

## STANDARISASI INSTRUMEN *INTEGRATED ASSESSMENT* HASIL BELAJAR BAHASA DENGAN PROGRAM QUEST

Pujiati Suyata, Nur Hidayanto, dan Agus Widyantoro  
Pascasarjana UNY dan UAD  
email: pujiati\_suyata@uny.ac.id

### Abstrak

Penelitian ini bertujuan menghasilkan model instrumen *integrated assessment* bentuk konvergen dan divergen hasil belajar bahasa Indonesia dan Inggris SMP yang terstandar dengan menggunakan *Item Response Theory*. Untuk tujuan tersebut dilakukan penelitian dengan desain penelitian dan pengembangan. Pada tahun I dikembangkan *learning continuum*, diikuti dengan pengembangan instrumen *integrated assessment* berdasarkan *learning continuum* tersebut. Selanjutnya dilakukan standarisasi instrumen menggunakan *Item Response Theory* 1-PL dan program *QUEST*. Hasil tahun I adalah 30 perangkat instrumen *integrated assessment* yang terstandar, buku panduan penyusunan instrumen *integrated assessment* hasil belajar bahasa dan panduan program *QUEST*. Pada tahun II, instrumen *integrated assessment* hasil belajar bahasa didiseminasikan ke Provinsi DIY, Kalsel, dan NTB.

**Kata kunci:** standarisasi *integrated assessment*, *Item Response Theory*, hasil belajar bahasa

### THE STANDARDIZATION OF INTEGRATED ASSESSMENT INSTRUMENTS FOR LANGUAGE LEARNING OUTCOMES USING THE QUEST PROGRAM

### Abstract

This study aims to produce a model of standardized convergent and divergent integrated-assessment instruments for Indonesian and English learning outcomes at junior high schools using Item Response Theory. It employed a research and development design. In the first year, a learning continuum was developed, followed by the development of integrated assessment instruments based on the learning continuum. Instrument standardization was then conducted using 1-PL Item Response Theory and the QUEST program. The results of the first year were 30 standardized integrated assessment instrument sets, a guidebook for construction of the integrated assessment instruments for language learning outcomes and the QUEST program guidebook. In the second year, the integrated assessment instruments for language learning outcomes were disseminated in provinces of Yogyakarta Special Territory, South Kalimantan, and West Nusa Tenggara.

**Keywords:** integrated assessment standardization, Item Response Theory, language learning outcomes

### PENDAHULUAN

Hasil belajar siswa, seperti Ujian Nasional misalnya, menunjukkan nilai mata pelajaran bahasa, Indonesia dan Inggris, tidak menggembirakan. Beberapa hal bisa diidentifikasi sebagai penyebab, salah satu

di antaranya sistem penilaian yang digunakan guruselama ini. Kegiatan penilaian melibatkan aktivitas pengukuran dan keberadaan instrumen. Salah satu jenis instrumen dalam penilaian bahasa adalah tes bentuk konvergen dan divergen.

Indonesia merupakan salah satu negara peserta program asesmen internasional. Data PISA menunjukkan bahwa Indonesia dari tahun ke tahun hasilnya belum seperti yang dihaarpkan. PISA tahun 2003, misalnya, kompetensi pemahaman membaca siswa Indonesia berada pada urutan 39 dari 41 negara peserta. Kondisi yang hampir sama juga terjadi pada PISA tahun 2007. Hasil penelitian Puspendik melaporkan bahwa salah satu penyebab rendahnya kompetensi pemahaman membaca siswa Indonesia adalah kurang terbiasanya mereka mengerjakan soal-soal kognitif tingkat tinggi (*Higher Ordering Test* atau *HOT*). Guru terbiasa memberikan soal-soal kognitif tingkat rendah, seperti pengetahuan (*to know*) dan pemahaman (*to comprehend*), sementara aplikasi (*to apply*), analisis (*to analysis*), evaluasi (*to evaluate*), dan kreasi (*to create*), yang merupakan kognitif tinggi (Pariñas, 2009: 14-14), jarang atau bahkan tidak pernah dilakukan. Karakteristik soal-soal *HOT*, seperti: 1) melibatkan berpikir tingkat tinggi, 2) permasalahan kompleks, dan 3) melibatkan lebih dari satu jenis kognitif tidak tampak dalam soal-soal buatan guru.

Secara umum tujuan pembelajaran bahasa Indonesia adalah tercapainya kompetensi berkomunikasi lewat keempat kompetensi berbahasa. Oleh karena itu, penilaian yang dilakukan juga harus mengukur kompetensi berbahasa tersebut. Dalam hal ini, bentuk penilaian yang terbaik dan disarankan adalah penilaian terpadu (*integrated assessment*) yang menekankan pada penilaian kompetensi berunjuk kerja bahasa seperti dalam berkomunikasi sehari-hari. Melalui penelitian ini, diungkap cara pengembangan *integrated assessment* tipe konvergen dan divergen hasil belajar bahasa, beserta teknik pembakuan dan penafsirannya. Diharapkan, kualitas guru akan meningkat secara nasional, khususnya untuk mata pelajaran Bahasa Indonesia dan Inggris.

Kurikulum berbasis kompetensi menekankan perlunya penilaian pembelajaran bahasa diwarnai oleh unjuk kerja berbahasa. Hal itu terjadi kana dalam penilaian bahasa harus memperhatikan hakikat dan fungsi bahasa. Pada hakikatnya, bahasa merupakan hasil budaya manusia dan berfungsi sebagai sarana komunikasi. Pendekatan penilaian yang sesuai adalah yang menekankan pada aspek kinerja berbahasa. Dengan demikian, penilaian pembelajaran bahasa tidak mengarah pada sistem bahasa, tetapi pada penggunaan bahasa secara benar sesuai dengan sistem itu. Dengan demikian, penilaian pembelajaran bahasa perlu menekankan pada *fungsi bahasa sebagai alat komunikasi*.

Karakteristik penilaian pembelajaran bahasa kedua adalah *otentik* (Brown, 2004: 28). Komunikasi berjalan secara apa adanya. Dengan demikian, penilaian pembelajaran bahasa sebagai alat komunikasi juga berjalan sewajarnya, sesuai dengan keadaan senyatanya. Jadi, data yang diperlukan adalah data otentik penggunaan bahasa.

Komunikasi nyata selalu melibatkan lebih dari satu kemampuan berbahasa, seperti komunikasi lewat HP akan melibatkan kompetensi mendengarkan dan berbicara. Karena itulah, ciri khas penilaian kompetensi berbahasa adalah terpadu, *integrated assesemnt*, antara kompetensi berbahasa yang satu dengan yang lain (Pappas, dkk, 1996:8).

Penilaian *integrated assessment* merupakan salah satu karakteristik penilaian dalam pembelajaran bahasa. Hal itu mengisyaratkan perlunya penilaian dalam pembelajaran bahasa diarahkan pada peristiwa-peristiwa berbahasa yang terjadi dalam situasi nyata. Dalam situasi nyata tersebut, terkait berbagai kompetensi berbahasa secara terintegrasi, untuk mendukung kelancaran komunikasi. Heaton (1988:15) menyebutnya *integrative tets*. Penilaian terintegrasi tersebut tampak jelas dalam *performance-based assessment*

(Brown, 2004: 28), seperti dalam diskusi akan terjadi kombinasi antara menyimak dan berbicara atau integrasi antara membaca dan menulis dalam meringkas suatu bacaan. Dengan demikian, dalam pembelajaran di sekolah, ditekankan kemampuan peserta didik menunjukkan kompetensi yang dimiliki secara nyata dan bermakna, berlandaskan berbagai kompetensi yang terintegrasi tersebut.

Jadi, penilaian model ini menekankan pada pengukuran kinerja, *doing something*, yang merupakan penerapan ilmu pengetahuan yang telah dikuasai secara teoretis. Fungsi bahasa sebagai alat komunikasi menjadi perhatian utama. Brown (2004: 10) menyebutnya *communicative language testing*. Meskipun demikian, tidak berarti penilaian pembelajaran mengabaikan sistem bahasa, sebab terganggunya penguasaan sistem tersebut dapat mengganggu praktik berbahasa dalam situasi nyata. Selain itu, penilaian pembelajaran tidak mengabaikan konteks. Strategi pemahaman konteks, yang disebut kompetensi pragmatik sangat berperan dalam berkomunikasi.

Salah satu kompetensi guru adalah pengembangan instrumen penilaian. Ada bermacam-macam instrumen penilaian, yang dibedakan berdasar fungsi, konten/isi, bentuk item, teknik pengujian, sistem penskoran, dan interpretasi terhadap hasil. Berdasarkan cara menginterpretasikan hasil, ada tes acuan patokan (CRT) yang berbeda dengan tes acuan norma (NRT). CRT digunakan untuk mengidentifikasi status individu berkenaan dengan standar pencapaian yang telah ditetapkan. Dalam NRT skor individu ditafsirkan berkenaan dengan distribusi penampilan kelompok yang diukur dengan instrumen yang sama.

Dalam pengembangan instrumen penilaian, ada beberapa tahap yang harus dilalui, yakni: (1) merancang, (2) mengujicoba, (3) menetapkan validitas, (4) menetapkan reliabilitas, dan (5) menginterpre-

tasi hasil. Kegiatan perancangan meliputi: (1) menetapkan tujuan, (2) menyiapkan tabel spesifikasi, (3) menulis item, dan (4) mengedit item. Kegiatan uji coba instrumen penilaian meliputi kegiatan: (1) analisis uji coba, dan (2) pemaknaan dan interpretasi. Tantangan terberat dalam mengembangkan instrumen penilaian justru pada lemahnya pemahaman tentang struktur dari substansi pengetahuan yang akan diukur.

Salah satu kualitas instrumen yang perlu dipenuhi oleh setiap pengembang instrumen adalah validitas. Salah satu di antaranya adalah validitas konstruk. Menurut Gorin (2001: 343-368), validitas konstruk merupakan sentral dalam penetapan mutu instrumen penilaian. Untuk instrumen bentuk tes, definisi konstruk yang komprehensif akan memfokuskan pengembangan item dan tes pada *ability* (kemampuan) dan *trait* (diri spesifik) yang akan diukur. Persoalan yang dihadapi dalam pemenuhan validitas konstruk dalam ranah kognitif bukan hanya terbatas pada kesesuaian item dan indikator pencapaian kompetensi, tetapi juga pada keberadaan kompetensi yang diukur, apakah kompetensi tersebut berada dalam satu dimensi ataukah tidak. Ternyata hal itu tidak mudah untuk dilakukan. Oleh karena itu, pemenuhan validitas konstruk dalam ranah kognitif dibatasi dalam pemenuhan validitas isi yang tergambar dalam *learning continuum*.

Reliabilitas diestimasi dengan koefisien yang menggambarkan derajat relatif kesalahan pengukuran dalam skor tes. Sementara itu, kesalahan pengukuran merupakan 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 & Haldyna, 1982: 43-47).

Perhitungan reliabilitas yang dijelaskan dalam teori tes klasik terkait dengan pengembangan tes menurut teori tes

klasik, yang mengacu pada NRT berbeda bila tes digunakan untuk CRT. Ada 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.

Instrumen penilaian kompetensi berpikir konvergen dapat diuji dengan menggunakan instrumen bentuk pilihan (*alternatif test*), sedangkan kemampuan berpikir divergen hanya dapat diukur dengan item tes bentuk uraian bebas. Menurut Roid & Haladyna (1982: 58-62) jawaban yang diharapkan dalam tes uraian terbuka harus luas dan komprehensif. Tes semacam ini kreatif karena menggabungkan beberapa jawaban. Dengan demikian, tes bentuk divergen sesuai dengan pengetahuan komprehensif yang harus diuji, dan pengetahuan kognitif yang dimaksudkan adalah umum dan luas.

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

Tantangan yang dihadapi dalam setiap pengukuran adalah terkait dengan panjang instrumen yang akan diujikan, banyaknya kriteria yang digunakan untuk menskala respons yang diberikan oleh siswa, dan ukuran sampel. Dalam pembakuan item, ukuran sampel ikut menentukan tingkat kestabilan yang dicapai. Menurut Han & Hambleton (2007:15-20) dan Theissen et al. (2001: 295-325), dalam model-model respons butir dikotomus, hanya ada benar dan salah (yaitu, 0 atau 1). Namun demikian, dalam beberapa situasi, dapat terjadi respons-respons lebih dari dua kategori. Sebagai contoh, suatu tes uraian terbuka dengan menggunakan

butir skala Likert, mungkin menghasilkan respons 5 kategori (sangat tidak baik, tidak baik, baik, dan sangat baik, yang dapat diberi kode dari 0 hingga 4).

Dalam teori pengukuran modern, ada model analisis butir soal dengan paradigma teori respons butir (*Item Response Theory*)/IRT), ada program satu parameter (1-PL) atau dikenal dengan Model Rasch. Program ini hanya memberikan layanan analisis data dengan skala dikotomus. Perluasan program Rasch adalah program Quest yang memberikan layanan untuk menganalisis data hasil pengujian yang menggunakan skala dikotomus, skala politomus, dan kombinasi keduanya. Program Quest juga ada yang tidak *under window* dengan langkah-langkah yang praktis sehingga dapat dilatihkan kepada guru yang sudah mengenal komputer. Program ini juga tidak menuntut banyaknya replikasi yang besar (Adam & Kho, 1996:89). Dengan demikian, guru dapat menganalisis hasil pengujian dalam konteks pengembangan instrumen penilaian hasil belajar yang terstandarkan menggunakan prinsip teori repons butir.

## METODE

Penelitian ini menerapkan model *Research and Development* dengan target akhir terdesiminasikannya hasil riset pada sejumlah provinsi. Pada tahun I, dilakukan: a) Pengembangan *learning continuum* bahasa Indonesia dan Inggris jenjang SMP, dengan mengacu pada Permendiknas nomor 22 tahun 2006 tentang Standar Isi. Selanjutnya dilakukan *focus group discussion* (FGD) dengan melibatkan pakar bidang studi dan pengukuran pendidikan. Hasil FGD diseminarkan dengan mengundang sejumlah guru bahasa Indonesia dan Inggris SMP, b) Penulisan instrumen sesuai *learning continuum*, c) Pembakuan item menggunakan program Quest dan d) berdasarkan teori dan hasil empiris pembakuan item, disusun panduan penulisan instrumen penilaian hasil belajar bahasa

dan panduan analisis data menggunakan program Quest.

Pada tahun II, kegiatan difokuskan pada tahap diseminasi instrumen berdasarkan buku panduan, yang diawali dengan kegiatan sosialisasi, kemudian dilanjutkan dengan pelaksanaan diseminasi dengan melatih guru yang tertunjuk di setiap provinsi yang dijadikan demplot. Kegiatan ini dimonitor dan dievaluasi keberhasilannya untuk selanjutnya dilaporkan kepada pihak yang terkait agar dapat disebarluaskan secara nasional.

Penelitian tahun I adalah kegiatan persiapan, dimulai dengan tahap pengkajian dan penyusunan *learning continuum* instrumen pengukur hasil belajar di SMP dilanjutkan penulisan instrumen pengukur hasil belajar untuk mata pelajaran yang bersangkutan, dilaksanakan di Fakultas Bahasa dan Seni, UNY. Adapun pelaksanaan uji coba dilakukan di SMP di Provinsi DIY. Lokasi analisis data dilaksanakan di Program Pascasarjana UNY. Penelitian tahun II, berupa kegiatan diseminasi hasil penelitian di tiga provinsi demplot yaitu di Provinsi DIY, Kalimantan Selatan, dan Nusa Tenggara Barat.

Pengumpulan data pada tahun I dilakukan dengan: 1) observasi partisipatif dan interview untuk pengumpulan data terkait SK dan KD/*learning continuum* mata pelajaran yang bersangkutan, 2) *forum group discussion (FGD)* untuk memperoleh SK dan KD/*learning continuum* mata pelajaran yang bersangkutan, 3) pengujian dilakukan selama uji coba instrumen pengukur hasil belajar untuk mata pelajaran bahasa Indonesia dan Inggris.

Pengumpulan data pada tahun II dilaksanakan melalui kegiatan 1) diseminasi instrumen pengukur hasil belajar untuk mata pelajaran bahasa Indonesia dan Inggris di SMP tempat guru mengajar, dan 2) observasi partisipatif dilakukan untuk monitoring dan evaluasi pelaksanaan program diseminasi.

Analisis data menggunakan pendekatan kualitatif dan kuantitatif. Analisis data secara kualitatif dalam bentuk analisis deskriptif digunakan untuk menganalisis hasil pengkajian SK dan KD dari silabus KTSP yang sudah ada, hasil perumusan *learning continuum* mata pelajaran, juga hasil review dan revisi instrumen pengukur hasil belajar. Analisis data secara kuantitatif menggunakan pendekatan *IRT* dikhotomis dan politomis dengan paket program QUEST (Adams & Kho, 1996:89).

## HASIL DAN PEMBAHASAN

Hasil tahun I tentang standarisasi *integrated assessment* hasil belajar bahasa model konvergen dan divergen sebagai berikut.

### Studi Pendahuluan

Hasil wawancara dengan para guru mengisyaratkan lemahnya kompetensi mereka dalam melakukan penilaian, khususnya penilaian yang terintegrasi antara satu aspek dengan aspek lain. Kelemahan berikutnya adalah kurang jelasnya pemahaman mereka terhadap pengembangan soal bentuk konvergen dan divergen. Padahal, penilaian semacam itu menjadi tuntutan karena mereka adalah guru dan fungsi bahasa sebagai alat komunikasi.

Dari wawancara juga ditemukan kurangnya pengetahuan guru dalam hal pengembangan instrumen yang terstandarisasi, khususnya terkait dengan analisis data dengan program pengukuran modern seperti Quest yang menggunakan *Item Respons Theory* satu parameter. Mereka bersemangat dan sangat antusias ingin mengetahui dan mengaplikannya dalam pengembangan instrumen.

### Pengembangan Instrumen

Telah dikembangkan 30 draft produk instrumen *integrated assessment*, 9 draft instrument bahasa Indonesia dan 21 in-

strumen bahasa Inggris. Pengembangan melalui serangkaian diskusi mengenai konsep dasar pengujian pembelajaran bahasa yang menuntut kompetensi siswa dalam berbahasa dan berbagai faktor terkait pengembangan instrumennya.

Hasil kegiatan tersebut 1) Pengembangan *Learning Continuum*, 2) Telaah *Learning Continuum*, 3) Seminar *Learning Continuum*, 4) Penyusunan Panduan *Integrated Item* Tes Hasil Belajar Bahasa Pola Konvergen dan Divergen dan Panduan Analisis Item Menggunakan Program Quest: Skala Dikotomus, Politomus, dan Kombinasinya, 5) Pelatihan Penulisan *Integrated Item* yang diikuti oleh guru Bahasa Indonesia SMP dan guru Bahasa Inggris SMP di Provinsi DIY, 6) Validasi Produk baik secara logis maupun empirik. Validasi logis dicapai melalui kegiatan *Focus Group Discussion (FGD)* dan telaah silang dengan melibatkan para guru mata pelajaran Bahasa Indonesia dan Inggris SMP. Sementara itu, validitas empiris

dicapai melalui uji coba di lapangan dan menganalisisnya dengan menggunakan program analisis *Quest*. Secara ringkas berdasarkan hasil uji coba di lapangan diperoleh hasil seperti tampak pada Tabel 1.

Tabel 1 menunjukkan bahwa soal-soal yang disusun oleh guru bahasa Indonesia dalam bentuk konvergen dan divergen, dengan item yang berjumlah antara 20-30 item. Soal mempunyai nilai reliabilitas yang berada pada rentang 0,78 – 0,91 dan hanya satu set soal yang reliabilitasnya 0,40. Semakin tinggi nilai reliabilitas semakin meyakinkan bahwa sampel uji coba sesuai atau *fit* dengan item yang diujikan. Semakin rendah, semakin banyak sampel uji coba yang tidak memberikan informasi yang diharapkan (tidak mengerjakan atau mengerjakan asal-asalan). Di samping reliabilitas berdasarkan item, reliabilitas juga dapat ditunjukkan berdasarkan *case* atau testi. Semakin tinggi nilainya, semakin meyakinkan bahwa pengukuran memberikan hasil yang konsisten.

Tabel 1. Kualitas Instrumen Bahasa Indonesia Berdasarkan Analisis *IRT* Menggunakan Program *QUEST*

No.	Nama Guru dan Sekolah	Bentuk Item	Jumlah Item	Reliability of Item Estimates	Infit MNSQ	Item Ditolak
1.	Sukinah SMP 12 YK	PG + Uraian	30	R = .83	Mean .10SD 1.16	3 (10%)
2.	Wahyudi SMP 1 GK	PG + Uraian	30	R = .89	Mean.99SD .22	0 ( 0%)
3.	Nuraini SMP 1 KP	PG + Uraian	30	R = .78	Mean 1.00SD .14	0 (0%_)
4.	Karyadi SMP 1 GK	PG + uraian	25	R = .80	Mean 1.00SD .09	0 (0%)
5.	Suprpto SMP 2 Wnsr	PG + uraian	20	R = .82	Mean 1.00SD .11	1 (5%)
6.	Nurudin SMP 1 Panjatan	PG + uraian	25	R = .80	Mean 1.00SD .14	1 (5%)
7.	Supriatmaja SMP Pandak	PG + uraian	20	R = .89	Mean 1.00SD .22	2 ( 10%)
8.	Rusmantara MTs Pundong	PG + uraian	25	R = .40*	Mean 1.00SD .11	0 (0%)
9.	Suprihatin SMP 3 Jetis Bantul	PG + uraian	20	R = .91	Mean 1.00SD .08	0(0%)

Tabel 2. Kualitas Instrumen Bahasa Inggris Berdasarkan Analisis IRT Menggunakan Program QUEST

No.	Nama Guru dan Sekolah	Bentuk Item	Jumlah Item	Reliabilitas	Infit MNSQ	Item Ditolak
1.	Asri Desire SMP 2 Bantul	PG + Uraian	30	R = .88	Mean 1.00SD .14	1 (3%)
2.	Erni SMP 1 Bantul	PG + Uraian	25	R = .80	Mean 1.00SD .18	3 ( 12%)
3.	Sunaryono SMP 3 KP	PG + Uraian	30	R = .88	Mean 1.00SD .04	0 (0%_)
4.	Laily F SMP GK	PG +uraian	30	R = .85	Mean 1.00SD .07	0 (0%)
5.	Nurwakhid SMP GK	PG +uraian	30	R = .91	Mean 1.00SD .14	2 (6%)
6.	Sudayanto SMP Sleman	PG +uraian	30	R = .84	Mean 1.00SD .07	3 (10%)
7.	Winarni SMP GK	PG +uraian	30	R = .94	Mean .95SD .57	6 (20%)
8.	Cahyo SMP GK	PG +uraian	20	R = .85	Mean .94SD .20	2 (10%)
9.	Dwi Wahyu SMP GK	PG +uraian	20	R = .84	Mean .99SD .15	2 (10%)
10.	Tri SMP GK	PG +uraian	20	R = .81	Mean 1.00SD .12	1 (5%)
11.	Fada N. SMP GK	PG +uraian	20	R = .79	Mean 1.00SD .10	0 (0%)
12.	Harini SMP GK	PG +uraian	20	R = .85	Mean .99SD .20	2 (10%)
13.	Tri Puji SMP GK	PG +uraian	20	R = .93	Mean 1.00 SD .10	0 (0%)
14.	Rien Endah SMP GK	PG +uraian	20	R = .72	Mean 1.00SD .08	1 (5%)
15.	Tri Sujarwati SMP GK	PG +uraian	20	R = .80	Mean 1.00SD .08	0 (0%)
16.	Sri Nanik SMP GK	PG +uraian	20	R = .80	Mean 1.00SD .20	3 (15%)
17.	Umi A. SMP GK	PG +uraian	20	R = .91	Mean .98SD .26	3 (15%)
18.	Yayuk SMP GK	PG +uraian	20	R = .89	Mean .99SD .19	2 (10%)
19.	Tumijo SMP GK	PG +uraian	20	R = .88	Mean 1.01 SD.19	3 (15%)
20.	Dwi K SMP GK	PG +uraian	20	R = .77	Mean 1.01SD .19	2 (10%)
21.	Sumartini SMP GK	PG +uraian	20	R = .73	Mean .99SD .24	5 (25%)*

Dari Tabel 1 juga dapat diamati bahwa dari nilai mean INFIT MNSQ dan SD secara keseluruhan, item sesuai dengan model Rach atau model 1- PL, berada pada rentang yang diperbolehkan, yaitu

antara  $\geq 0,77 - \leq 1,30$ . Dilihat dari setiap item, setiap set soal, ada 5 set diterima semula (penolakan 0%), 2 set ada penolakan 5%, dan 2 set penolakan 10%.

Tabel 2 menunjukkan bahwa soal-soal yang disusun oleh guru bahasa Inggris dalam bentuk konvergen dan divergen, dengan item yang berjumlah antara 20-30. Soal mempunyai nilai reliabilitas cukup baik, yang berada pada rentang 0,72 – 0,94. Semakin tinggi nilai reliabilitas semakin meyakinkan bahwa sampel uji coba *fit* dengan item yang diujikan. Di samping reliabilitas berdasarkan item, reliabilitas juga ditunjukkan berdasarkan *case* atau testi. Semakin tinggi nilainya, semakin meyakinkan bahwa pengukuran reliabel dan memberikan hasil yang konsisten.

Dari Tabel 2 juga dapat diamati bahwa dari nilai mean INFIT MNSQ dan SD secara keseluruhan, item sesuai dengan model Rach atau model 1- PL, berada pada rentang yang diperbolehkan, yaitu antara  $\geq 0,77$  –  $\leq 1,30$ . Dilihat dari setiap set soal, ada 5 set diterima semua (penolakan 0%) atau semua itemnya memenuhi persyaratan item tes model 1-PL, 1 set ada penolakan 3%, dan 3 set penolakan 5-6%, 7 set penolakan 10-12%, 3 set penolakan 15%, 2 set penolakan 20%, dan 1 set penolakan 25%.

Kegiatan penelitian Tahun II adalah sosialisasi dan desiminasi hasil penelitian Tahun I di tiga provinsi, yakni Provinsi Kalimantan Selatan, Nusa Tenggara Barat, dan DIY. Dalam perkembangannya, ada guru dari provinsi lain yang ingin mengikuti kegiatan ini, yaitu Provinsi Jateng, Kalbar, Kalteng, Riau, dan Lampung. Dari setiap provinsi diundang guru-guru SMP Mapel Bahasa Indonesia dan Inggris lewat MGMP. Dari Provinsi Kalimantan Selatan diundang 18 guru Bahasa Indonesia dan 18 guru Bahasa Inggris. Dari Provinsi NTB, peserta Mapel Bahasa Inggris ada 17 dan Bahasa Indonesia 17 pula. Sementara itu, dari Provinsi DIY/Jateng, peserta Mapel Bahasa Indonesia 19 dan Bahasa Inggris 19.

Kegiatan sosialisasi dan desiminasi diawali dengan tugas membaca buku 1) panduan penulisan item *integrated assess-*

*ment* dan 2) panduan analisis program *Quest*. Peserta diminta membaca panduan dengan cermat dalam waktu 1 jam untuk setiap panduan. Langkah berikutnya adalah presentasi dan penjelasan buku panduan oleh tim peneliti diikuti tanya jawab dan diskusi. Selanjutnya kegiatan praktik analisis item menggunakan program *Quest* dengan data simulasi yang telah disediakan. Kegiatan berikutnya adalah pemaknaan hasil analisis dengan program *Quest*. Pada akhir kegiatan, para peserta diminta mengisi angket terkait pemahaman buku panduan.

Setelah membaca buku panduan penulisan item *integrated assessment*, para peserta dari Provinsi Kalimantan Selatan pada umumnya dapat memahami semua yang mereka baca. Hal tersebut terlihat dari jawaban angket yang menyatakan bahwa pemahaman item-item dalam angket, yang memahami seluruhnya ada (90 – 100%), termasuk kategori baik sekali. Ada satu item yang belum seluruhnya memahami, yaitu tentang evaluasi otentik, yang memahami seluruhnya baru mencapai 82%.

Hasil sosialisasi dari NTB, setelah membaca buku panduan pemahaman guru-guru SMP, Bahasa Indonesia dan Inggris, Nusa Tenggara Barat terhadap item-item yang terkait dengan evaluasi pembelajaran yang ditanyakan dalam angket, yang memahami seluruhnya berada pada rentang antara 80 – 92%, suatu pemahaman yang berada pada kategori baik. Ada sebagian kecil di antara mereka yang belum seluruhnya memahami atau belum paham, namun teman-teman mereka yang sudah paham, yang tergabung dalam MGMP, dapat melatih mereka.

Hasil analisis angket pemahaman buku panduan dari Provinsi DIY dan Jateng, peserta sosialisasi dan desiminasi dari provinsi tersebut adalah guru-guru bahasa Indonesia dan Inggris SMP yang sedang menempuh studi S-2 di Pascasarjana UNY. Peserta tidak hanya dari Provinsi



DIY dan Jateng, tetapi juga dari provinsi lain. Dari Kalimantan Barat ada 5 orang, Lampung 3 orang, Kalteng 2 orang, dan Riau 1 orang. Dengan demikian, sosialisasi dan desiminasi ini tidak hanya pada 4 provinsi, melainkan lebih dari itu, yaitu 8 provinsi..

Hasil sosialisasi tampak bahwa pemahaman mereka setelah membaca buku panduan penulisan soal bahasa yang terintegrasi adalah antara 90%-100%, termasuk kategori baik sekali. Hanya sedikit di antara mereka yang belum sepenuhnya memahami, namun hal tersebut dapat ditutup dengan bertanya pada teman-teman yang sudah memahami seluruhnya. Mereka berada dalam satu kelas sehingga bisa belajar dari sesama teman. Dalam kegiatan tersebut mereka memperoleh buku panduan penulisan item *integrated assessment*, dalam kegiatan sosialisasi dan demesiminasi ini, para guru juga memperoleh buku panduan analisis data menggunakan program analisis modern, yaitu program Quest.

Hasil analisis terhadap pemahaan buku panduan analisis Quest sebagai berikut. Pertama, dari Provinsi Kalimantan Selatan dapat diketahui bahwa pemahaman guru bahasa SMP Provinsi Kalimantan Selatan tentang item-item buku panduan analisis menggunakan program Quest pada umumnya baik, ( 80 %- 90%), dan satu item yang pemhamannya sangat baik (95%), yaitu tentang *CTT (Classical Test Theory)*. Dari beberapa pelatihan terdahulu, mereka cukup memahami analisis menggu akan tes klasik. Setelah membaca buku panduan analisis menggunakan program Quest, mereka lebih terbuka pemikirannya bahwa program klasik ternyata ada kelemahannya dan program Quest yang menggunakan teori modern lebih baik karena dapat menutup kelemahan tes klasik.

Kedua, hasil analisis pemahaman panduan analisis data Program Quest Provinsi NTB adalah para guru sudah

memahami analisis menggunakan program klasik yang telah mereka kenal sebelumnya. Dengan buku panduan tersebut, pemahaman yang semula masih setengah-setengah menjadi lebih baik. Sebanyak 85% mengaku telah memahami program *CTT*. Selain itu, pemahaman mereka terhadap analisis menggunakan program modern, Quest, menjadi terbuka. Dengan membaca buku panduan tersebut mereka mengatakan cukup memahami (70% - 80%).

Ketiga, hasil analisis pemahaman para guru SMP di Provinsi DIY dan Jateng setelah mereka membaca buku panduan analisis data menggunakan program Quest adalah pemahaman guru DIY dan Jateng berada pada kisaran antara 90% - 100%, suatu pemahaman yang sangat baik. Pada latihan aplikasi program Quest dengan data simulasi, dengan dipandu peneliti, mereka dengan antusias mengikutinya. Hasilnya semua peserta dapat melakukannya dengan baik.

Selain praktik dengan data simulasi, guru-guru di Provinsi DIY dan Jateng juga melakukan praktik analisis data menggunakan program Quest dengan data sesungguhnya. Data diperoleh dengan melakukan penyusunan soal sesuai rambu-rambu yang disarankan dalam buku panduan penulisan soal *interated assessment*, dimulai dari mempelajari *learning continuum*, menyusun kisi-kisi, telaah kisi-kisi, uji coba soal, dan analisis menggunakan program Quest. Oleh karena lokasi mereka yang jauh dari Yogyakarta, tidak semua di antara mereka dapat melakukannya. Untuk uji coba soal, misalnya mereka harus kembali ke sekolah asal yang jauh tempatnya. Namun demikian, bagi yang bisa dijangkau, mereka melakukannya sesuai prosedur yang seharusnya. Tabel 3 menampilkan hasil analisis data menggunakan Program Quest.

Dari Tabel 3 dapat dilihat bahwa semua guru dapat menganalisis datanya dengan mengaplikasikan program Quest.

Dari 10 perangkat tes yang disusun, ternyata 5 perangkat semua itemnya memenuhi persyaratan item tes model 1-PL. Sebanyak 3 perangkat hanya memiliki  $> 0 - \leq 5\%$  item yang tidak memenuhi persyaratan item 1-PL, dan 2 perangkat tes memiliki  $5 - \leq 10\%$  item yang tidak memenuhi item tes model 1-PL. Kualitas soalnya baik dengan *reliabilitas of estimated* antara 0.77 - 0.89. Semakin mendekati 1.0 reliabilitas semakin baik. Dilihat dari Infit

MNSQ, item yang disusun guru Bahasa Indonesia termasuk baik, karena masuk dalam ukuran antara .77 - 1.30. Hampir semua item dapat diterima, item ditolak hanya berkisar antara 0%- 8%.

Selanjutnya disajikan hasil analisis menggunakan program Quest Guru Bahasa Inggris sebagaimana tampak pada Tabel 4. Dari Tabel 4 dapat dilihat bahwa semua guru dapat menganalisis datanya dengan mengaplikasikan program Quest.

Tabel 3. Kualitas Instrumen *Integrated Assessment* Bahasa Indonesia SMP Berdasarkan Analisis IRT Menggunakan Program QUEST

No.	Nama Guru	Bentuk Item	Jumlah Item	Reliability of Item Estimates	Infit MNSQ	Item Ditolak
1.	Dewi Sulistyaningsih	PG +uraian	25	r = .83	Mean .99SD -.21	2 (8% )
2.	Evi Zuliani	PG +uraian	25	r = .87	Mean 1.00SD -.11	0 (0%)
3.	Irawati Wahyuni	PG +uraian	25	r = .88	Mean 1.00SD -.00	0 (0%_)
4.	Mushofa	PG+uraian	40	r = .88	Mean 1.00SD .12	0 (0%)
5.	Mustajib Sofyan	PG +uraian	40	r = .86	Mean 0.84SD -.19	3 (7.5% )
6.	Nurjanah Yuli	PG+uraian	25	r = .89	Mean .99SD .12	1 (4%)
7.	Puji Kurniawan	PG+uraian	25	r = .62**	Mean 1.00SD .07	0 ( 0%)
8.	Romafi	PG+iraian	25	r = .77	Mean 1.00SD .10	0 ( 0%)
9.	Sri Juwariyah	PG	20	r = .88	Mean .99SD .14	1 (5% )
10.	Sutopo	PG+uraian	40	r = .85	Mean 1.00SD -.12	1(2.5%)

Tabel 4. Kualitas Instrumen *Integrates Assessment* Bahasa Inggris SMP Berdasarkan Analisis IRT Menggunakan Program QUEST

No.	Nama Guru	Bentuk Item	Jumlah Item	Reliability of Item Estimates	Infit MNSQ	Item Ditolak
1.	Ari Priyono	PG	22	r = .85	Mean 1.00SD -.09	0 (0% )
2.	Cesilla Usnaini	PG	22	r = .87	Mean 1.00SD -.11	0 (0%)
3.	Heni Eko Purnomo	PG	20	r = .87	Mean 1.00SD -.10	0 (0%_)
4.	Nanik Agustina	PG	20	r = .85	Mean 1.00SD .143	2 (10%)
5.	Sri Rejeki	PG +uraian	25	r = .89	Mean 0.97SD -.16	1 (4% )
6.	Purba Nurfauzi	PG+uraian	30	r = .89	Mean .99SD .08	0 (0%)
7.	Sukhadman	PG+uraian	25	r = .87	Mean 1.00SD .13	0 ( 0%)
8.	Sunarto	PG+uraian	25	r = .70	Mean 1.00SD .12	0 ( 0% )
9.	Yuli Arisandi	PG+uraian	25	r = .88	Mean .99SD .14	1(4% )
10.	Zaenah Syafitri	PG+uraian	25	r = .78	Mean .99SD -.15	1(4% )

Dari 10 perangkat tes yang disusun, ternyata 6 perangkat semua itemnya memenuhi persyaratan item tes model 1-PL. Sebanyak 3 perangkat hanya memiliki  $>0 - \leq 5\%$  item yang tidak memenuhi persyaratan item 1-PL, dan 1 perangkat tes memiliki  $.5 - \leq 10\%$  item yang tidak memenuhi item tes model 1-PL. Kualitas soalnya baik dengan *reliabilitas of estimated* antara 0.77 - 0.89. Semakin mendekati 1.0 reliabilitas semakin baik. Dilihat dari Infit MNSQ, item yang disusun guru Bahasa Indonesia termasuk baik, karena berada di antara .77 - 1.30. Hampir semua item dapat diterima, item ditolak hanya berkisar antara 0% - 10% .

### Pembahasan

Produk yang dikembangkan pada penelitian Tahun 1 berupa model standarisasi *integrated assessment* hasil belajar bahasa Indonesia dan Inggris SMP bentuk konvergen dan divergen. Model tersebut meliputi: (1) *learning continuum* pembelajaran bahasa Indonesia dan Inggris SMP, (2) kisi-kisi soal ulangan bahasa Indonesia dan Inggris SMP, dan (3) soal tes bahasa Indonesia dan Inggris SMP yang terintegrasi, model konvergen dan divergen, yang terstandarisasi. Selanjutnya, model yang dikembangkan dilengkapi dengan: (1) panduan penyusunan *integrated assessment* hasil belajar bahasa dan (2) panduan analisis soal dengan menggunakan program Quest. Panduan-panduan tersebut sangat membantu para guru dalam mengembangkan penilaian hasil belajar bahasa yang terintegrasi dan sekaligus terstandarisasi.

Prosedur pengembangan telah dilakukan dengan melewati serangkaian kegiatan pengembangan, yaitu: (1) studi awal pengembangan, (2) penyusunan draft model, dan (3) validasi model. Pada studi awal, kegiatan dimulai dari survei lapangan dalam rangka *need assessment*, kajian teori dan konsep tentang *integrated assessment* hasil belajar bahasa serta stan-

darisasi instrument, *FGD*, dan diakhiri dengan penyusunan draft model.

Penyusunan draft model dimulai dari penyusunan kisi-kisi soal, menelaah kisi-kisi, mengembangkan butir soal, dan menelaah butir soal atas kesesuaiannya dengan kisi-kisi yang diacunya. Setelah semua itu dilakukan, diyakini bahwa secara teori, draft model sudah cukup valid.

Validasi model empiris dilakukan dengan cara mengujikan soal yang disusun kepada siswa SMP. Hasil analisis dengan menggunakan program Quest mencerminkan hal-hal sebagai berikut. Pertama, dilihat dari hasil uji empiris dapat diketahui bahwa seluruh item soal mempunyai *reliability of item estimates* yang tinggi, yaitu telah sesuai antara sampel dengan item yang diujikan. Hanya satu set soal yang agak rendah, dengan nilai reliabilitas .40. Kedua, dari hasil wawancara dengan guru, terungkap bahwa ulangan dilaksanakan setelah pelajaran olahraga. Dengan demikian, siswa kurang fokus pada ulangan karena masih berkeringat dan lelah. Ditinjau dari kesesuaian item dengan model, terungkap bahwa 97% set soal mempunyai kesesuaian yang tinggi (antara 80%- 100%). Hanya 1 set yang item tidak fitnya 25% atau penerimaan sebanyak 75%.

Melihat perangkat tes yang disusun guru dan dikaitkan dengan hasil analisis secara empiris menggunakan program Quest tampak bahwa semua guru sudah berhasil memiliki pengalaman dan mampu menyusun item hasil belajar bahasa yang terintegrasi bentuk konvergen dan divergen yang memenuhi standar. Mayoritas instrumen tes sudah menunjukkan hasil yang memuaskan. Bahkan banyak instrumen tes yang semua itemnya sudah memenuhi syarat model 1-PL. Selain itu, hasil yang diperoleh para guru peserta juga sudah dapat menunjukkan secara empiris bahwa panduan yang disusun oleh peneliti dengan memperhatikan masukan

dari para peserta sudah dapat digunakan sebagai acuan kerja para guru.

Pada penelitian Tahun II, para guru sudah melakukan penyusunan kisi-kisi dan menulis item sesuai kisi-kisi. Namun ada beberapa guru yang menulis item dahulu baru menyusun kisi-kisi berdasarkan item. Dengan buku panduan tersebut, pemahaman tentang hal itu menjadi lebih baik. Terkait item bentuk konvergen, pada umumnya mereka telah melakukan hal itu. Namun penyusunan soal bahasa konvergen yang otentik belum banyak dilakukan. Dengan sosialisasi ini pengetahuan mereka menjadi bertambah. Penyusunan item divergen secara benar belum banyak dilakukan oleh para guru. Pemahaman mereka tentang hal itu masih perlu diluruskan. Dengan adanya sosialisasi dan desiminasi ini, pemahaman mereka menjadi mantap.

Terkait item yang terintegrasi, pada umumnya guru-guru belum menerapkannya dalam soal-soal ulangan. Padahal, soal-soal model tersebut adalah khas soal bahasa, karena fungsi bahasa selain sarana berpikir adalah untuk berkomunikasi. Dalam komunikasi, berbagai aspek bahasa akan terintegrasi, dan itulah yang semestinya dilatihkan pada siswa, antara lain dalam bentuk soal ulangan. Dengan adanya kegiatan sosialisasi dan desiminasi ini, pemahaman mereka terkait penyusunan soal ujian bahasa meningkat. Terkait buku panduan penulisan item, pemahaman para guru agak terkendala karena kurangnya contoh. Hal ini menjadi masukan bagi perbaikan buku panduan. Dengan contoh-contoh yang lebih banyak dan bervariasi, para guru akan lebih mudah memahaminya.

Sementara itu, terkait buku panduan analisis item menggunakan program Quest adalah selama ini program analisis item yang mereka kenal adalah *ITEMAN* atau *Anates*. Program tersebut adalah program analisis yang menggunakan teori klasik. Sekarang mereka baru mengetahui

bahwa program klasik tersebut mempunyai kelemahan dan program teori modern, seperti Quest mempunyai banyak keunggulan. Selain itu, di dalam program Quest juga ada program klasiknya. Jadi, sekali pakai, dapat dihasilkan pengukuran berdasarkan teori klasik maupun teori modern. Dengan data simulasi, dipandu tim peneliti, para guru peserta sosialisasi dan desiminasi sudah dapat melakukan analisis data menggunakan program Quest dan sekaligus memaknai hasilnya. Kondisi ini sangat menggembirakan dan semoga kemampuan para guru ini dapat disebarluaskan kepada teman-teman mereka yang belum berkesempatan mengikuti kegiatan ini.

Dengan dikuasainya program analisis menggunakan program Quest, para guru berharap dapat mengumpulkan item-item yang baik, yang nantinya dapat menjadi kumpulan soal yang sudah diketahui kualitasnya, akhirnya dapat menjadi bank soal. Harapan ini sangat mungkin dapat dilaksanakan mengingat kemampuan guru yang telah dilatih dalam kegiatan ini cukup bagus. Pada umumnya para peserta mengusulkan agar pelatihan analisis dengan program Quest ini dapat dilanjutkan pada kesempatan lain, agar mereka dapat berlatih lebih intensif lagi. Selain itu, khususnya untuk segi pemaknaan, akan lebih mendalam lagi jika kegiatan semacam ini dapat dilanjutkan. Hal ini menunjukkan bahwa para guru antusias menyambut sosialisasi dan desiminasi ini. Terkait panduan analisis Quest, para peserta mengusulkan agar disajikan cara konversi dari menu *excel* ke *notepad*. Selain itu, penjelasannya akan lebih mudah diterima jika digunakan bahasa yang mudah dicerna. Hal ini menjadi masukan bagi perbaikan buku panduan yang sedang disosialisasikan.

Peserta sosialisasi dan desiminasi dari Provinsi DIY dan Jateng telah melakukan analisis item dengan soal yang dibuat sendiri dengan rambu-rambu penulisan

soal sesuai panduan penulisan soal. Setelah itu mereka menganalisisnya dengan program Quest. Ternyata mereka berhasil melakukannya dan hasilnya bagus. Dibandingkan dengan peserta sosialisasi dan desiminasi dari provinsi lain, peserta provinsi Jateng dan DIY terlihat lebih baik dari yang lain. Hal itu dapat dipahami mengingat mereka adalah guru-guru terpilih yang telah diseleksi oleh proyek PM2TK. Selain itu, mereka sadar bahwa kegiatan ini dapat membantu penyelesaian studi mereka di S-2.

### SIMPULAN

Simpulan penelitian Tahun I sebagai berikut. Pertama, penelitian ini sudah berhasil mengembangkan 30 set perangkat instrumen *integrated assessment* hasil belajar bahasa Indonesia dan Inggris SMP yang terstandarisasi tipe konvergen dan divergen, yang sesuai dengan *learning continuum* hasil belajar. Standarisasi menggunakan teori respons butir 1-PL lewat program Quest. Kedua, perangkat instrumen terdiri atas: (a) *learning continuum*, (b) kisi-kisi soal, (c) soal ulangan hasil belajar bahasa Indonesia dan Inggris yang terintegrasi antara satu aspek kompetensi berbahasa dengan aspek berbahasa lain, dan (d) kunci soal. Ketiga, penelitian ini sudah menghasilkan Panduan Pengembangan Instrumen *Integrated Assessment* Hasil Belajar Bahasa Model Konvergen dan Divergen. Panduan ini telah memandu guru dalam mengembangkan item soal yang sesuai dengan rambu-rambu penyusunan *integrated assessment*. Keempat, penelitian juga telah menghasilkan Panduan Penggunaan Program Quest untuk Analisis Butir Soal Hasil Belajar Bahasa Model Konvergen dan Divergen. Panduan tersebut telah teruji kelayakannya dan diperbaiki sesuai dengan masukan yang diberikan para guru. Panduan tersebut telah dijadikan pegangan bagi guru dalam melakukan analisis dan menginterpretasi hasil analisis dari hasil ulangan harian/midsemester

Selanjutnya berdasarkan hasil analisis dan pembahasan kegiatan Tahun II disimpulkan sebagai berikut. Pertama, pada Tahun II, penelitian ini telah berhasil melakukan sosialisasi dan desiminasi model pengembangan instrumen *integrated assessment* hasil belajar bahasa mata pelajaran Bahasa Indonesia dan Inggris yang tersatandakan dengan sasaran guru-guru Bahasa Indonesia dan Inggris SMP di tiga provinsi, yaitu Provinsi Kalimantan Selatan, Nusa Tenggara Barat, dan DIY. Lebih dari itu, sasaran sosialisasi dan desiminasi yang semula guru-guru dari tiga provinsi, ternyata meluas menjadi 8 provinsi yaitu Provinsi Jateng, Kalbar, Kalteng, Riau, dan Lampung. Kedua, penelitian ini menghasilkan buku "Panduan Penyusunan Item *Integrated Assessment*" untuk bidang bahasa Indonesia dan Inggris SMP. Dengan buku panduan tersebut, pemahaman para guru dalam hal penyusunan soal ujian bahasa yang otentik dan terintegrasi menjadi lebih baik dan mantap. Ketiga, penelitian ini juga menghasilkan buku "Panduan Analisis Data Menggunakan Program *QUEST*". Buku ini dapat menjadi pegangan guru dalam melakukan analisis menggunakan program pengukuran modern model Rasch (1-PL) sekaligus memaknai hasilnya. Keempat, melalui penelitian Stranas ini, ternyata kompetensi guru dalam evaluasi dapat ditingkatkan. Guru adalah ujung tombak pembaharuan. Dengan diberdayakannya guru, kualitas pengujian akan meningkat dan pada gilirannya kualitas pendidikan akan meningkat pula.

### UCAPAN TERIMA KASIH

Artikel ini merupakan salah satu produk penelitian Stranas, Hibah Dikti, tahun 2012-2013. Oleh karena itu, disampaikan ucapan terima kasih kepada Direktur DP2M, Dikti yang telah membiayai penelitian ini. Terima kasih juga disampaikan kepada Rektor UNY melalui Lembaga Penelitian dan Pengabdian Masyarakat.

UNY yang telah memfasilitasi penelitian ini sehingga semua proses penelitian dapat berjalan sesuai ketentuan. Selanjutnya, disampaikan ucapan terima kasih pada guru-guru SMP DIY, Kalsel, NTB, Kalbar, Kalteng, Lampung, Riau, dan Jateng yang telah berpartisipasi dalam penelitian ini. Diharapkan penelitian ini bermfaat bagi segenap civitas akademika, khususnya dalam asesmen pembelajaran bahasa.

#### DAFTAR PUSTAKA

- Adams, R.J. & Kho, Seik-Tom. 1996. *Acer Quest Version 2.1*. Camberwell, Victoria: The Australian Council for Educational Research.
- Brown, Douglas H. 2004. *Language Assessment Principles and Classroom Practices*. New York: Longman.
- 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.
- 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.
- Heaton, J.B. 1988. *Writing English Language Testing*. London and New York: Longman
- Pappas, Christine C, Kiefer, Barbara Z, Levstik Linda. 1996. *An Integrated Language Perspective in the Elementary School*. London: Longman
- Pariñas, Neil. 2009. Revised Taxonomy: Reframing our understanding of knowledge and Cognitive Procecess. *The Assessment Handbook: Continuing Education Program. Volume 1, May 2009*.
- Roid, G.H. & Haladyna, Th.M. (1982). *A Technology for Test-item Writing*. Orlando: Academic Press, Inc.
- Theissen, 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". In: Thissen, D. & Wainer, H. 2000. *Test Scoring*. Mahwah, New Jerrsey: Lawrence Erlbraum Associates, Publishers.