

BAB I. PENDAHULUAN

A. LATAR BELAKANG

Ujian Nasional merupakan salah satu cara untuk menilai kualitas pembelajaran. Hasil Ujian Nasional tahun 2010 menunjukkan bahwa nilai mata pelajaran bahasa kurang baik. Padahal, bahasa adalah pintu gerbang pengetahuan. Hasil belajar yang lain dapat diprediksi dari hasil tersebut. Beberapa hal bisa diidentifikasi sebagai penyebab, salah satu di antaranya sistem penilaian yang digunakan di sekolah. Bentuk dan cara penilaian dalam banyak hal memberikan pengaruh penting bagi proses pembelajaran, dan karenanya menentukan capaian kompetensi.

Secara umum tujuan pembelajaran bahasa Indonesia dalam KTSP adalah capaian kemampuan berkomunikasi lewat saluran keempat kemampuan berbahasa. Karena itulah, penilaian yang dilakukan juga harus mengukur kemampuan berbahasa itu. Bentuk penilaian yang terbaik dan disarankan dalam kurikulum tersebut adalah penilaian terpadu (*integrated assessment*) yang menekankan pada penilaian kemampuan berunjuk kerja bahasa sebagaimana halnya dalam berkomunikasi sehari-hari.

Menurut Permendiknas Nomor 20 Tahun 2007, dari segi teknik dan instrumen penilaian yang digunakan, setiap satuan pendidikan dituntut memiliki guru yang mampu melakukan pengujian pendidikan dengan instrumen yang benar-benar akuntabel. Kondisi tersebut belum tampak, guru perlu diberdayakan dalam hal itu, khususnya guru bahasa Indonesia dan Inggris, terutama dalam hal *integrated assessment* hasil belajar bahasa yang terstandarkan.

Selama ini penilaian hasil belajar terstandar yang dikenalkan kepada guru adalah tipe konvergen, dan terbatas pada butir soal pilihan ganda. Padahal banyak hasil belajar yang baru dapat diungkap secara tuntas bila digunakan butir soal bentuk uraian. Bentuk soal uraian tersebut berpeluang melatih peserta didik berpikir divergen. Melalui penelitian ini, suatu panduan cara pengembangan *integrated assessment* tipe konvergen dan divergen hasil belajar bahasa, beserta teknik pembakuan dan penafsirannya, akan dibuat. Hasilnya, kualitas guru akan meningkat secara nasional, khususnya untuk mata pelajaran Bahasa Indonesia dan Inggris. Dengan manfaat yang begitu besar, penelitian ini penting sekali untuk dilakukan. Pada skala yang lebih luas, hasil penelitian ini juga dapat digunakan untuk mata pelajaran lain.

B. TUJUAN KHUSUS

Secara khusus, tujuan yang akan dicapai melalui penelitian ini adalah dihasilkannya instrumen *integrated assessment* hasil belajar bahasa Indonesia dan Inggris tipe konvergen dan divergen yang terstandarisasi jenjang SMP, namun demikian hal itu juga dapat dipergunakan untuk jenjang pendidikan dan bahasa lain.

Model tersebut dapat memberdayakan guru SMP di seluruh Indonesia dalam penyusunan instrumen *integrated assessment* hasil belajar bahasa Indonesia dan Inggris yang terstandarkan, menurut teori respons butir. Dari hasil pembuktian secara empiris, dikembangkan panduan model standarisasi *integrated assessment* hasil belajar bahasa Indonesia dan Inggris. Secara rinci, tujuan tersebut dijabarkan sebagai berikut.

1. Tahun Pertama

- a. Mengembangkan instrumen *integrated assessment* hasil belajar bahasa Indonesia dan Inggris sesuai dengan *learning continuum* hasil belajar, khususnya untuk jenjang SMP sebagai *abstract continuum* pengukuran instrumen untuk kedua mata pelajaran tersebut, dan disertai bukti empiris. Instrumen distandarkan menggunakan teori respons butir.
- b. Menyusun panduan penilaian hasil belajar tipe konvergen dan divergen berdasarkan teori dan hasil pembuktian empiris instrumen yang telah diperoleh.

2. Tahun dua

- a. Mendiseminasikan pengembangan penilaian hasil belajar yang terstandarisasi di provinsi yang dijadikan demplot, yang dalam hal ini adalah Provinsi DIY, Kalimantan Selatan, dan Nusa Tenggara Barat.
- b. Finalisasi buku panduan penilaian hasil belajar tipe konvergen dan divergen disertai hasil diseminasi pada tiga provinsi untuk disebarluaskan kepada pihak terkait (Dinas Pendidikan dan LPMP) ke seluruh provinsi di Indonesia.

C. Urgensi (Keutamaan) Penelitian

Masalah penilaian hasil belajar merupakan salah satu masalah penting dalam pembelajaran, termasuk pembelajaran bahasa. Tercapai tidaknya tujuan belajar dapat dilihat dari hasil penilaian yang dilakukan. Demikian juga, kualitas guru dalam mengajar

dan melakukan penilaian dapat dilihat dari kegiatan evaluasi. Bagaimana model, cara, bentuk, dan instrumen evaluasi hasil pembelajaran dalam banyak hal mempengaruhi capaian pembelajaran siswa. Karena itulah, perlu dicari cara dan model penilaian hasil belajar yang tepat agar capaian belajar siswa optimal.

Dengan demikian, peran penilaian hasil belajar bahasa haruslah mendapatkan perhatian serius, tidak sekedar dianggap numpang lewat dan digarap sambil lalu tanpa memikirkan dampak yang sebenarnya cukup besar. Sebagaimana halnya pembelajaran bahasa yang menekankan fungsi komunikatif, penilaian hasil belajar bahasa harus pula ditekankan untuk mengukur kemampuan komunikasi dengan bahasa dalam situasi sewajarnya. Untuk mengukur sesuai dengan kompetensi itu, diperlukan model yang berbeda dengan cara yang telah lazim dilakukan. Salah satu model yang dikemukakan di sini adalah penilaian secara terpadu, atau *integrated assessemnet* hasil belajar bahasa.

Integrated assessment tidak mengukur pengetahuan bahasa demi bahasa itu sendiri, melainkan bagaimana penggunaan bahasa itu dalam komunikasi yang sewajarnya. Model penilaian itu menuntut siswa untuk dapat mempergunakan bahasa sebagaimana fungsi bahasa sebagai alat berkomunikasi, yang memadukan berbagai unsur dan kompetensi komunikasi. Jadi, tekanan penialian adalah kemampuan yang bersifat produktif dan reseptif, baik lisan maupun tertulis secara terpadu.

Model penilaian terpadu atau *integrated assessmenet* belum banyak dikenal dan dipergunakan oleh guru bahasa di sekolah, padahal model inilah yang lebih diutamakan untuk mengukur hasil pembelajaran. Oleh karena itu, penelitian pengembangan tentang penilaian terpadu dalam pembelajaran bahasa sangat penting untuk dilakukan. Dengan mempergunakan pengukuran model terpadu ini, tinggi rendahnya skor hasil pengukuran dijamin dapat mencerminkan kemampuan berbahasa sebenarnya.

Dalam pengembangan instrumen penilaian diperlukan pemenuhan teknik dan prosedur pengembangan serta proses standadisasinya. Untuk pemenuhan teknik dan prosedur pengembangannya, harus ada tujuan penilaian yang memuat aspek yang akan dinilai yang dirumuskan dalam *learning continuum* mata pelajaran. Ditinjau dari segi penskalaan, sampai sekarang belum banyak dikaji penskalaan hasil pengukuran pendidikan menggunakan metode teori respons item. Padahal, tersedia program yang praktis untuk menganalisis tipe item berskala dikotomus, politomus, dan kombinasinya.

Kurikulum tingkat satuan pendidikan (KTSP) memberikan peluang guru, termasuk guru bahasa Indonesia dan Inggris, untuk memperkaya ide, membantu mengoptimalkan kemampuan berpikir peserta didik dalam berpikir konvergen dan divergen, serta mendukung kemampuan berpikir kritis dan kreatif. Namun, ditinjau dari pengembangan proses pembelajaran yang selama ini terjadi, jarang guru mengembangkan pembelajaran semacam itu.. Selama ini peserta didik lebih difokuskan pada pola berpikir konvergen.

Bahasa, baik bahasa Indonesia maupun Inggris, merupakan pintu gerbang pengetahuan. Penguasaan akan bahasa tersebut, akan membuka pintu pengetahuan seluas-luasnya. Jika penguasaan bahasa baik, yang dapat diketahui dari penilaian hasil belajar bahasa, baik secara divergen maupun konvergen, maka penguasaan hasil belajar secara keseluruhan akan baik pula. Oleh sebab itulah, guru bahasa perlu mempunyai kompetensi memadai dalam mengembangkan penilaian hasil belajar yang terstandarkan.

Dari uraian di atas, tampak bahwa diperlukan pengembangan kualitas guru di seluruh tanah air, khususnya dalam pengembangan sistem penilaian hasil belajar bahasa yang terstandarkan. Penelitian ini dapat mengantarkan guru, khususnya guru Bahasa Indonesia dan Inggris pada jenjang SMP dalam memanfaatkan sistem penilaian yang terstandardisasi untuk meningkatkan kualitas pengujian dan pembelajaran.

BAB II. TINJAUAN PUSTAKA

Penerapan model *integrated assessment* dalam pembelajaran bahasa dewasa ini tidak terlepas dari kebijakan penerapan kurikulum berbasis kompetensi (KBK) yang kemudian dikenal dengan sebutan Kurikulum Tingkat Satuan Pendidikan (KTSP). Penerapan penilaian terpadu tersebut juga merupakan konsekuensi logis dari penggunaan pendekatan pembelajaran yang direkomendasikan KTSP, yaitu pembelajaran kontekstual (*Contextual Teaching and Learning*). Terkait dengan perubahan kurikulum, terjadi juga perubahan paradigma dalam penilaian hasil belajar.

A. Paradigma Baru dalam Penilaian Pembelajaran

Penerapan pembelajaran berbasis kompetensi, yang sekarang bernama KTSP, merupakan tuntutan yang tidak terelakkan saat ini mengingat adanya persaingan global pada era globalisasi. Kemampuan atau kompetensi sumber daya manusia menjadi hal yang menentukan dalam persaingan tersebut. Tugas sekolah adalah mengembangkan kompetensi siswa seoptimal mungkin agar siswa mampu bersaing di dunia kerja. Terkait akan hal itu, kriteria keberhasilan dilihat dari kompetensi dasar yang dikuasai siswa.

Penilaian semacam itu merupakan penilaian berbasis kompetensi, suatu hal baru yang berbeda dengan cara sebelumnya. Seperti dikatakan Direktorat Jenderal Pendidikan Tinggi (2005), *assessment* dalam konteks KBK berbeda dengan *assessment* dalam konteks kurikulum yang lain. Perbedaan antara lain pada (a) hasil belajar dinyatakan dengan kompetensi yang dapat ditampilkan, (b) hasil belajar juga mencakup aspek afektif yang terintegrasi dalam mata pelajaran, (c) penilaian menggunakan acuan kriteria, dan (d) penilaian dilakukan secara berkelanjutan.

Karena penilaian berbasis kompetensi tidak untuk membandingkan keberhasilan seseorang dengan orang lain, maka harus memiliki rujukan (*referene*) yang jelas dan pasti. Dengan demikian, penilaian berbasis kompetensi membandingkan tingkat kompetensi yang telah dikuasai seseorang dengan kompetensi yang telah ditetapkan sebagai rujujanya, bukan membandingkan seseorang dengan kelompoknya. Acuan yang digunakan adalah Acuan Kriteria dan bukan Acuan Norma seperti yang digunakan oleh

kurikulum sebelum KBK. Dalam pembelajaran di sekolah digunakan Standar Kelulusan Minimal (SKM) atau Kriteria Kelulusan Minimal (KKM).

B. Karakteristik Penilaian Pembelajaran Bahasa

Dalam era kurikulum berbasis kompetensi, penilaian pembelajaran bahasa juga diwarnai oleh hal itu. Oleh karena penilaian berbasis kompetensi berfokus pada hasil (*output*), bukan pada masukan ataupun proses, maka penilaian pembelajaran bahasa juga diarahkan untuk menentukan penguasaan siswa atas kompetensi yang harus dikuasainya. Oleh karena itu, perlu dilakukan penilaian awal untuk mengetahui seberapa jauh kompetensi yang akan dipelajari telah dikuasai siswa (Suyata, 2005).

Meskipun demikian, tidak berarti bahwa dalam penilaian tersebut proses tidak penting. Proses tetap penting dalam rangka menunjang hasil. Hasil tidak akan menjadi baik, jika proses tidak berjalan baik. Namun demikian tujuan akhir tetap pada hasil yang berupa capaian kompetensi yang dikuasai siswa.

Karena penilaian berbasis kompetensi dilaksanakan terhadap setiap individu untuk menentukan penguasaan kompetensi tertentu, maka penilaian pembelajaran bahasa dilakukan pada setiap siswa. Kegiatan penilaian dapat dilakukan dalam situasi kelompok, misalnya untuk menilai kemampuan berdiskusi dalam memecahkan masalah, mengukur kompetensi berbicara dan menyimak, namun sasaran penilaian tetap pada kemampuan secara individual.

Dalam penilaian berbasis kompetensi, dimungkinkan siswa melakukan evaluasi diri. Hal itu dapat memberikan hasil yang lebih bermakna, baik bagi guru maupun siswa, karena mampu memotivasi mereka dalam menjalankan fungsi dan peran masing-masing. Penilaian terbuka tersebut juga berlaku untuk penilaian bahasa. Dalam penilaian pembelajaran bahasa, *penilaian terbuka* tersebut menjadi suatu hal penting mengingat sifat bahasa yang *aplikatif*. Pembelajaran bahasa dinilai bukan dari teori berbahasa, melainkan pada bagaimana siswa berbahasa (Suyata, 2007).

Hal itu dilakukan karena dalam penilaian bahasa harus memperhatikan hakikat dan fungsi bahasa. Pada hakikatnya, bahasa merupakan hasil budaya manusia yang selanjutnya juga berfungsi sebagai sarana komunikasi. Pendekatan penilaian yang sesuai adalah yang menekankan pada aspek kinerja dan atau kemahiran berbahasa. Dengan demikian, penilaian pembelajaran bahasa tidak mengarah pada sistem bahasa, melainkan pada

bagaimana menggunakan bahasa secara benar sesuai dengan sistem itu. Secara pragmatis (Heaton, 1998) menjelaskan bahasa lebih merupakan satu bentuk kinerja dan performansi daripada sebuah sistem ilmu. Pandangan ini mengarahkan penilaian pembelajaran bahasa haruslah menekankan pada *fungsi bahasa sebagai alat komunikasi* daripada sebagai sistem bahasa.

Karakteristik penilaian pembelajaran bahasa berikutnya adalah *autentik* (Heaton, 1988). Pada dasarnya komunikasi berjalan secara apa adanya. Dengan demikian, penilaian bahasa sebagai alat komunikasi juga akan berjalan sewajarnya, sesuai dengan keadaan senyatanya. Jadi, data yang diperlukan adalah data nyata penggunaan bahasa.

Pada kenyataannya komunikasi nyata akan melibatkan lebih dari satu kemampuan berbahasa, seperti komunikasi lewat telepon melibatkan kompetensi mendengarkan dan berbicara. Dengan demikian, karakteristik penilaian kompetensi berbahasa adalah *terpadu, integrated assessemnt*, antara kemampuan berbahasa yang satu dengan yang lain (Pappas, dkk, 1996).

Terkait dengan hal tersebut, standar penilaian pembelajaran bahasa meliputi (a) kompetensi *mendengarkan*, di antaranya berdaya tahan dalam berkonsentrasi mendengarkan berbagai konteks, memahami dan peka terhadap gagasan, pandangan, dan perasaan orang lain, serta mampu memberikan pendapat, (b) kompetensi *berbicara*, di antaranya mampu berdiskusi, meyakinkan orang, menjelaskan suatu respons, dan mengkritik dalam berbagai keperluan, (c) kompetensi *membaca*, di antaranya membaca berbagai ragam teks, menganalisis informasi dan gagasan, memberikan komentar, menyeleksi, dan menyimpulkan, dan (d) kompetensi *menulis*, di antaranya menulis karangan atau laporan penyelesaian tugas.

C. *Integrated Assessment* dalam Pembelajaran Bahasa

Seperti dikatakan di depan, penilaian *integrated assessment* merupakan salah satu karakteristik penilaian dalam pembelajaran bahasa. Hal itu mengisyaratkan bahwa penilaian dalam pembelajaran bahasa mengarah pada peristiwa-peristiwa berbahasa yang terjadi dalam situasi nyata yang berjalan secara wajar. Dalam situasi tersebut, selalu terkait berbagai unsur dan kompetensi berbahasa secara terintegrasi, yang mendukung kelancaran berkomunikasi. Heaton (1998) dan Weir (1990) menyebutnya *integrative tests*. Penilaian terintegrasi tersebut tampak jelas dalam *performance-based assessment* (Brown, 2004),

seperti kombinasi antara menyimak dan berbicara atau integrasi antara membaca dan menulis. Dalam pembelajaran di sekolah, ditekankan kemampuan peserta didik mendemonstrasikan kemampuan yang dimiliki secara nyata dan bermakna, berlandaskan berbagai kompetensi yang terintegrasi tersebut.

Jadi, penilaian model ini menekankan pada pengukuran kinerja, *doing something*, melakukan sesuatu yang merupakan penerapan ilmu pengetahuan yang telah dikuasai secara teoretis. Fungsi bahasa sebagai alat komunikasi menjadi perhatian utama. Brown (2004) menyebutnya *communicative language testing*. Meskipun demikian, tidak berarti mengabaikan sistem bahasa, sebab terganggunya penguasaan sistem bahasa akan mengganggu penggunaan bahasa dalam situasi nyata. Demikian juga, tidak berarti mengabaikan konteks. Strategi pemahaman konteks, yang biasa disebut pragmatik sangat berperan dalam komunikasi.

Dengan demikian, *integrated assessment* merupakan penilaian holistik (Yeager, 1991). Bentuk keterpaduan dapat bermacam-macam, salah satu di antaranya seperti tampak pada tabel berikut (Suyata, 1996)

Tabel 1: *Integrated Assessment* Antar-aspek Hasil Belajar Bahasa

	Kebhsn	Menymk	Berbicra	Membac	Menuls	Brstra	Prgmtk
Kebahasaan	-	v	V	v	v	V	v
Menyimak		-	V	v	v	V	v
Berbicara			-	v	v	V	v
Membaca				-	v	V	v
Menulis					-	v	v
Bersastra						-	v
Pragmatik							-

Bentuk keterpaduan lain dapat terjadi antara lebih dari dua aspek, seperti menyimak berita TV, menuliskan pokok berita, dan memaknai kata sulit dalam berita tersebut. Dapat juga dilakukan penilaian pemahaman wacana, memaknai kalimat inversi dalam wacana, dan memenuliskan ide pokok wacana tersebut.

Dalam *integrated assessment* hasil belajar bahasa, penilaian dilakukan dalam aspek kognitif, afektif, dan psikomotorik. Penilaian kognitif ditujukan untuk menilai kemampuan berpikir dalam memecahkan masalah, penilaian afektif untuk menilai kompetensi yang

terkait dengan perasaan, sikap, atau motivasi, dan penilaian psikomotorik untuk menilai kinerja. Khusus untuk penilaian kognitif, Pariñas (2009), menyarankan digunakannya *HOT (Higher Order Thinking)* yang dalam Bloom, berupa aplikasi, analisis, evaluasi, dan kreasi. Dalam penelitian ini, *integrated assessment* akan diterapkan pada penilaian bahasa Indonesia dan Inggris.

D. Pengembangan Instrumen Penilaian Hasil Belajar

Pengembangan instrument penilaian merupakan bagian yang tak terpisahkan dari kompetensi guru. Pendidik harus mengakses apa yang dipandang penting, bukan membereskan apa yang dengan mudah dapat diukur (Bacon, 1995). 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) 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 instrument penilaian, ada beberapa tahap yang harus dilalui, yakni: (1) perancangan tes, (2) ujicoba tes, (3) penetapan validitas, (4) penetapan reliabilitas, dan (5) interpretasi skor tes. Kegiatan perancangan tes tercakup di dalamnya yakni: (1) penetapan tujuan, (2) penyiapan tabel spesifikasi, (3) menyeleksi format item yang sesuai, (4) menulis item, dan (5) mengedit item. Kegiatan uji coba instrument penilaian meliputi kegiatan: (1) analisis item pengujian uji coba pertama, (2) analisis item pengujian uji coba kedua, dan (3) penyiapan format siap pakai untuk pengujian. Tantangan terbesar dalam mengembangkan instrument penilaian justru pada lemahnya pemahaman tentang struktur dari substansi pengetahuan yang akan diukur (Ebel & Fresbie, 1986:32-36).

Dalam pengembangan instrumen penilaian yang mengacu pada acuan kriteria, tiga langkah yang harus dilakukan yaitu: (1) mengemukakan maksud/tujuan pembelajaran, (2) menspesifikasikan domain prestasi yang mencerminkan maksud/tujuan pembelajaran, dan (3) mengembangkan item tes. Langkah ketiga bisa mencakup teknologi penulisan item, dan dua langkah sebelumnya mencakup syarat-syarat yang harus dipikirkan dengan saksama—sifat tujuan pembelajaran beserta spesifikasinya, khususnya dalam bentuk

tujuan-tujuan pembelajaran (Roid & Haladyna, 1982) . Stark et al. (2001) menjelaskan prosedur pengembangan instrumen menggunakan teori respons item (*item response theory* atau IRT) dapat mengatasi keterbatasan-keterbatasan CTT.

E. Kajian Validitas

Pada umumnya, suatu instrumen penilaian untuk mencapai representasi yang memadai guna mewakili domain prestasi. merupakan masalah validitas isi sebagaimana dikemukakan oleh Cronbach tahun 1970 juga Helmstadter tahun 1964, dan jenis format yang paling tepat adalah jenis yang memungkinkan jumlah maksimum item untuk dibuat sampel dalam suatu periode waktu tertentu (Roid & Haladyna, 1982).

Validitas konstruk adalah sentral untuk menetapkan mutu instrumen penilaian (Embretson & Gorin, 2001). Definisi konstruk yang komprehensif akan memfokuskan pengembangan item dan tes pada *ability* (kemampuan) dan *trait* (diri spesifik) yang akan diukur. Ada tiga jenis model kognitif, yakni model kognitif penguasaan domain, model kognitif spesifikasi tes, dan model kognitif performans tugas (Gorin, 2006).

Kualitas instrumen tidak terlepas dari pemvalidasiannya secara empirik. Banyak faktor yang dapat menimbulkan *error* pengukuran. Meskipun pengembangan tes sudah dilakukan dengan benar namun ada sumber kekeliruan yang pokok khususnya yang berkaitan dengan analisis pengujian, yang dapat dibagi menjadi tiga sub-elemen, yakni: penskoran, analisis item, dan penyamaan/*equating* (Allalouf, 2007).

F. Kajian Reliabilitas

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).

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) 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) menyajikan suatu prosedur untuk menghitung reliabilitas tes acuan norma jika datanya berupa data kategorik.

Di dalam IRT, varians kekeliruan berkebalikan dari fungsi informasi. Secara lebih terperinci, untuk semua penempuh ujian dengan kemampuan teta yang sama, varians eror (ragam galat) (kondisional) adalah kebalikan dari fungsi informasi teta. Logika untuk perspektif ini bahwa varians eror (ragam galat) tidak ada hubungannya dengan interpretasi semantik kata "informasi" atau "kebalikan/invers".

Smith (2003) memasalahkan reliabilitas pada pengukuran-pengukuran yang dilakukan pada asesmen kelas. Dengan adanya perubahan dari asesmen tradisional ke asesmen alternatif dalam setting asesmen kelas, maka tuntutan berpindah dari perhatian ke pengukuran yang lebih tradisional ke asesmen kelas terutama dalam kaitannya dengan pendekatan dari segi reliabilitasnya.

G. Peskalaan dan Konsekuensi Model Analisis

Instrumen penilaian kemampuan berpikir konvergen dapat diuji menggunakan instrument 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) jawaban yang diharapkan dalam tes uraian terbuka harus luas dan komprehensif. Tes semacam ini kreatif karena menggabung-gabungkan jawaban yang bisa didefinisikan. Dengan demikian, tes ini dimaksudkan sesuai dengan situasi-situasi dimana pengetahuan komprehensif itu harus diuji, dan kunci pemilihan bentuk tes ini adalah menjamin bahwa pengetahuan kognitif yang dimaksudkan 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.

Suatu jawaban model memang esensial untuk penskoran yang tepat. Jawaban model harus mengandung elemen-elemen isi pertanyaan tes uraian. Jika isinya perlu diatur dengan cara tertentu, jawaban model harus mencerminkan pengaturan itu. Akan bijaksana

juga jika dibuat daftar poin-poin utama sehingga jawabannya harus mengandung dan menggunakan nilai-nilai untuk masing-masing jawaban model.

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

Menurut Han & Hambleton (2007) juga Theissen et al. (2001), dalam model-model respons butir dikotomus, hanya jenis data responsnyalah yang benar (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).

Program QUEST memberikan layanan untuk menganalisis data hasil pengujian yang menggunakan skaladikotomus, 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.

BAB III. TUJUAN DAN MANFAAT PENELITIAN

Banyak manfaat yang diperoleh dari penelitian ini, baik secara teoretis maupun paraktis.

A. Manfaat Teoretis

Secara teoretis, manfaat peneltian ini adalah memberikan pendalaman teori dan konsep pengujian pada guru, terkait penyusunan instrumen *integrated assessment* dalam pengujian hasil belajar bahasa yang terstandarkan. Standarisasi instrumen memerlukan konsep yang cukup rumit, terutama pada pemahaman konsep tentang analisis respons butir.

Pendalaman konsep juga terkait pemenuhan teknik dan prosedur pengembangan serta proses standadisasinya. Untuk pemenuhan teknik dan prosedur pengembangannya, harus ada tujuan penilaian yang memuat aspek yang akan dinilai yang dirumuskan dalam *learning continuum* mata pelajaran. Ditinjau dari segi penskalaan, sampai sekarang belum banyak dikaji penskalaan hasil pengukuran pendidikan menggunakan metode teori respons butir. Padahal, tersedia program yang praktis untuk menganalisis tipe butir berskala dikotomus, politomus, dan kombinasinya.

B. Manfaat Praktis

Secara praktis, penelitian ini bermanfaat secara strategis dalam skala nasional.

Bagi dunia pengujian, akan diperoleh intrumen pengujian bahasa yang terstandar, model konvergen dan divergen, yang dapat dipertanggungjawabkan kualitasnya. Dengan kualitas instrumen yang akuntabel, dapat diketahui kompetensi siswa yang sesungguhnya. Kompetensi siswa di bidang bahasa yang baik akan membawa pengaruh positif bagi hasil belajar mata pelajaran lain.

Bagi guru, model yang ditemukan dapat memberdayakan guru SMP di seluruh Indonesia dalam penyusunan instrumen *integrated assessment* hasil belajar bahasa Indonesia dan Inggris tipe konvergen dan divergen yang terstandarkan, menurut teori respons butir. Dari hasil pembuktian secara empiris, dikembangkan panduan model standarisasi *integrated assessment* hasil belajar bahasa Indonesia dan Inggris. Soal yang berkualitas akan mempengaruhi hasil pengukuran, dan kualitas hasil pengukuran mempengaruhi hasil evaluasi pendidikan secara keseluruhan. Model tersebut difokuskan

pada bahasa Indonesia dan Inggris di jenjang SMP, namun demikian model tersebut juga dapat dipergunakan untuk jenjang pendidikan dan bahasa lain.

Bagi Dinas pendidikan dan LPMP, penelitian ini bermanfaat membantu lembaga tersebut meningkatkan kualitas guru di bidang pengujian yang akuntabel. Dengan adanya buku panduan penilaian hasil belajar tipe konvergen dan divergen yang dihasilkan penelitian ini, lembaga tersebut dapat memanfatkannya untuk melatih guru dalam penusunan instrumen pengujian bahasa yang terstandarkan.

Bagi Direktorat Pendidikan Dasar dan Menengah, hasil penelitian ini bermanfaat bagi pencapaian hasil ujian nasional (UN) secara nasional. Bahasa Indonesia dan Inggris adalah dua mata pelajaran yang di UN kan. Jika kedua bahasa tersebut mempunyai skor yang baik, karena sudah dilatih oleh guru masing-masing di sekolah dengan instrumen pengujian yang terstandar, dapat diprediksi hasil UN mata pelajaran lain akan baik juga.

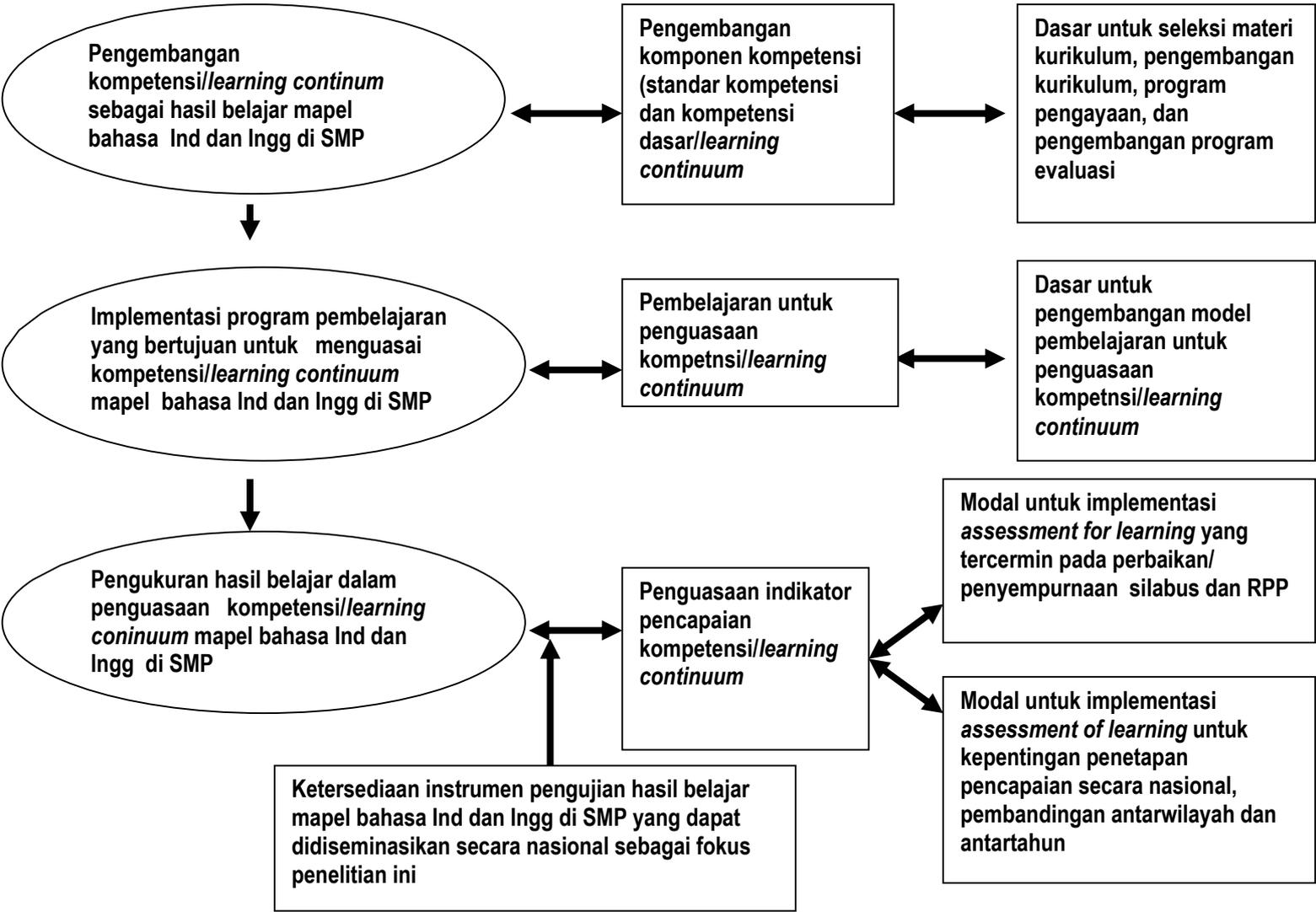
BAB IV. METODE PENELITIAN

A. Kerangka Konseptual

Instrumen penilaian hasil belajar, termasuk hasil belajar bahasa, harus dirancang dan disusun sesuai dengan tujuan pembelajaran yang ditargetkan. Tujuan pembelajaran tersebut dirumuskan dalam kompetensi atau *learning continuum*. Asumsi yang digunakan adalah sebagai berikut.

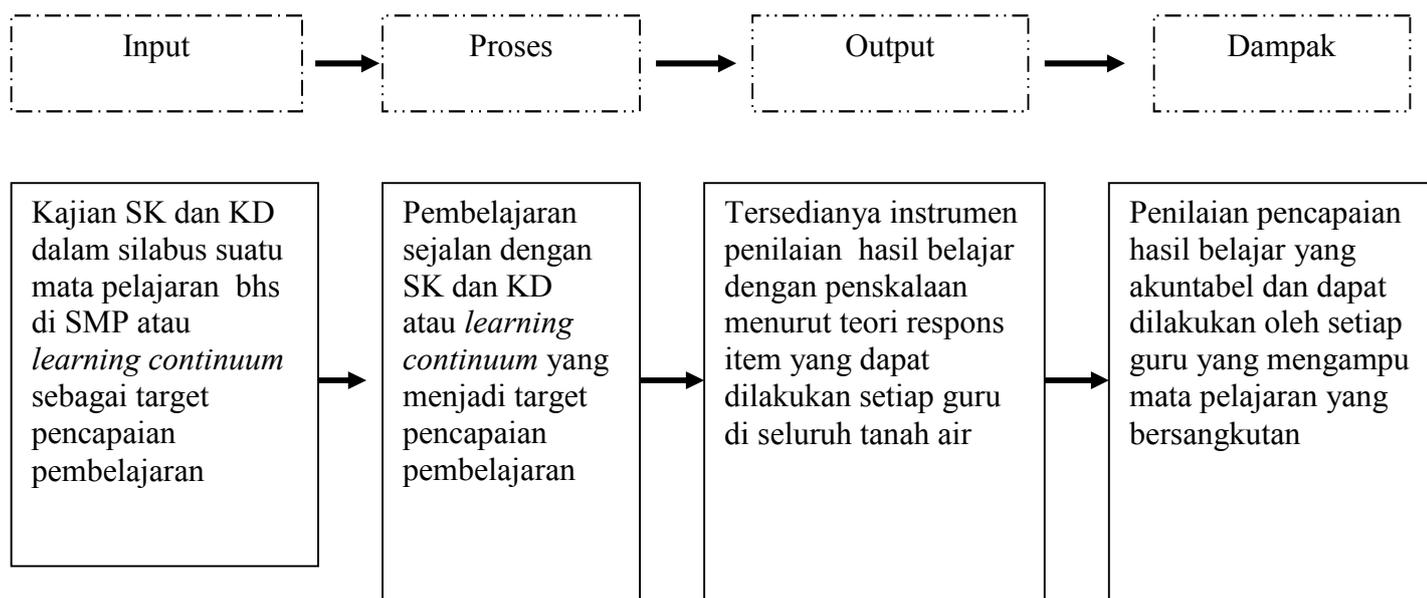
1. Hasil belajar yang ditargetkan harus dirumuskan dalam kompetensi/*learning continuum* yang secara operasional dapat dijabarkan ke dalam indikator pencapaian.
2. Asesmen merupakan bagian yang integral dari program pembelajaran. Oleh karena itu, instrumen pengukur yang digunakan dalam pendidikan dari segi teknik dan prosedur penyusunannya harus dapat dipertanggungjawabkan akuntabilitasnya yang dikembangkan berdasar indikator capaian kompetensi/*learning continuum*.
2. Data hasil pengujian dapat diinterpretasi sehingga benar-benar dapat mencerminkan tingkat capaian penguasaan kompetensi/*learning continuum*.
3. Data hasil pengujian dapat disajikan dalam bentuk skala yang dapat digunakan untuk perbandingan antartahun dengan memanfaatkan prinsip *equating*.

Dengan asumsi dasar di atas, maka hasil-hasil pengujian dalam suatu mata pelajaran, termasuk pelajaran bahasa Indonesia dan Inggris, akan menjadi bagian yang tak terpisahkan dari sistem pembelajaran yang dikemas oleh guru dalam upaya meningkatkan prestasi siswa. Berikut kerangka konsep pengembangan instrumen penilaian hasil belajar beserta metode penskalaannya dan kedudukannya dalam fokus penelitian yang berskala nasional



Gambar 1. Kerangka konsep pengembangan instrumen penilaian hasil belajar Bahasa Indonesia dan Inggris beserta metode penskalaannya dan kedudukannya dalam fokus penelitian yang berskala nasional

Kerangka kerja model *input-proses-output-dampak* sebagai pendekatan dalam penelitian ini akan memberikan gambaran yang lebih jelas mengenai keseluruhan kinerja yang akan dilakukan beserta hasil yang diharapkan. Bila digambarkan dalam bentuk diagram akan tersaji sebagai berikut.



Gambar 2. Model input-proses-output-dampak dalam pengembangan instrumen pengukur hasil belajar di SMP

B. Langkah-Langkah Penelitian

Ada dua pola dalam langkah penelitian hibah stranas ini, sesuai dengan tahapan penelitian tahun pertama dan tahun kedua.

Untuk tahun pertama, pengembangan *learning continuum* bahasa Indonesia dan Inggris sebagai tahapan *base line* yang merupakan tahap dasar untuk menemukan *abstract continuum* sesuai dengan hakikat pengukuran, yang dalam hal ini berupa pengujian hasil belajar sebagai prestasi yang dicapai peserta didik selama menjalani program pembelajaran. Dalam hal ini, kedudukan *learning continuum* sebagai *abstract continuum* merupakan kendali sistem penilaian dalam konteks *assessment for learning* selama peserta didik berada dalam proses mengikuti program pembelajaran yang ditempuh dalam suatu satuan pendidikan, dan akan menjadi kendali *assessment of learning* saat peserta didik mengakhiri program pembelajaran yang ditempuhnya untuk dinyatakan berhasil/lulus atau gagal/tidak lulus. Dalam pengembangan *learning continuum* peneliti mengacu kepada idealisme akademik baik mengacu kepada referensi yang ada maupun berdasarkan kebijakan nasional sesuai dengan Permendiknas Nomor 22 tahun 2006 tentang Standar Isi Pendidikan Dasar dan Menengah serta Permendiknas Nomor 23 tahun 2006 tentang Standar Kompetensi Lulusan Pendidikan Dasar dan Menengah, yang dalam hal ini difokuskan pada mata pelajaran bahasa Indonesia dan Inggris di SMP. Selanjutnya

dilakukan *focus group discussion* (FGD) dengan melibatkan pakar bidang studi dan pakar pendidikan bidang studi, dan pakar penilaian pendidikan, khususnya pakar pengukuran pendidikan. Hasil FGD selanjutnya diseminarkan dengan mengundang sejumlah guru pengampu mata pelajaran yang bersangkutan untuk memperoleh pertimbangan praktis sesuai dengan karakteristik satuan pendidikan

Selanjutnya, diteruskan dengan tahap riset dengan langkah pokok pada tersusunnya buku panduan penilaian hasil belajar yang terstandarisasi khususnya untuk mata pelajaran Bahasa Indonesia dan Inggris, yang didasarkan pada hasil analisis secara empirik dari instrumen yang telah disusun mulai dari penulisan kisi-kisi, penulisan item, penelaahan, dan perakitan, serta uji coba untuk memperoleh bukti empirik.

Indikator keberhasilan tahun I:

- Tersusunnya *learning continuum* pelajaran bahasa Indonesia dan Inggris
- Tersusunnya instrumen *integrated assessment* hasil belajar bahasa Indonesia dan Inggris yang terstandar menurut teori respons butir.
- Berkembangkannya buku panduan penilaian hasil belajar yang terstandarisasi khususnya untuk mata pelajaran Bahasa Indonesia dan Inggris

Pada tahun kedua, kegiatan difokuskan pada tahap diseminasi instrumen berdasarkan buku panduan, yang diawali dengan kegiatan sosialisasi, kemudian dilanjutkan dengan pelaksanaan diseminasi dengan membimbing setiap guru yang tertunjuk di setiap provinsi yang dijadikan demplot untuk dilatih dan mempraktikkannya di SMP, tempat guru Bahasa Indonesia dan Inggris mengampu mata pelajarannya. Kegiatan ini dimonitor dan dievaluasi keberhasilannya untuk selanjutnya dilaporkan kepada pihak yang terkait agar dapat disebarluaskan secara nasional.

Indikator keberhasilan tahun II :

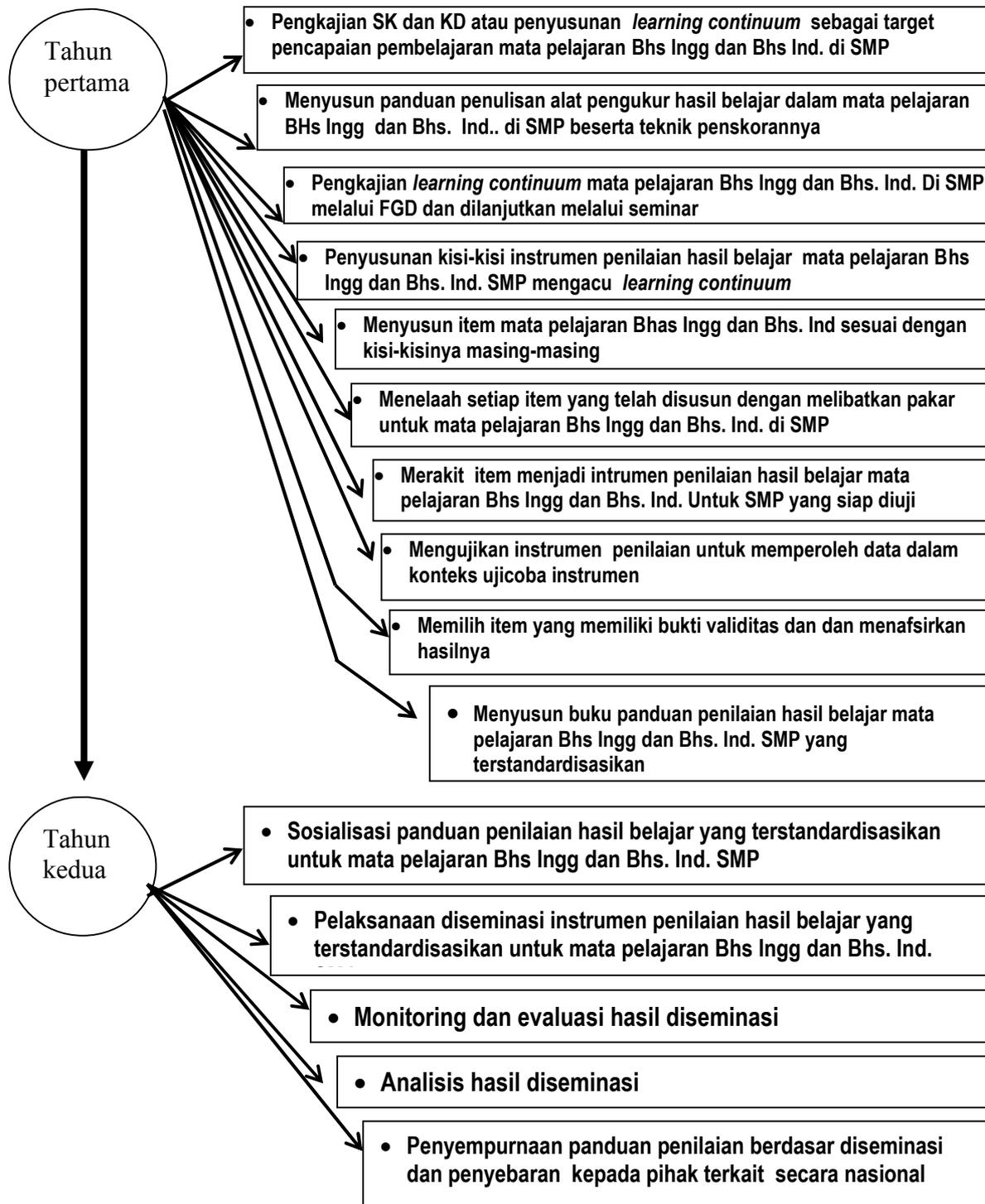
- Tersusunnya buku panduan final *integrated assessment* hasil belajar bahasa Indonesia dan Inggris yang terstandarkan.

-Terdiseminasiakannya dan tersosialisaikannya penyusunan *integrated assessment* hasil belajar bahasa Indonesia dan Inggris yang testandarkan ke seluruh demplot penelitian.

1. Penelitian Tahun Pertama

- a. Mengembangkan *learning continuum* hasil belajar untuk mata pelajaran Bahasa Inggris dan Indonesia SMP sebagai *abstract continuum* pengukuran yang merupakan acuan dalam pengembangan instrument penilaian hasil belajar untuk kedua mata pelajaran tersebut. *Learning continuum* tersebut memuat tujuan penilaian mata pelajaran sesuai dengan aspek yang akan dinilai untuk mata pelajaran yang bersangkutan.
- b. Menelaah *learning continuum* kedua mata pelajaran tersebut melalui *focus group discussion* (FGD).
- c. Menyeminarkan *learning continuum* hasil *focus group discussion* (FGD) dengan mengundang praktisi berupa guru Bahasa Inggris dan Indonesia di SMP.
- d. Melaksanakan penyusunan instrumen penilaian hasil belajar pola konvergen dan divergen untuk mata pelajaran Bahasa Inggris dan Indonesia di SMP dengan menggunakan *learning continuum* sebagai acuan.
- e. Melaksanakan ujicoba instrumen penilaian hasil belajar pola konvergen dan divergen mata pelajaran Bahasa Inggris dan Indonesia SMP tertunjuk menurut rekomendasi Dinas Pendidikan DIY.
- f. Melakukan analisis hasil ujicoba instrumen pengukur hasil belajar pola konvergen dan divergen mata pelajaran Bahasa Inggris dan Indonesia SMP tertunjuk.
- g. Menyusun panduan penilaian hasil belajar pola konvergen dan divergen berdasarkan hasil ujicoba instrumen yang telah dilakukan.

Bila dibuat bagan alur akan tampak tahapan sebagai berikut.



Gambar 3. Tahap penelitian selama dua tahun untuk pengembangan panduan penyusunan instrumen penilaian hasil belajar yang terstandardisasikan untuk mata pelajaran Bahasa Inggris dan Indonesia di SMP

1. Penelitian Tahun Pertama

- h. Mengembangkan *learning continuum* hasil belajar untuk mata pelajaran Bahasa Inggris dan Indonesia SMP sebagai *abstract continuum* pengukuran yang merupakan acuan dalam pengembangan instrument penilaian hasil belajar untuk kedua mata pelajaran tersebut. *Learning continuum* tersebut memuat tujuan penilaian mata pelajaran sesuai dengan aspek yang akan dinilai untuk mata pelajaran yang bersangkutan.
- i. Menelaah *learning continuum* kedua mata pelajaran tersebut melalui *focus group discussion* (FGD).
- j. Menyeminarkan *learning continuum* hasil *focus group discussion* (FGD) dengan mengundang praktisi berupa guru Bahasa Inggris dan Indonesia di SMP.
- k. Melaksanakan penyusunan instrumen penilaian hasil belajar pola konvergen dan divergen untuk mata pelajaran Bahasa Inggris dan Indonesia di SMP dengan menggunakan *learning continuum* sebagai acuan.
- l. Melaksanakan ujicoba instrumen penilaian hasil belajar pola konvergen dan divergen mata pelajaran Bahasa Inggris dan Indonesia SMP tertunjuk menurut rekomendasi Dinas Pendidikan DIY.
- m. Melakukan analisis hasil ujicoba instrumen pengukur hasil belajar pola konvergen dan divergen mata pelajaran Bahasa Inggris dan Indonesia SMP tertunjuk.
- n. Menyusun panduan penilaian hasil belajar pola konvergen dan divergen berdasarkan hasil ujicoba instrumen yang telah dilakukan.

Secara rinci tahap penelitian beserta hasil dan analisis data yang digunakan adalah sebagai berikut.

Tabel 2. Tahapan kegiatan penelitian, hasil yang ingin dicapai, pendekatan yang dilakukan, metode pengumpulan Data dan Teknik Analisis Data untuk Penelitian Tahun Pertama

No	Kegiatan Penelitian	Hasil yang Ingin dicapai	Pendekatan yang Digunakan	Metode Pengumpulan Data	Teknik Analisis Data
1.	Meninjau SK dan KD dalam KTSP atau merumuskan <i>learning continuum</i> untuk mata pelajaran Bhs. Ing dan Bahasa Indonesia di SMP	<ul style="list-style-type: none"> Hasil tinjauan SK dan KD dari KTSP atau rumusan <i>learning continuum</i> untuk mata pelajaran Bhs Ingg dan Bahasa Indonesia di SMP 	Pendekatan kualitatif	Interview dengan guru mata pelajaran dan studi dokumen	Analisis deskriptif kualitatif
2	Mengkaji SK dan KD dalam KTSP untuk merumuskan <i>learning continuum</i> untuk mata pelajaran Bhs. Ing dan Bahasa Indonesia di SMP melalui FGD	<ul style="list-style-type: none"> Hasil kajian SK dan KD yang layak dimasukkan ke dalam <i>learning continuum</i> untuk mata pelajaran Bhs Ingg dan Bahasa Indonesia di SMP 	Pendekatan kualitatif	Studi dokumentasi	Analisis deskriptif kualitatif
3	Menelaah rumusan <i>learning continuum</i> untuk mata pelajaran Bhs Ingg dan Bahasa Indonesia di SMP melalui FGD	<ul style="list-style-type: none"> Hasil telaah <i>learning continuum</i> untuk mata pelajaran Bhs Ingg dan Bahasa Indonesia di SMP yang menjadi dasar untuk merumuskan <i>learning continuum</i> untuk memperoleh masukan dari para praktisi 	Pendekatan kualitatif	<i>Forum Group Discussion</i>	Analisis deskriptif kualitatif
4	Menelaah rumusan <i>learning continuum</i> untuk mata pelajaran Bhs. Ingg dan Bahasa Indonesia melalui seminar yang diikuti para praktisi berua guru SMP untuk mata pelajaran yang bersangkutan	<ul style="list-style-type: none"> <i>Learning continuum</i> untuk mata pelajaran Bhs Ingg dan Bahasa Indonesia yang siap dipakai sebagai acuan untuk mengembangkan instrument penilaian pola konvergen dan divergen untuk mata pelajaran yang bersangkutan 	Pendekatan kualitatif		Analisis deskriptif kualitatif
5	Mengujicobakan instrumen pengukur hasil belajar untuk mata pelajaran Bhs. Ingg dan Bahasa Indonesia di SMP	<ul style="list-style-type: none"> Terujicobakannya instrumen pengukur hasil belajar untuk mata pelajaran Bhs Inggi dan Bahasa Indonesia di SMP 	Pendekatan kuantitatif	Observasi partisipatif	Analisis deskriptif kualitatif

No	Kegiatan Penelitian	Hasil yang Ingin dicapai	Pendekatan yang Digunakan	Metode Pengumpulan Data	Teknik Analisis Data
6	Mermonitor dan terevaluasinya ujicoba instrumen pengukur hasil belajar untuk mata pelajaran Bhs Ingg dan Bahasa Indonesia di SMP	<ul style="list-style-type: none"> • Termonitor dan terevaluasinya ujicoba instrumen pengukur hasil belajar untuk mata pelajaran Bhs Ingg dan Bahasa Indonesia di SMP 	Pendekatan kuantitatif	Observasi partisipatif	Analisis deskriptif kualitatif
7.	Menganalisis hasil uji coba instrumen pengukur hasil belajar untuk mata pelajaran Bhs Ingg dan Bahasa Indonesia di SMP	<ul style="list-style-type: none"> • Teranalisisnya data hasil uji coba instrumen pengukur hasil belajar untuk mata pelajaran Bhs Ingg dan Bahasa Indonesia di SMP 	Pendekatan kuantitatif	Data empirik	Analisis IRT
8	Menyusun panduan penilaian untuk mata pelajaran Bhs Ingg dan Bahasa Indonesia di SMP berdasarkan konsep teoretis dan hasil empiris	<ul style="list-style-type: none"> • Tesusunnya panduan penilaian untuk mata pelajaran Bhs Ingg dan Bahasa Indonesia di SMP berdasarkan konsep teoretis dan hasil empiris 	Pendekatan kualitatif	Kajian teori dan bukti empirik	Analisis deskriptif

2. Penelitian Tahun Kedua

- a. Mengadakan persiapan diseminasi implikasi pengembangan penilaian hasil belajar yang terstandarisasi berdasarkan hasil uji coba pada tahun pertama dengan menggandakan sejumlah peserta, mempersiapkan acara pertemuannya dengan para guru di tiga provinsi demplot. Dalam hal ini, panduan penilaian hasil belajar terstandarisasi adalah untuk mata pelajaran Bhs. Inggris dan Bahasa Indonesia. Panduan tersebut memuat (1) *learning continuum*, (b) cara penyusunan kisi-kisi, (c) cara penyusunan item pola konvergen dan pola divergen, (d) cara menelaah instrumen penilaian hasil belajar secara kualitatif, (e) cara menganalisis instrumen penilaian hasil belajar secara kuantitatif menggunakan data lapangan menggunakan program QUEST, baik untuk data dikotomus maupun data politomus, (f) cara menafsirkan hasil analisis kuantitatif untuk kepentingan *assessment for learning* maupun untuk kepentingan *assessment of learning* pada satuan pendidikan jenjang SMP untuk mata pelajaran Bahasa Inggris dan Indonesia.
- b. Menyelenggarakan sosialisasi implementasi pemakaian panduan pengembangan penilaian hasil belajar yang terstandarisasi.

- c. Membimbing guru mengembangkan instrumen penilaian hasil belajar mengikuti langkah-langkah yang ada di dalam panduan.
- d. Membimbing guru melakukan analisis hasil pengujian di sekolahnya masing-masing sesuai dengan mata pelajaran yang diampu, untuk kepentingan *assessment for learning* maupun untuk kepentingan *assessment of learning*.
- e. Membuat kesimpulan hasil diseminasi pada SMP tertunjuk sesuai dengan rekomendasi Dinas Pendidikan setempat pada beberapa provinsi yang dijadikan demplot, yang dalam hal ini adalah Provinsi DIY, Kalimantan Barat, dan Nusa Tenggara Barat.
- f. Menyebarluaskan panduan pengembangan instrument penilaian pola konvergen dan divergen dalam mata pelajaran Bahasa Inggris dan Indonesia secara luas kepada pihak-pihak yang berkepentingan (Dinas Pendidikan, LPMP, SMP) melalui jalur internet.

Secara rinci tahap beserta hasil dan analisis data yang digunakan adalah sebagai berikut.

Tabel 3. Tahapan kegiatan penelitian, hasil yang ingin dicapai, pendekatan yang dilakukan, metode pengumpulan Data dan Teknik Analisis Data untuk Penelitian Tahun Kedua

No	Kegiatan Penelitian	Hasil yang Ingin dicapai	Pendekatan yang Digunakan	Metode Pengumpulan Data	Teknik Analisis Data
1..	Mempersiapkan sosialisasi diseminasi di tiga provinsi demplot	<ul style="list-style-type: none"> Tersiapkannya kegiatan diseminasi implementasi pengembangan penilaian hasil belajar sesuai panduan yang disiapkan di tiga provinsi demplot 	Pendekatan kualitatif	Observasi partisipatif	Analisis deskriptif kualitatif
2	Mendiseminasikan pengembangan penilaian hasil belajar untuk mata pelajaran Bhs Ingg dan Bhs. Indonesia di SMP pada tiga provinsi demplot	<ul style="list-style-type: none"> Terlaksananya diseminasi pengembangan penilaian hasil belajar untuk mata pelajaran Bhs Ingg dan Bhs. Indonesia di SMP pada tiga provinsi demplot 	Pendekatan kualitatif	Observasi partisipatif	Analisis deskriptif kualitatif
3	Monitoring dan evaluasi pelaksanaan diseminasi instrumen pengukur hasil belajar di SMP untuk mata pelajaran yang bersangkutan	<ul style="list-style-type: none"> Termonitor dan terevaluasinya pelaksanaan diseminasi pengembangan penilaian hasil belajar untuk mata pelajaran Bhs Ingg dan Bhs. Indonesia di SMP pada tiga provinsi demplot 	Penelitian tindakan kelas	Observasi dan refleksi	Analisis deskriptif kualitatif
4	Teranalisisnya data hasil diseminasi pengembangan penilaian hasil belajar untuk mata pelajaran Bhs. Ingg dan Bhs. Indonesia di SMP pada tiga provinsi demplot	<ul style="list-style-type: none"> Teranalisisnya data hasil diseminasi pengembangan penilaian hasil belajar untuk mata pelajaran Bhs Ingg dan Bhs. Indonesia di SMP pada tiga provinsi demplot 	Pendekatan kualitatif	Observasi Partisipatif	Analisis deskriptif kualitatif
5	Tersosialisasikannya hasil diseminasi pengembangan penilaian hasil belajar untuk mata pelajaran Bhs. Ingg dan Bhs. Indonesia di SMP pada tiga provinsi demplot ke pihak terkait yang berkepentingan	<ul style="list-style-type: none"> Tersosialisasikannya hasil diseminasi pengembangan penilaian hasil belajar untuk mata pelajaran Bhs Ingg dan Bhs. Indonesia di SMP pada tiga provinsi demplot ke pihak terkait yang berkepentingan 	Pendekatan kualitatif	Observasi Partisipatif	Analisis deskriptif kualitatif

C. Lokasi Penelitian

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

D. Metode Pengumpulan Data

Pengumpulan data pada tahun pertama dilaksanakan melalui kegiatan sebagai berikut.

1. Pengumpulan data melalui observasi partisipatif dan interview dilakukan untuk pengumpulan data sebagai dasar untuk peninjauan SK dan KD/*learning continuum* mata pelajaran yang bersangkutan sebagai target pembelajaran.
2. Pengumpulan data melalui *forum group discussion* (FGD) untuk memperoleh SK dan KD/*learning continuum* mata pelajaran yang bersangkutan sebagai target pencapaian hasil belajar.
3. Pengumpulan data melalui pengetesan dilakukan selama uji coba instrumen pengukur hasil belajar untuk mata pelajaran yang bersangkutan di SMP.

Pengumpulan data pada tahun kedua dilaksanakan melalui kegiatan sebagai berikut.

1. Pengumpulan data melalui kegiatan diseminasi instrumen pengukur hasil belajar untuk mata pelajaran yang bersangkutan di SMP tempat guru mengajar.
3. Pengumpulan data melalui observasi partisipatif dan dilakukan untuk monitoring dan evaluasi pelaksanaan program diseminasi.

D. Teknik Analisis Data

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 yang bersangkutan di SMP, juga hasil review dan revisi instrumen pengukur hasil belajar untuk mata pelajaran yang bersangkutan. Analisis data secara kuantitatif menggunakan pendekatan IRT dikhotomus dan politomus dengan paket program QUEST (Adams & Kho, 1996).

BAB V. HASIL PENELITIAN DAN PEMBAHASAN

A. HASIL PENELITIAN

Penelitian pengembangan standarisasi *integrated assessment* hasil belajar bahasa model konvergen dan divergen ini dilakukan dengan serangkaian kegiatan. Kegiatan dan hasil kegiatan tersebut adalah sebagai berikut.

1. Studi Pendahuluan

Studi pendahuluan menganalisis permasalahan mengenai lemahnya kompetensi guru dalam evaluasi pembelajaran, mengkaji teori-teori terkait konsep *integrated assessment* dalam pengujian bahasa, dan model-model penilaian yang relevan dengannya. 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 Response Theory* satu parameter. Mereka bersemangat dan sangat antusias ingin mengetahui dan mengaplikannya dalam pengembangan instrumen.

Hasil kajian pada studi pendahuluan dijadikan dasar untuk mengembangkan draft produk instrumen, yaitu instrumen penilaian terintegrasi yang terstandarisasi bentuk konvergen dan divergen dalam pengujian bahasa Indonesia dan Inggris SMP.

2. Pengembangan Draft Produk Instrumen

Pengembangan draft produk instrumen dilakukan melalui serangkaian diskusi mengenai konsep dasar pengujian pembelajaran bahasa yang menuntut kompetensi siswa dalam berbahasa dan berbagai faktor terkait pengembangan instrumennya. Hasil kegiatan tersebut sebagai berikut

a. Pengembangan *Learning Continuum*

Kegiatan pengembangan draft produk diawali dengan pengembangan *learning continuum* untuk mata pelajaran Bahasa Indonesia dan bahasa Inggris jenjang SMP yang memuat aspek kompetensi untuk mata pelajaran yang bersangkutan. *Learning continuum* dalam hal ini berkedudukan sebagai *abstract continuum* pengukuran yang merupakan acuan dalam pengembangan instrumen penilaian hasil belajar untuk kedua mata pelajaran tersebut. *Learning continuum* mata pelajaran dapat dilihat pada Lampiran.

b. Telaah *Learning Continuum*

Selanjutnya *learning continuum* yang disusun 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 Bahasa Indonesia dan tiga Pakar Pendidikan bahasa Inggris sebagai penelaah. *Learning continuum* yang telah ditelaah inilah yang dijadikan acuan bagi guru untuk mengembangkan instrumen pengukur hasil belajar. Hasil telaah terlampir pada Lampiran 1a untuk Bahasa Indonesia dan 1b untuk Bahasa Inggris.

c. Seminar *Learning Continuum*

Learning continuum yang telah ditelaah oleh pakar diseminarkan dengan mengundang praktisi, yaitu guru bahasa Indonesia dan Inggris SMP, pengurus MGMP Bahasa Indonesia dan Inggris di Prov. DIY. Masukan-masukan dari para praktisi digunakan untuk menyempurnakan *learning continuum* yang tersusun. Masukan dari para praktisi ada pada Lampiran.

d. Penyusunan Panduan

Kegiatan selanjutnya, tim peneliti menyusun panduan. Ada dua panduan yang disusun, yaitu (1) Panduan Penyusunan *Integratted Item* Tes Hasil Belajar Bahasa Pola Konvergen dan Divergen, (2) Panduan Analisis Item Menggunakan Program QUEST: Skala Dikotomus, Politomus, dan Kombinasinya. Panduan Penyusunan *Integratted Item* Tes Hasil Belajar Bahasa Pola Konvergen dan Divergen bertujuan memandu para guru Bahasa Indonesia dan Inggris SMP dalam menyusun item-item yang terintegrasi, satu aspek dengan aspek yang lain bentuk konvegen dan divegen. Item disusun sedemikian rupa agar situasi berkomunikasi secara nyata yang melibatkan lebih dari satu aspek terwujud. Panduan kedua, yaitu Panduan Analisis Item Menggunakan Program QUEST:

Skala Dikotomus, Politomus, dan Kombinasinya bertujuan memandu para guru melakukan analisis dengan program QUEST dan memaknai hasilnya.

e. Pelatihan Penulisan *Integrated* Item

Setelah panduan disusun, diundang guru Bahasa Indonesia dan bahasa Inggris SMP yang berasal dari SMP Kabupaten dan Kota, Prov. DIY melalui MGMP. Guru-guru yang diundang terutama pengurus MGMP dengan harapan mereka dapat menularkan ilmu yang diperoleh kepada para anggota MGMP lain. Guru-guru diundang dalam rangka pelatihan penulisan *integrated item* model konvergen dan divergen. Adapun daftar guru peserta penulis soal adalah sebagai berikut.

Tabel 4: Peserta Standarisasi *Integrated Assessment* Bahasa Indonesia

No	Nama Lengkap	Nama Instansi (Lengkap)
1	Ratnasari, S.Pd	SMP 16 Yogyakarta
2	Sukinah, S.Pd	SMP 12 Yogyakarta
3	Dra. Septiana Listiyaningrum	SMP 8 Yogyakarta
4	Tri Yennie Suryaningsih, S.Pd.	SMP Muh 9 Yogyakarta
5	Sri Mulyani, S.Pd.	SMP 7 Yogyakarta
6	Endang Cahyaningsih	SMP 8 Yogyakarta
7	Widig Cagyono, S.Pd	SMP 9 Yogyakarta
8	Riyanto, S.Pd.	SMP 2 Depok, Sleman
9	M. Th. E. Rina Listiana, S.Pd, M.Hum	SMP 1 Ngaglik, Sleman
10	Cahyaningsih, M.Pd	SMP 4. Pakem, Sleman
11	Sri Mariyani, M.Pd.	SMP Muh 1, Godean, Sleman
12	Sudaryanto, S.Pd	SMP 1 Depok, Sleman
13	Haryono, S.Pd	SMP 2 Bambanglipura
14	Rusmantara, S.Pd.	M. Ts Pundong, Bantul
15	Suprihatin, S.Pd.	SMP 3 Jetis, Bantul

Tabel 5. Peserta Standarisasi *Integrated Assessment* Bahasa Inggris

No	Nama Lengkap	Nama Instansi (Lengkap)
1	Ani Pujiastuti, M.Hum.	SMP 1 Bambanglipura, Bantul
2	Erni Etik Suryati, M.Pd	SMP 1 Sedayu, Bantul
3.	Asri Desire Ulfa, S.Pd.	SMP 2 Pleret, Bantul
4.	Drs. Supriatmaja	SMP 1 Pandak, Bantul
5.	Wahyudi. S.Pd	SMP 1 Semanu, Gunungkidul
6.	Kariyadi, S.Pd	SMP 1 Karangmaja, Gunungkidul
7.	Suprpta, S.Pd.	SMP 2 Wonosari, Gunungkidul
8.	Laily Amon Fajariyah, S.Pd.	SMP 5 Panggang, Gunungkidul
9.	Nurkhamid, S.Pd.	SMP 4 Panggang, Gunungkidul
10	Caecilia Winarni, S.Pd.	SMP 1 Semanu, Gunungkidul
11	Sunaryono, S.Pd.	SMP 5 Samigaluh, Kulonprogo
12	Nuruidin, S.Pd.	SMP 1 Panjatan, Kulonprogo
13	Nirnaeni Amperawati, S.Pd.	SMP 1 Pengasih, Kulonprogo
14	Munafir, M.Pd..	SMP 4 Pengasih, Kulonprogo
15	Kuntit Aliningsih, S.Pd.	SMP 4 Pengasih, Kulonprogo

Ketigapuluh guru tersebut bertanggungjawab menyusun : (1) Kisi-kisi soal, (2) Instrumen tes untuk ulangan harian/ulangan tengah semester menggunakan panduan yang telah disusun peneliti, dan (3) Kunci jawaban. Selama pelaksanaan penulisan item tes, peneliti memperhatikan masukan-masukan guru untuk menyempurnakan panduan yang disusun peneliti.

Hasil pada langkah ini adalah (1) 15 kisi-kisi instrumen penilaian hasil belajar Bahasa Indonesia dan 15 kisi-kisi hasil belajar Bahasa Inggris yang terintegrasi antara satu aspek dengan aspek berbahasa lain. (2) Instrumen tes untuk ulangan harian/ulangan tengah semester sebanyak 15 set untuk bahasa Indonesia dan 15 set untuk bahasa Inggris, dan (3) kunci jawaban sebanyak 15 set untuk bahasa Indonesia dan 15 set untuk bahasa Inggris. Hasil pengembangan kisi-kisi soal, kunci soal, dan instrumen tes ada pada Lampiran.

3. Validasi Produk

Tahap ketiga dari pengembangan produk ini adalah validasi produk. Validasi produk dilakukan, 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. Berikut penjelasan terkait hal itu.

a. Telaah Item Silang

Item tes yang telah disusun dirangkai menjadi perangkat tes dan dilakukan telaah silang antar- penulis sebelum soal diujikan. Telaah dilakukan dengan menyesuaikan soal dengan kisi-kisi yang disusun dan memperhatikan substansi materi, konstruksi, serta bahasa. Hasil telaah berkisar pada kurang sesuainya item dengan indikator yang ada, konstruksi item yang kurang mencerminkan integrasi antar- aspek berbahasa, dan perlunya perbaikan dari segi bahasa.

b. Focus Group Discussion

Masukan dari telaah silang antar-penulis didiskusikan untuk memperoleh keputusan final bagi perbaikan item yang disusun. Salah satu topik diskusi adalah bagaimana mengintegrasikan secara tepat, dalam bentuk soal. antara aspek kebahasaan

dengan kompetensi membaca atau integrasi antara aspek membaca dan menulis. Hasil diskusi antara lain, soal disusun menggunakan bentuk integratif dan pragmatik. Unsur pragmatik perlu diperhatikan sebab pemilihan kata dan kalimat yang tepat akan mempengaruhi pemaknaan dan pemahaman bacaan dan tulisan. Selanjutnya, masukan-masukan dari kegiatan *Focus Group Discussion (FGD)* dan telaah teman sejawat digunakan untuk menyempurnakan item yang disusun.

c. Uji Coba Item

Selanjutnya perangkat tes yang telah tersusun diujikan di sekolah masing-masing, baik sebagai ulangan harian atau ulangan tengah semester. Lembar jawaban siswa digunakan sebagai data empirik untuk bahan analisis kualitas item yang dikembangkan. Pada uji coba soal, dimasukkan 5 item untuk bahasa Indonesia dan 5 item untuk bahasa Inggris yang berfungsi sebagai Anchor.

d. Pelatihan Analisis Item dengan Program QUEST

Setelah dilaksanakan pengujian di sekolah masing-masing, selanjutnya data hasil ulangan dianalisis menggunakan program QUEST berpedoman pada panduan yang telah disusun peneliti. Pada awal kegiatan, para guru diminta membaca buku panduan analisis QUEST. Selanjutnya, tim peneliti menjelaskan teori dan konsep analisis item menggunakan program QUEST, suatu program analisis modern dengan menggunakan *Item Respons Theory* 1 parameter.

Berikutnya, peserta dilatih melakukan analisis dengan data masing-masing. Pada analisis QUEST pertama ini, belum semua guru dapat melakukannya dengan baik. Sebagian guru yang sudah berhasil menjadi tutor bagi teman yang masih belum berhasil. Pada akhirnya, dari 30 orang guru yang berhasil melakukan analisis QUEST baru sebanyak 16 orang dan dua hari kemudian ada 3 orang lagi yang mengirimkan hasilnya.

Bagi guru bahasa, tampaknya program QUEST yang menggunakan teori pengukuran modern masih agak menyulitkan, pada pelatihan pertama belum berhasil semuanya. Para guru menjadi penasaran dan mengusulkan agar dibuka pelatihan gelombang kedua sebab mereka ingin melakukannya sampai bisa. Karena itulah dibuka program pelatihan analisis QUEST yang kedua.

Pada pelatihan kedua, program pelatihan agak diperbaiki, yaitu aplikasi program QUEST dengan data simulasi dahulu, kemudian baru dengan data sesungguhnya. Ternyata, dengan cara seperti itu, pelatihan lebih berhasil dan semua guru puas karena berhasil melakukan analisis dengan program analisis modern, yaitu QUEST. Analisis dengan program tersebut memang menjadi dambaan mereka sebab program tersebut dianggap praktis karena dapat menganalisis bentuk soal yang dikhotomus, politomus, maupun kombinasinya sekaligus.

Analisis dilakukan Laboratorium Komputer, FBS, UNY. Selama pelaksanaan analisis, para guru memberikan masukan agar panduan yang disusun peneliti lebih operasional, terutama istilah-istilah yang digunakan masih sulit dipahami oleh guru-guru bahasa. Hasil akhir panduan, setelah memperhatikan masukan dari para guru peserta dilampirkan pada Lampiran 3.

Hasil kerja guru berupa 30 perangkat tes Bahasa Indonesia dan Inggris SMP dilampirkan pada Lampiran. Semua perangkat tes sudah disempurnakan berdasarkan masukan dari kegiatan telaah silang antarpeserta. Hasil analisis menggunakan program QUEST terhadap item tes yang disusun guru Bahasa Indonesia dan Inggris dilampirkan pada Lampiran . Jika dilihat dari tanggal analisis dapat diketahui bahwa setelah guru dijelaskan secara teoretik cara menganalisis menggunakan panduan yang sudah ada, di antara para guru ada yang sudah dapat berhasil melakukan analisis sendiri sebelum jadwal pertemuan dan ada pula yang melakukan setelah dijelaskan pada waktu pertemuan. Namun ada pula yang baru berhasil setelah tiga hari pertemuan. Secara ringkas berdasarkan hasil uji coba di lapangan diperoleh hasil sebagai berikut.

Tabel 6. Kualitas Instrumen Bahasa Indonesia Berdasarkan Analisis IRT Menggunakan Program QUEST

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

Tabel 3 menunjukkan bahwa soal-soal yang disusun oleh guru bahasa Indonesia adalah 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.

Dari Table 3 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 semua (penolakan 0%), 2 set ada penolakan 5%, dan 2 set penolakan 10%.

**Tabel 7. Kualitas Instrumen Bahasa Inggris
Berdasarkan Analisis IRT Menggunakan Program QUEST**

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

Tabel 4 menunjukkan bahwa soal-soal yang disusun oleh guru bahasa Indonesia adalah dalam bentuk konvergen dan divergen, dengan item yang berjumlah antara 20-30.

Soal mempunyai nilai reliabilitas yang berada pada rentang 0,72 – 0,94. 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 ditunjukkan berdasarkan case atau testi. Semakin tinggi nilainya, semakin meyakinkan bahwa pengukuran memberikan hasil yang konsisten.

Dari Table 4 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%.

B. Pembahasan

Produk yang dikembangkan 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.

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 survey lapangan dalam rangka need assessment, kajian teori dan konsep tentang *integrated assessment* hasil belajar bahasa serta standarisasi 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. 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. 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.

BAB VI SIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah diperoleh, simpulan penelitian Tahun I ini sebagai berikut.

1. 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*.
2. 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.
3. 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*.
4. 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

B. Saran

Dari hasil penelitian dan simpulan dapat ditarik beberapa saran sebagai berikut.

1. Hasil penelitian ini dapat digunakan bagi pengembangan instrumen pengujian hasil belajar bahasa lain, seperti bahasa Jawa, Perancis, Jerman, atau yang lain. Demikian juga hasil penelitian ini dapat digunakan untuk pengujian hasil belajar bahasa pada jenjang yang lain, seperti SD, SMA, atau SMK.
2. Sesuai rencana, pada tahun kedua penelitian ini akan dilakukan kegiatan sosialisasi dan desiminasi di 3 provinsi, yaitu NTB, Kalsel, dan DIY. Dalam

melaksanakan kegiatan tersebut perlu kerjasama yang baik antara peneliti dan Dinas Pendidikan dan MGMP setempat dalam pengaturan waktu agar tidak berbenturan dengan kegiatan lain dan jadwal sekolah. Selain itu, jika pelaksanaan di dua tempat di luar Jawa harus sampai kegiatan mengujikan hasil tes sebagai ulangan harian disekolahnya masing-masing, maka pelaksanaan harus dilakukan dalam dua tahap dengan waktu yang berbeda. Hal ini perlu menjadi perhatian terkait waktu dan dana yang digunakan.

3. Bagi peneliti berikutnya, hasil penelitian ini dapat dijadikan acuan bagi pengembangan instrumen hasil belajar lain yang terstandarisasi. Dengan instrumen standar, kualitas pengujian akan lebih baik, dan itulah yang sangat diharapkan oleh semua pihak.

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.
- Brouwn, Douglas H. (2004). *Language Assessment Principles and Classroom Practices*. New York: Longman.
- 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.
- 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.
- Heaton, J.B. (1998). *Writing English Language Testing*. London and New York: Longman.
- Kupermintz, H. (2004). On the reliability of categorically scored examinations [Versi elektronik]. *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
- 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.

- 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
- Suyata, Pujiati. (1996). *Teori dan pelaksanaan evaluasi pembelajaran bahasa indonesia*. Jakarta: Direktorat Pendidikan Menengah Umum, Depdiknas.
- Suyata, Pujiati.(2005). "Cara Melakukan Penilaian Berbasis Kompetensi yang Menyatu dengan Pembelajaran Bahasa". Bahan penataran Himpunan Evaluasi Pendidikan Indonesia (HEPI) bekerjasama dengan Satker Pembinaan Pendidikan PLP, Dinas Pendidikan, Provinsi DIY.
- Suyata, Pujiati.(2007). *Paradigma baru dalam pembelajaran bahasa*. Yogy: FBS, UNY.
- Thissen, D., Nelson, L, & Surygert, K.A. (2001). Item response theory applied to combination of multiple-choise and constructed response items—Approximation methods for scale score. In: Thissen, D. & Wainer, H. (2001). *Test Scoring*. Mahwah, New Jerrsey: Lawrence Erlbraum Associates, Publishers.