sebagian SMA RSBI sudah

memperkaya kurikulum sekolahnya dengan mengacu kepada kurikulum di negara maju seperti AS.

Learning continuum untuk mata pelajaran Biologi dan Bahasa Indonesia. *Learning continuum* untuk Biologi disusun mulai dari SMP karena setelah dilakukan komparasi dengan tabel spesifikasi *Biological Science Curriculum Study* (BSCS) sebagian materi Biologi SMA menurut rumusan Standar Kompetensi dan Kompetensi Dasar yang ada di Standar Isi (Permendiknas No. 20 Tahun 2006) diajarkan mulai *grade* VI. Selanjutnya, *learning continuum* yang telah disusun ditelaah masing-masing oleh tiga orang pakar.

Learning continuum yang telah disusun selanjutnya ditelaah oleh pakar melalui *Focus Group Discussion* (FGD), dengan tujuan agar terjadi sinkronisasi masukan secara langsung melalui pertemuan tersebut. Dalam hal ini, ada tiga pakar Pendidikan Biologi dan tiga pakar Pendidikan Bahasa Indonesia sebagai penelaah. *Learning continuum* yang telah ditelaah inilah yang dijadikan acuan bagi guru untuk mengembangkan instrumen pengukur hasil belajar.

Masukan dari pakar Pendidikan Biologi selain membenahi posisi KD setelah direlokasikan/diplotkan sesuai dengan tabel spesifikasi BSCS adalah membenahi rumusan KD terutama untuk KD yang dimungkinkan dirumuskan di atas rumusan KD yang ada dalam standar isi menurut Permendiknas No. 22 Tahun 2006 (Depdiknas, 2006). Diskusi secara mendalam dilakukan mengingat sebagian rumusan KD yang ada di level SMA justru dimasukkan dalam KD pada level SMP menurut tabel spesifikasi BSCS. Masukan untuk *learning continuum* bahasa Indonesia dari ketiga pakar Pendidikan Indonesia adalah pada sekuensi KD dan tingkatan rumusan KD berdasarkan sifat hierarkinya. Hal ini dikarenakan memang bahasa Indonesia seperti bidang bahasa pada umumnya memiliki hierarki yang ketat dari suatu kemampuan di bawahnya dengan suatu kemampuan yang harus ada di atasnya atas dasar kompleksitas konten dan tingkat kesukaran kemampuan yang harus didemonstrasikan. *Learning continuum* keterampilan proses sains dalam Biologi menggunakan rumusan Bambang Subali (2009).

Kegiatan selanjutnya tim peneliti menyusun panduan, berupa (1) panduan analisis data pengukuran pendidikan menggunakan program

Quest untuk memperoleh bukti empiris kesahihan setelah panduan disusun, diundang 15 guru Biologi Madrasah Aliah (MA) dan 15 guru bahasa Indonesia SMA yang berasal dari SMA di DIY melalui MGMP.

Ketiga puluh guru tersebut bertanggung jawab menyusun instrumen tes untuk ulangan harian/ulangan tengah semester menggunakan panduan yang telah disusun peneliti. Sebagaimana telah dikemukakan pada metode penelitian bahwa model yang digunakan adalah model terpakai, dalam arti bahwa panduan yang disusun oleh peneliti digunakan sebagai acuan oleh peserta dalam menulis butir soal dan sekaligus disempurnakan atas dasar masukan dari para penulis soal. Oleh karena itu, selama pelaksanaan penulisan butir tes, peneliti memperhatikan masukan-masukan guru untuk menyempurnakan panduan yang disusun peneliti. Masukan yang terbanyak adalah permintaan contoh soal divergen beserta rubrik dan pedoman penskorannya, meskipun sudah diberi rambu-rambunya.

Selanjutnya, butir tes yang telah disusun dirangkai menjadi perangkat tes dan dilakukan telaah silang antarpengulis sebelum soal diujikan. Selanjutnya perangkat tes tersebut diujikan di sekolahnya masing-masing, baik sebagai ulangan harian atau ulangan tengah semester. Setelah dilaksanakan pengujian di sekolahnya masing-masing selanjutnya data hasil ulangan tersebut dianalisis menggunakan program Quest berpedoman panduan yang telah disusun peneliti. Guru peserta menganalisis data berpedoman pada panduan yang telah disusun peneliti. Sama seperti dalam pemakaian panduan penulisan soal, panduan analisis ini juga dikembangkan dengan model uji coba terpakai, dalam arti bahwa panduan yang disusun oleh peneliti digunakan sebagai acuan oleh peserta dalam melakukan analisis data dan sekaligus disempurnakan atas dasar masukan dari para peserta. Oleh karena itu, selama pelaksanaan analisis peneliti merekam masukan dari para guru. Masukan paling banyak adalah agar diberi cara penafsiran hasil analisis program Quest yang lebih operasional. Untuk kepentingan penskoran siswa juga menggunakan model uji coba terpakai, dalam arti bahwa untuk hasil tes juga dicapai untuk menentukan skor siswa dengan cara membuang (tidak memasukkan) butir yang tidak *fit* dalam model.

Hasil analisis menggunakan program Quest terhadap butir tes yang disusun guru Bahasa Indonesia berdasarkan hasil uji coba di lapangan

menunjukkan bahwa dari 15 perangkat tes yang disusun guru dan telah ditelaah silang ternyata 6 perangkat tes semua butirnya memenuhi persyaratan butir tes model 1-PL. Sebanyak 5 perangkat tes hanya memiliki $>0-\leq 5\%$ butir yang tidak memenuhi persyaratan butir tes model 1-PL, ada 3 perangkat tes yang memiliki $>5-\leq 10\%$ butir yang tidak memenuhi persyaratan butir tes model 1-PL dan ada 1 perangkat tes yang memiliki $\geq 10\%$ butir yang tidak memenuhi persyaratan butir tes model 1-PL. Satu perangkat tes yang memiliki butir $\geq 10\%$ tidak memenuhi syarat model 1-PL ada 20%.

Hasil analisis menggunakan program Quest terhadap butir tes yang disusun guru Biologi berdasarkan hasil uji coba di lapangan menunjukkan bahwa dari 15 perangkat tes yang disusun guru dan telah ditelaah silang ternyata 9 perangkat tes semua butirnya memenuhi persyaratan butir tes model 1-PL. Sebanyak 4 perangkat tes hanya memiliki $>0-\leq 5\%$ butir yang tidak memenuhi persyaratan butir tes model 1-PL, ada 1 perangkat tes yang memiliki $> 5-\leq 10\%$ butir yang tidak memenuhi persyaratan butir tes model 1-PL dan ada 1 perangkat tes yang memiliki $\geq 10\%$ butir yang tidak memenuhi persyaratan butir tes model 1-PL. Satu perangkat tes yang memiliki $\geq 10\%$ butir yang tidak memenuhi syarat model 1-PL ada 28%.

Sebagaimana dikemukakan dalam rumusan tujuan bahwa tujuan penelitian ini adalah mengembangkan panduan penilaian hasil belajar mata pelajaran Biologi dan Bahasa Indonesia di SMA sehingga dapat dijadikan pegangan bagi para guru di lapangan untuk mengembangkan penilaian yang terstandarkan yang butirnya dianalisis menurut teori respons butir, terutama bagi guru Biologi dan guru Bahasa Indonesia SMA di seluruh Indonesia.

Hasil utama yang diperoleh tahun pertama ini adalah dikembangkannya instrumen penilaian oleh para guru dengan mengacu kepada *learning continuum* hasil belajar, khususnya untuk mata pelajaran Biologi dan Bahasa Indonesia SMA sebagai *abstract continuum* pengukuran instrumen untuk kedua mata pelajaran tersebut, dan disertai bukti empiris. Kedua adalah tersusunnya panduan penilaian hasil belajar pola konvergen dan divergen berdasarkan teori dan didukung hasil pembuktian empiris instrumen yang telah diperoleh. Ketiga adalah penyusunan panduan penggunaan program

Quest. Perangkat tersebut divalidasi secara empiris melalui uji coba terpakai.

Melihat perangkat tes yang telah berhasil disusun guru dikaitkan dengan hasil analisis secara empiris menggunakan program Quest menunjukkan bahwa semua guru sudah berhasil memiliki pengalaman melakukan penilaian yang terstandarkan, yakni mampu menyusun kisi-kisi berdasarkan *learning continuum* yang tersedia—*learning contonuum* juga telah dirumuskan dan di-*review* pakar—dan menyusun butir berdasarkan kisi-kisi. Selanjutnya, butir yang disusun telah direview silang sesama guru sebidang sebagai persyaratan analisis secara kualitatif. Kemudian setelah mengujicobakan di lapangan, guru menganalisis butir menggunakan program Quest, sehingga diperoleh butir yang cocok dengan model Rasch. Dengan langkah tersebut maka akuntabilitas pengukuran telah dapat terpenuhi sebagaimana yang diharapkan. Pengalaman ini tentu akan sangat membantu guru ketika ditugasi untuk membantu satuan pendidikan melakukan analisis untuk memenuhi bukti empiris di lapangan. Hal ini ke depan sangat dimungkinkan mengingat sudah banyak satuan pendidikan SMA yang melakukan ujian sekolah terhadap mata pelajaran yang diujikan melalui Ujian Nasional (UN) bahkan dengan soal-soal yang lebih sulit tingkatannya dari UN. SMA RSBI yang menggunakan tes dari luar juga dapat diketahui kualitasnya setelah dikerjakan oleh siswanya.

Hasil yang diperoleh oleh para guru peserta juga sudah dapat menunjukkan secara empiris bahwa panduan yang disusun oleh peneliti dan setelah diaplikasikan di lapangan dengan memperhatikan masukan dari para peserta sudah dapat digunakan sebagai acuan kerja para guru. Hasil tahun pertama menunjukkan pula bahwa mayoritas instrumen tes sudah menunjukkan hasil yang memuaskan. Bahkan banyak instrumen tes yang semua butirnya sudah memenuhi syarat model 1-PL. Hal ini berarti bahwa guru sudah dapat memahami panduan yang disusun peneliti. Sebagaimana dikemukakan di latar belakang, bahwa sampai saat ini belum tersedia petunjuk analisis menggunakan program Quest menggunakan pengantar berupa bahasa Indonesia, sehingga belum ada panduan praktis yang dapat dipakai sebagai acuan para guru di lapangan. Dengan hasil yang diperoleh melalui penelitian tahap pertama ini, berarti bahwa panduan penggunaan

program Quest—yang sampai sekarang belum ada terbitan panduan praktis dalam bahasa Indonesia—sudah berhasil digunakan para guru di lapangan.

Hasil tahun kedua, diperoleh beberapa temuan. Temuan lapangan setelah dilakukan sosialisasi, pertama bahwa semua guru peserta menyatakan pada dasarnya selama ini semua guru sudah menyusun kisi-kisi dan menyusun butir tes sesuai kisi-kisi, namun tidak ada guru yang menyatakan pernah meninjau ulang SKL dalam Permendiknas Nomor 23 tahun 2006 maupun SK dan KD dalam Permendiknas Nomor 22 tahun 2006 (Depdiknas, 2006). Selama ini penyusunan kisi-kisi juga mengacu kepada silabus yang dibahas dalam MGMP. Dengan demikian, antarSMA relatif tidak ada perbedaan yang mencolok. Tidak ada guru yang menyatakan pernah menyusun butir tes pola divergen untuk pengukuran hasil belajar, baik pada ulangan harian maupun ulangan umum. Dengan adanya sosialisasi hibah strategis nasional tentang penyusunan soal pola konvergen dan divergen semua guru memiliki persepsi yang positif.

Temuan kedua, bahwa para guru menyatakan bahwa untuk ulangan harian tidak akan ada masalah bila mereka mengenalkan soal pola konvergen dan divergen untuk dipadukan menjadi satu guna mengukur hasil belajar peserta didik. Namun, dalam hal ulangan umum kecil peluangnya akan ada soal pola divergen mengingat di Provinsi DIY khususnya dikenal adanya penyelenggaraan ulangan umum bersama.

Temuan ketiga, mengenai analisis butir menggunakan program Quest, semua peserta memiliki persepsi yang sama, yakni bahwa program tersebut memiliki keunggulan dibanding program yang selama ini mereka kenal. Selama ini mereka mengenal program Anates dan Itemen untuk menganalisis butir, namun keduanya hanya mendasarkan pada teori tes klasik. Sebagian kecil di antara peserta pernah memperoleh pengenalan analisis butir menggunakan pendekatan teori tes modern (IRT) namun umumnya tidak pernah menggunakan lagi sehingga sudah lupa. Dengan adanya pemaparan tentang kelebihan program Quest, maka ada kemudahan melakukan analisis karena sekali analisis menggunakan perpaduan teori tes klasik dan teori tes modern (IRT).

Hasil diseminasi di ketiga provinsi diperoleh temuan, pertama bahwa berpedoman buku Panduan Penulisan Butir Pola Konvergen dan

Divergen yang telah disusun dan diujicobakan pada tahun pertama, para guru peserta tidak mengalami kesulitan dalam menyusun kisi-kisi dan butir pola konvergen. Namun, dalam penyusunan butir pola divergen sebagian guru baik pada diseminasi di Provinsi Daerah Istimewa Yogyakarta, Kalimantan Barat, dan Nusa Tenggara Barat masih mengalami kesulitan, terutama dalam membedakan rubrik dan pedoman penskoran. Para peserta mengusulkan ada penjelasan tentang perbedaan dari keduanya disertai penambahan contoh.

Temuan kedua, sebagian peserta mengalami kesulitan saat melakukan pembuatan data simulasi sebagai contoh hasil ulangan dalam bentuk *excel* untuk diubah ke dalam bentuk *notepad* karena peserta tidak pernah/jarang menggunakannya. Adapun temuan ketiga yakni sebagian peserta mengalami kesulitan membaca hasil menggunakan menu *notepad* dikarenakan *setting* program di dalam komputer ada yang belum menyediakannya, ditambah dengan faktor keterampilan peserta menggunakan komputer menjadi kendala di luar persiapan peneliti.

Berdasarkan tindak lanjut kegiatan diseminasi, yakni penawaran kepada para guru untuk mencoba menyusun kisi-kisi dan butir serta menganalisisnya berdasarkan data riil di lapangan diperoleh kenyataan bahwa pertama tidak semua peserta memanfaatkan kesempatan yang ada. Sebagian besar guru peserta yang mau mencoba melaksanakannya, masih ada sebagian guru peserta yang mengalami kebingungan dalam menyiapkan file kendali (*control file*) untuk eksekusi. Namun, setelah dilakukan korespondensi melalui email dapat diselesaikan dengan baik.

Hasil diseminasi tahun ketiga di LPMP tujuh provinsi diperoleh temuan, pertama bahwa kedua buku panduan yang telah disusun dan diujicobakan pada tahun pertama dan telah didiseminasikan pada tahun kedua di tiga provinsi dapat dipahami dengan baik oleh para widyaiswara setelah memperoleh tambahan penjelasan. Temuan berikutnya, sebagian widyaiswara mengalami kesulitan saat melakukan pembuatan data simulasi sebagai contoh hasil ulangan dalam bentuk *excel* untuk diubah ke dalam bentuk *notepad* karena peserta tidak pernah/jarang menggunakannya. Temuan yang lain, sebagian peserta mengalami kesulitan membaca hasil menggunakan menu *notepad* dikarenakan *setting* program di dalam komputer

ada yang belum menyediakannya. Berdasarkan tindak lanjut kegiatan diseminasi, yakni penawaran kepada para widyaiswara untuk menganalisis data melalui tugas rumah dapat diselesaikan dengan baik setelah dilakukan korespondensi melalui email.

Sebagaimana dikemukakan dalam rumusan tujuan bahwa tujuan penelitian ini adalah mengembangkan panduan penilaian hasil belajar mata pelajaran Biologi dan Bahasa Indonesia di SMA sehingga dapat dijadikan pegangan bagi para guru di lapangan untuk mengembangkan penilaian yang terstandarkan yang butirnya dianalisis menurut teori respons butir, terutama bagi guru Biologi dan guru Bahasa Indonesia SMA di seluruh Indonesia. Dalam hal ini, yang dimaksud dengan pengembangan hasil penilaian yang terstandarkan adalah bahwa hasil penilaian diperoleh dengan prosedur yang dapat dipertanggungjawabkan dibuktikan dengan adanya penyusunan *learning continuum* sebagai *abstract continuum* pengukuran yang dijadikan acuan guru menyusun kisi-kisi, dilanjutkan dengan me-nyusun butir sesuai kisi-kisi. Butir yang telah disusun ditelaah untuk me-ngetahui kualitasnya melalui analisis kualitatif, dilanjutkan dengan peng-ujian dan dianalisis secara kuantitatif. Dengan demikian, skor siswa yang diperoleh benar-benar memenuhi kaidah pengukuran.

Melihat perangkat tes yang telah berhasil disusun guru dikaitkan dengan hasil analisis secara empiris menggunakan program Quest menunjukkan bahwa semua guru sudah berhasil memiliki pengalaman melakukan penilaian yang terstandarkan. Kegiatan tersebut diwujudkan dalam kemampuannya menyusun kisi-kisi. Selanjutnya, butir yang disusun telah di-*review* silang sesama guru sebidang sebagai persyaratan analisis secara kualitatif. Kemudian setelah mengujicobakan di lapangan, guru juga sudah mampu menganalisis menggunakan program Quest—sebagai bukti bahwa butir tes memenuhi persyaratan *fit* dengan model yang digunakan, yakni model *Rasch-1PL*. Dengan langkah tersebut maka akuntabilitas pengukuran telah dapat terpenuhi sebagaimana yang diharapkan. Pengalaman ini tentu akan sangat membantu guru jika mereka ditugasi untuk membantu satuan pendidikan melakukan analisis untuk memenuhi bukti empiris di lapangan. Di masa mendatang hal ini sangat dimungkinkan mengingat sudah banyak satuan pendidikan SMA yang melakukan ujian sekolah terhadap mata

pelajaran yang diujikan melalui Ujian Nasional (UN) bahkan dengan soal-soal yang lebih sulit tingkatannya, juga dapat diketahui kualitasnya setelah dikerjakan oleh siswanya.

Hasil yang diperoleh oleh para guru peserta juga sudah dapat menunjukkan secara empiris bahwa panduan yang disusun oleh peneliti dan setelah diaplikasikan di lapangan dengan memperhatikan masukan dari para peserta sudah dapat digunakan sebagai acuan kerja para guru. Hasilnya menunjukkan bahwa mayoritas instrumen tes sudah menunjukkan hasil yang memuaskan. Bahkan, banyak instrumen tes yang semua butirnya sudah memenuhi syarat model 1-PL. Hal ini berarti bahwa guru sudah dapat memahami panduan yang disusun peneliti. Sebagaimana dikemukakan di latar belakang, bahwa sampai saat ini belum tersedia petunjuk analisis menggunakan program Quest menggunakan pengantar berupa bahasa Indonesia, sehingga belum ada panduan praktis yang dapat dipakai sebagai acuan para guru di lapangan. Dengan hasil yang diperoleh melalui penelitian tahap pertama ini, berarti bahwa panduan penggunaan program Quest—yang sampai sekarang belum ada terbitan panduan praktis dalam bahasa Indonesia—sudah berhasil digunakan para guru di lapangan.

Hasil wawancara dengan guru peserta dari SMA Galur Kulonprogo DI Yogyakarta yang butirnya terbanyak ditolak (28%) besar kemungkinan diakibatkan tidak tepatnya waktu pelaksanaan ulangan harian. Menurut penuturan guru, ulangan harian dilakukan tepat setelah ulangan tengah semester. Akibatnya, sebagian siswa mengeluh dan mereka tidak serius dalam mengerjakannya. Namun demikian, melihat hasil secara keseluruhan maka panduan yang disusun dan telah mendapat masukan dari guru peserta sudah dapat dijadikan panduan yang operasional.

Memperhatikan temuan pada tahun kedua, maka permasalahan di lapangan yang berkait dengan penyusunan kisi-kisi berbasis *learning continuum* tidak akan dapat direalisasikan sepanjang otonomisasi yang diharapkan melalui Kurikulum Tingkat Satuan Pendidikan (KTSP) belum disikapi satuan pendidikan sebagai suatu bentuk otonomisasi dalam mengembangkan kurikulum secara operasional sesuai dengan potensi peserta didik sebagaimana dituntut di dalam Permendiknas Nomor 22 tahun 2006. Terlebih dengan adanya kebijakan penyelenggaraan ulangan umum ber-

sama yang tidak disikapi sebagai tindakan mengukur kompetensi minimal, maka yang terjadi para guru akan mengambil sikap menyamakan KTSP dengan sekolah lain.

Dalam hal aplikasi penggunaan program Quest untuk menganalisis data, kesulitan memahami menu beserta pemanfaatannya ketika melakukan konversi merupakan permasalahan yang umum di lapangan. Keadaan ini hanya dapat diatasi jika pemanfaatan program analisis terhadap data untuk memperoleh bukti secara empiris dicanangkan sebagai suatu tuntutan yang harus dipenuhi dalam setiap satuan pendidikan. Dalam hal ini, kebijakan yang mulai dilaksanakan dalam tahun ini bahwa setiap satuan pendidikan wajib melaksanakan ujian sekolah bagi mata pelajaran yang diujikan secara nasional akan menjadi momen penting. Tentunya secara operasional guru peserta diseminasi maupun panduan yang disempurnakan melalui kegiatan diseminasi ini diharapkan dapat dijadikan modal bagi satuan pendidikan untuk memberikan bukti empiris atas kualitas tes yang diujikan melalui ujian sekolah.

Simpulan

Berdasarkan hasil penelitian yang telah diperoleh dapat ditarik beberapa simpulan yakni 1) penelitian ini sudah menghasilkan *learning continuum* mata pelajaran Biologi dan mata pelajaran Bahasa Indonesia SMA yang dapat dijadikan acuan bagi guru untuk mengembangkan indikator yang ditindaklanjuti dengan penyusunan instrumen hasil belajar di SMA pada kedua mata pelajaran tersebut; 2) penelitian ini sudah menghasilkan Panduan Penyusunan Butir Tes Pola Konvergen dan Divergen yang sudah dapat dijadikan pedoman bagi guru dalam menyusun instrumen hasil belajar dengan kedua pola tersebut, dan telah diujicobakan di sekolahnya masing-masing, baik pada tahun pertama maupun pada tahun kedua; 3) penelitian ini sudah menghasilkan “Panduan Analisis Data Pengukuran Pendidikan untuk Memperoleh Bukti Empiris Kesahihan Menggunakan Program Quest” dengan menggunakan bahasa Indonesia sebagai bahasa pengantar dan sudah divalidasi melalui kegiatan diseminasi di tiga provinsi (DIY, Kalbar, dan NTB) sehingga dapat dijadikan pegangan bagi guru

dalam melakukan analisis dan menginterpretasikan hasil analisis dari hasil terutama ulangan akhir semester dan ulangan kenaikan kelasserta ujian sekolahnya masing-masing menggunakan program tersebut; 4) bukti empiris menunjukkan bahwa mulai dari penyusunan *learning continuum*, kisi-kisi dan soal yang disusun guru, analisis hasil uji coba lapangan tahun pertama maupun pada kegiatan diseminasi tahun kedua menunjukkan bahwa guru sudah mampu melakukan penilaian yang terstandarkan.

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; pp. 36-43.
- Atherton, J.S. (2005). *Learning and Teaching: Convergent and Divergent Learning*. Diakses pada tanggal 03-Des-2006: <http://www.learningandteaching.info/learning/converge.htm>.
- Bond, T.G. & Fox, Ch.M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. 2nd ed. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Brennan, R.L. (2001). Some problems, pitfalls, and paradoxes in educational measurement [Versi elektronik]. *Educational Measurement, Issues and Practice*. Washington: Winter 2001. Vol. 20, Iss. 4; pp.6-18.
- BSCS. BSCS is a non-profit organization. © Copyright (2010) BSCS All Rights Reserved (Biology Syllabus <http://www.rickovernaval.org/ourpages/auto/2009/9/3/54143865/Biol>)
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.

- Depdiknas. (2006). *Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 22 Tahun 2006 Tentang Standar Isi untuk Satuan Pendidikan dasar dan Menengah.*
- Depdiknas. (2006). *Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 23 Tahun 2006 tentang Standar Kompetensi Lulusan untuk Satuan Pendidikan dasar dan Menengah.*
- Depdiknas. (2006). *Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 24 Tahun 2006 tentang Pelaksanaan Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 22 Tahun 2006 tentang Standar Isi untuk Satuan Pendidikan dasar dan Menengah dan Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 23 Tahun 2006 tentang Standar Kompetensi Lulusan untuk Satuan Pendidikan dasar dan Menengah.*
- Depdiknas. (2007). *Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 20 Tahun 2007 Tentang Standar Penilaian Pendidikan untuk Satuan Pendidikan Dasar dan Menengah.*
- Edward, J.R. & Bagozzi, R.P. (2000). On the nature and direction of relationship constructs and measurement. *Psychological Methods*. 2000. Vol. 5. No. 2; pp.155-174.
- Gorin, J.S. (2006). Test Design with cognition in mind. *Educational Measurement, Issues and Practice*. Washington: Winter 2006. Vol. 25, Iss. 4; pp.21-35.
- Han, Kyung T. & Hambleton, R.K. (2007). *User's manual for wingen2: Windows software that generates IRT model parameters and item response.* (Media elektronik]. Massachusetts: Center for Educational Assessment.
- Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal*, Spring 2002, Vol.39, No. 1, pp.69-95.

- IEA. (2003). *TIMSS 2003 technical report. Finding from IEA's trends in international mathematics and science study at the fourth and English grades*. Boston: TIMSS & PIRLS International Study Center.
- Jehlen, A. (2007). Testing how the sausage is made. *NEA Today*. Washington: Apr 2007. Vol. 25, Iss. 7; pp.29-34.
- Keeves & Alagumalai. (1999). New Approach to measurement. Dalam: Masters, G.N. & Keeves, J.P. (eds.). *Advances in Measurement in Educational Research and Assessment* (pp.23-42). Amsterdam: Pergamon, An imprint of Elsevier Science.
- Keeves, J.P. & Masters, G.N. (1999). Introduction Dalam: Masters, G.N. & Keeves, J.P. (eds.). *Advances in measurement in educational research and assessment*(pp.1-22). Amsterdam: Pergamon, An imprint of Elsevier Science.
- Muraki, E. & Bock, R.D. (1998). *Parscale: IRT item analysis and test scoring for rating scale data*. Chicago: Scientific Software International, Inc.
- Roid, G.H. & Haladyna, Th.M. (1982). *A technology for test-item writing*. Orlando: Academic Press, Inc.
- Subali, Bambang. (2009). *Pengukuran keterampilan proses sains pola divergen dalam mata pelajaran biologi SMA di provinsi DIY dan Jawa Tengah. Disertasi tidak diterbitkan*. Yogyakarta: Program Pascasarjana UNY
- Thissen, D., Nelson, L, & Surygert, K.A. (2001). Item response theory applied to combination of multiple-choice and constructed response items—Approximation methods for scale score. Dalam: Thissen, D. & Wainer, H. (2001). *Test Scoring* (pp.293-324). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Wright & Masters, G.N. (1982). *Rating scale analysis*. Chicago: Mesa Press.