

## PENGEMBANGAN TES KEMAMPUAN BERPIKIR TINGKAT TINGGI FISIKA (PysTHOTS) PESERTA DIDIK SMA

<sup>1)</sup>Edi Istiyono, <sup>2)</sup>Djemari Mardapi, <sup>3)</sup>Suparno

<sup>1),3)</sup>Jurusan Pendidikan Fisika FMIPA Universitas Negeri Yogyakarta

<sup>2)</sup>Jurusan Pendidikan Teknik Elektro FT Universitas Negeri Yogyakarta

<sup>1)</sup>edi\_istiyono@uny.ac.id, <sup>2)</sup>djemarimardapi@gmail.com, <sup>3)</sup>suparno\_mipa@uny.ac.id

### Abstrak

Penelitian dilakukan untuk mengembangkan instrumen kemampuan berpikir tingkat tinggi fisika (PhysTHOTS) peserta didik SMA dan mendapatkan karakteristik PhysTHOTS. Kisi-kisi instrumen disusun berdasarkan aspek dan subaspek kemampuan berpikir tingkat tinggi, yang selanjutnya digunakan untuk menyusun *item-item*. Instrumen terdiri atas dua perangkat tes yang masing-masing memiliki 26 *item* termasuk delapan *anchor item* dan telah divalidasi oleh ahli pengukuran, ahli pendidikan fisika, ahli fisika, dan praktisi. Instrumen yang telah divalidasi diujicobakan pada 1.001 siswa dari sepuluh SMAN di Daerah Istimewa Yogyakarta. Data politomus dianalisis menggunakan *Partial Credit Model* (PCM). Hasil uji coba menunjukkan bahwa semua *item* sebanyak 44 dan instrumen PhysTHOTS terbukti *fit* dengan PCM, reliabilitas instrumen sebesar 0,95, indeks kesukaran *item* mulai -0,86 sampai 1,06 yang berarti semua *item* dalam kategori baik. Dengan demikian, PhysTHOTS memenuhi syarat digunakan untuk mengukur kemampuan berpikir tingkat tinggi fisika peserta didik SMA.

**Kata kunci:** pengembangan instrumen, tes kemampuan berpikir tingkat tinggi, fisika, politomus, dan PCM

## DEVELOPING HIGHER ORDER THINKING SKILL TEST OF PHYSICS (PhysTHOTS) FOR SENIOR HIGH SCHOOL STUDENTS

<sup>1)</sup>Edi Istiyono, <sup>2)</sup>Djemari Mardapi, <sup>3)</sup>Suparno

<sup>1),3)</sup>Jurusan Pendidikan Fisika FMIPA Universitas Negeri Yogyakarta

<sup>2)</sup>Jurusan Pendidikan Teknik Elektro FT Universitas Negeri Yogyakarta

<sup>1)</sup>edi\_istiyono@uny.ac.id, <sup>2)</sup>djemarimardapi@gmail.com, <sup>3)</sup>suparno\_mipa@uny.ac.id

### Abstrak

The objectives of this research were to develop an instrument for measuring senior high school students' physics higher order thinking skills (PhysTHOTS) and to obtain the characteristics of the PhysHOTS. The instrument blue print was developed based on the aspects and sub-aspects of high order thinking skills, then it was used to develop the items. Two sets of instrument consisting of 26 items and each, including eight anchor items were then validated by promotors, measurement experts, physics specialists, physics education experts, and practitioners. The validated instruments were then tried out on 1,001 students of ten senior high schools throughout Special Province of Yogyakarta. The polytomous data were analyzed according to the Partial Credit Model (PCM). The results show that the 44 items and PhysTHOTS were fit to the PCM, the reliability of the test was 0.95, the items' difficulty indexes were between -0.86 and 1.06. Therefore, the PhysTHOTS are qualified to measure senior high school students' physics higher order thinking skills.

**Keywords:** instrument development, physics test of higher order thinking skills, polytomous, and PCM

## Pendahuluan

Dewasa ini dunia berada pada era globalisasi. Pada era ini persaingan cukup ketat, yakni persaingan kualitas sumber daya manusia (SDM). Kualitas SDM bangsa ditentukan oleh tingkat pendidikan bangsa tersebut. Peningkatan kualitas pendidikan dimulai dari peningkatan kualitas pembelajaran. Peningkatan kualitas pembelajaran dapat dimulai dengan menyusun tujuan pembelajaran yang tepat.

Salah satu tujuan Mata Pelajaran Fisika di SMA agar peserta didik memiliki kemampuan mengembangkan kemampuan bernalar dalam berpikir analisis induktif dan deduktif dengan menggunakan konsep dan prinsip fisika untuk menjelaskan berbagai peristiwa alam dan menyelesaikan masalah, baik secara kualitatif maupun kuantitatif (BSNP, 2006, p.160). Hal ini diperkuat dengan anjuran pemerintah dalam kurikulum tingkat satuan pendidikan. Penilaian (asesmen) hendaknya direncanakan untuk mengukur pengetahuan dan konsep, keterampilan proses sains (KPS), dan penalaran tingkat tinggi (Pusat Kurikulum, 2007, pp.23-24). Dengan demikian, melalui pembelajaran fisika diharapkan peserta didik dapat mengembangkan diri dalam berpikir. Peserta didik dituntut tidak hanya memiliki kemampuan berpikir tingkat rendah (*lower order thinking*), tetapi sampai pada kemampuan berpikir tingkat tinggi (*higher order thinking, HOT*).

Berkenaan dengan kemampuan berpikir tingkat tinggi ini, fakta menunjukkan bahwa prestasi fisika yang diukur pada aspek *reasoning* Indonesia berada pada ranking 40 dari 42 negara (TIMSS & PIRLS International Study Center, 2012, p.48). Hal senada dinyatakan Efendi (2010, p.393) bahwa berdasarkan hasil TIMSS dapat ditarik kesimpulan sebagai berikut: (1) rata-rata capaian fisika siswa Indonesia ditinjau dari aspek kognitif (*knowing, applying, reasoning*) masih rendah; (2) kecenderungan capaian fisika siswa Indonesia selalu menurun pada tiap aspek kognitif sehingga kemampuan fisika siswa Indonesia harus ditingkatkan pada semua aspek, khususnya aspek *reason-*

*ing* dengan cara membekali siswa kemampuan berpikir tingkat tinggi. Dengan demikian, prestasi fisika yang menuntut kemampuan berpikir tingkat tinggi siswa sekolah menengah Indonesia, di kancah internasional masih rendah. Prestasi belajar fisika rendah dapat disebabkan karena proses pembelajaran atau model asesmennya yang tidak tepat. Dalam hal ini hanya akan dibahas tentang asesmennya, karena asesmen yang tepat dapat mendorong siswa untuk belajar dengan berpikir tingkat tinggi.

Berdasarkan teori perkembangan Piaget, tahap operasional formal adalah tahap anak mulai berusia sebelas tahun. Pada tahap ini anak sudah mulai dapat mengembangkan kemampuan untuk memanipulasi konsep abstrak melalui penggunaan proposisi dan hipotesis (Piaget, 2005, p.122 dan Reedal, 2010, p.7). Usia siswa SMA antara 15 sampai 18 tahun, sehingga kemampuan berpikir tingkat tinggi mereka sudah mapan.

Piaget mengatakan bahwa kematangan dan kesiapan seseorang harus menunggu serta harus cocok antara pengaruh dari luar dan perkembangan di dalam dirinya (*match*), tetapi tidak demikian menurut Vygotsky. Ada sesuatu di atas tahap perkembangan itu (*plus one matching*). Ada daerah-daerah yang sangat sensitif untuk diaktualisasikan dalam diri anak yang dinamakan Zone Proximal Development (ZPD) (Albert, Corea, & Macadino, 2012, p.14). Dengan menerapkan konsep ZPD pada pendidikan, maka pembelajaran akan memajukan perkembangan anak. Salah satu wujud konkret implikasi dari teori Vygotsky adalah dilaksanakannya akselerasi belajar bagi anak berbakat, pendidikan bagi anak-anak yang mempunyai kemampuan intelektual luar biasa, dalam proses pembelajaran harus selalu meningkatkan kadar mental atau berpikir tingkat tinggi.

Menurut taksonomi Bloom yang telah direvisi proses kognitif terbagi menjadi kemampuan berpikir tingkat rendah (*Lower Order Thinking*) dan kemampuan berpikir tingkat tinggi (*Higher Order Thinking*). Kemampuan yang termasuk LOT adalah kemampuan mengingat (*remember*), memahami (*understand*), dan menerapkan (*apply*), sedangkan

*HOT* meliputi kemampuan menganalisis (*analyze*), mengevaluasi (*evaluate*), dan menciptakan (*create*) (Anderson & Krathwohl, 2001, p.30). Taksonomi Bloom sudah lama diterapkan dalam bidang pendidikan dan sudah lama digunakan. Taksonomi Bloom masih digunakan dalam banyak kurikulum dan bahan pengajaran (Brookhart, 2010, p.39, Schraw and Robinson, 2011, pp.158-159). Dengan demikian, kemampuan berpikir tingkat tinggi fisika (*Physics Higher Order Thinking*) meliputi kemampuan fisika dalam menganalisis, mengevaluasi, dan menciptakan.

Menurut Brookhart (2010, p.5) kemampuan berpikir tingkat tinggi (HOTS) adalah (1) berpikir tingkat tinggi berada pada bagian atas taksonomi kognitif Bloom, (2) tujuan pengajaran di balik taksonomi kognitif yang dapat membekali peserta didik untuk melakukan transfer pengetahuan, (3) mampu berpikir artinya peserta didik mampu menerapkan pengetahuan dan keterampilan yang mereka kembangkan selama belajar pada konteks yang baru. Dalam hal ini yang dimaksud “baru” adalah aplikasi konsep yang belum terpikirkan sebelumnya oleh peserta didik, namun konsep tersebut sudah diajarkan, ini berarti belum tentu sesuatu yang universal baru. Berpikir tingkat tinggi berarti kemampuan peserta didik untuk menghubungkan pembelajaran dengan hal-hal lain yang belum pernah diajarkan.

Untuk memantau proses, kemajuan, dan perbaikan hasil belajar peserta didik secara berkesinambungan, diperlukan penilaian. Penilaian pendidikan adalah proses pengumpulan dan pengolahan informasi untuk menentukan pencapaian hasil belajar peserta didik (Peraturan Menteri Pendidikan Nasional No 20, 2007). Penilaian dapat dilakukan secara lisan ataupun tertulis. Penilaian secara tertulis dilakukan dengan tes tertulis. Secara garis besar ada dua bentuk soal tes tertulis, yaitu: memilih jawaban dan mensuplai jawaban. Soal tes tertulis yang jawabannya dengan memilih jawaban antara lain: pilihan ganda, dua pilihan (benar-salah, ya-tidak), menjodohkan, dan sebab-akibat.

Perlu diketahui bahwa model penilaian juga berpengaruh terhadap kemampuan berpikir siswa. Menurut Van den Berg (2008, p.15) bahwa kurikulum memiliki potensi yang kaya untuk mengembangkan keterampilan berpikir tingkat tinggi peserta didik. Guru harus merencanakan dengan baik dan melibatkan peserta didik dalam kegiatan pembelajaran yang dapat mendorong dan mengembangkan kemampuan berpikir tingkat tinggi tersebut. Penilaian dapat diimplementasikan untuk membantu siswa dalam meningkatkan kemampuan berpikir tingkat tinggi mereka. Hal ini didukung pendapat lain, bahwa pertanyaan berpikir tingkat tinggi dapat mendorong siswa untuk berpikir secara mendalam tentang materi pelajaran (Barnett & Francis (2012, p.209). Berdasarkan dua pendapat ini dapat disimpulkan bahwa tes kemampuan berpikir tingkat tinggi dapat memberikan rangsangan kepada siswa untuk mengembangkan kemampuan berpikir tingkat tinggi juga.

Nitko & Brookhart (2011, p.223) menjelaskan bahwa ketentuan dasar penilaian kemampuan berpikir tingkat tinggi adalah menggunakan tugas-tugas yang memerlukan penggunaan pengetahuan dan keterampilan dalam situasi baru. Untuk melakukan penilaian terhadap kemampuan *higher order thinking* harus menggunakan bahan-bahan baru. Salah satu cara yang dapat dilakukan adalah dengan menggunakan set-set *item* yang bergantung pada konteks.

Kenyataan bahwa tes pilihan ganda lebih banyak digunakan dari pada bentuk tes yang lain. Hal ini karena tes pilihan ganda memiliki kelebihan-kelebihan, antara lain: (1) materi yang diujikan dapat mencakup sebagian besar bahan pembelajaran, (2) jawaban siswa dapat dikoreksi dengan mudah dan cepat, (3) jawaban setiap pertanyaan sudah pasti benar atau salah, sehingga penilaian objektif (Sudjana, 1990, p.49). Walaupun ada juga kelemahan tes ini, yaitu: (1) kemungkinan peserta didik untuk melakukan tebakan jawaban masih cukup besar dan (2) proses berpikir siswa tidak dapat dilihat dengan nyata (Sudjana, 1990, p.49). Di samping itu, kelemahan lain tes objektif yakni:

(1) percaya diri yang tinggi pada testi dan (2) terjadinya kecurangan (*cheating*).

Untuk menghindari kecurangan (*cheating*), misalnya kerja sama dengan peserta didik lain, maka format (set) tes yang dikerjakan peserta didik yang berdekatan sebaiknya berbeda. Oleh karena itu, diperlukan minimal dua perangkat tes.

Berdasarkan hasil survei pendahuluan dengan cara melakukan wawancara dengan guru-guru fisika SMA di Daerah Istimewa Yogyakarta (DIY), diperoleh informasi bahwa sebagian besar di sekolah, baik pada tes tengah semester maupun tes akhir semester umumnya menggunakan tes pilihan ganda biasa. Jadi, tes pilihan ganda masih merupakan primadona dalam mendapatkan data prestasi belajar fisika siswa SMA. Kenyataan bahwa tes pilihan ganda yang digunakan di SMA untuk tes hasil belajar mata pelajaran fisika baru mengukur kemampuan mengingat, memahami, dan menerapkan. Jadi, tes pilihan ganda yang digunakan di SMA baru mengukur kemampuan berpikir tingkat rendah (*Lower Order Thinking (LOT)*) belum mengukur kemampuan berpikir tingkat tinggi Fisika (*PhysHOT*).

Penilaian dalam pendidikan menggunakan dua macam teori pengukuran, yakni: teori pengukuran klasik dan teori pengukuran modern. Teori Tes Klasik disebut juga dengan *Classical True-Score Theory*, dinamakan Teori Tes Klasik karena unsur-unsur teori ini sudah dikembangkan dan diaplikasikan sejak lama, namun tetap bertahan hingga sekarang (Suryabrata, 2002, p.21). Menurut teori pengukuran klasik penskoran hasil tes biasanya dilakukan secara parsial berdasarkan langkah-langkah yang harus ditempuh untuk menjawab benar suatu butir soal. Penskoran dilakukan per langkah dan skor per *item* peserta diperoleh dengan menjumlah skor peserta didik tiap langkah, dan kemampuan diestimasi dengan skor mentah. Model penskoran seperti ini belum tentu tepat, karena tingkat kesulitan tiap langkah tidak diperhitungkan.

Penilaian hasil ujian peserta didik didasarkan pada tahap-tahap yang dapat di-

selesaikan peserta didik. Walaupun hanya menyelesaikan tahap awal saja, peserta ujian sudah mendapatkan nilai. Nilai tertinggi tentu saja didapatkan ketika peserta ujian telah menyelesaikan semua tahapan soal ujian dalam butir tersebut. Prosedur penilaian tersebut sebenarnya sama dengan bagaimana individu merespon butir dalam skala psikologi. Misalnya, sebuah butir yang menyediakan empat kategori respons dari '*tidak pernah*', '*jarang*', '*sering*', dan '*selalu*' analog dengan tahap penyelesaian. Menyelesaikan soal hanya sampai tahap pertama analog dengan kategori '*tidak pernah*' sedangkan kalau sudah sampai tahap akhir, analog dengan kategori '*selalu*'. Asumsi ini kemudian dikembangkan menjadi *partial credit model (PCM)*. Ketika diasumsikan bahwa sebuah *item* mengikuti pola kredit parsial maka kemampuan individu lebih tinggi diharapkan memiliki skor yang lebih tinggi daripada individu yang memiliki kemampuan rendah (Widhiarso, 2010, p.6). Menurut Wright & Masters, PCM juga sesuai untuk menganalisis respon pada pengukuran berpikir kritis dan pemahaman konseptual dalam sains (Van der Linden & Hambleton, 1997, pp. 101-102).

PCM dikembangkan untuk menganalisis *item* tes yang memerlukan beberapa langkah penyelesaian. PCM dapat diberikan pada langkah-langkah yang dapat dikerjakan oleh individu. Dengan demikian, PCM cocok untuk dikenakan pada tes prestasi, termasuk soal fisika yang membutuhkan tahap identifikasi permasalahan hingga solusi akhir.

PCM merupakan pengembangan dari Model IRT 1 *parameter of logistic (1-PL)* dan termasuk keluarga Model Rasch. Model dikotomis sederhana dalam Model Rasch merupakan kasus khusus dari PCM. Model dikotomis dan PCM dapat dikatakan campuran dalam satu analisis (Wu & Adams, 2007). PCM merupakan pengembangan dari Model Rasch *item* dikotomis yang diterapkan pada *item* politomis. Model Rasch *item* dikotomis yang hanya berisi satu parameter lokasi *item* (tingkat kesulitan) kemudian dikembangkan dengan menjabarkan lokasi

butir menjadi beberapa kategori. Asumsi pada PCM yakni setiap butir mempunyai daya beda yang sama. PCM mempunyai kemiripan dengan *Graded Response Model* (GRM) pada *item* yang diskor dalam kategori berjenjang, namun indeks kesukaran dalam setiap langkah tidak perlu urut, suatu langkah dapat lebih sukar dibandingkan langkah berikutnya. Ini berarti, PCM merupakan pengembangan Model Rasch dikotomus menjadi politomus dengan satu parameter logistik yaitu tingkat kesulitan.

Skor kategori pada PCM menunjukkan banyaknya langkah untuk menyelesaikan dengan benar butir tersebut. Skor kategori yang lebih tinggi menunjukkan kemampuan yang lebih besar daripada skor kategori yang lebih rendah. Pada PCM, jika suatu butir memiliki dua kategori, maka Persamaan probabilitas menjadi persamaan model Rasch, seperti persamaan yang dinyatakan oleh Hambleton and Swaminathan (1985). Oleh karena itu, PCM dapat diterapkan pada butir politomus dan dikotomus.

Hasil survei pendahuluan menunjukkan bahwa penskoran hasil tes pilihan ganda menggunakan model dikotomus, artinya jika *item* benar diberi skor 1 dan jika salah diberi skor 0. Penskoran belum menggunakan model politomus yang lebih adil karena mempertimbangkan langkah-langkah penyelesaian tes. Model penskoran dikotomus ini belum menghargai tahap-tahap penyelesaian soal, karena dengan tingkat kesalahan yang berbeda mendapatkan skor yang sama yakni 0. Dengan demikian skoring model ini tentu kurang adil.

Tujuan tes untuk mengetahui kemampuan peserta didik dan posisinya dalam kelompok tersebut. Jika tes yang ditempuh peserta didik tidak semuanya sama, tes terdiri atas dua atau lebih perangkat, maka hasil tes tidak dapat dibandingkan antarpeserta didik. Agar hasil tes dapat dibandingkan untuk perangkat tes yang terdiri atas lebih dari satu perangkat, maka perlu ada *anchor item* untuk penyetaraan (*equating*) hasil tes.

Hambleton & Swaminathan (1991, pp.123-143) menyatakan penyetaraan skor tes atau *equating* adalah tindakan mengkon-

versi skor tes yang satu (skor tes X) menjadi skor yang metrik (yang sesuai dengan ukuran) dari tes yang lain (skor tes Y). Menurut Kolen & Brannen (1995, p.2) penyetaraan (*equating*) skor tes adalah proses statistik yang digunakan untuk melakukan penyesuaian skor antara suatu tes dengan tes yang lain. Skor tes yang sudah disesuaikan melalui *equating* bersifat *interchangeable* sehingga dapat diambil keputusan yang adil yang didasarkan pada paket tes yang berbeda. Dengan demikian, berarti agar hasil tes testi yang mengerjakan perangkat tes yang berbeda dari dua atau lebih perangkat tes dapat dibandingkan, maka harus dilakukan penyetaraan tes.

Berdasarkan uraian tersebut, untuk mengukur kemampuan berpikir tingkat tinggi fisika digunakan tes berbentuk pilihan ganda beralasan yang dinamakan Tes kemampuan berpikir tingkat tinggi fisika (*Physics Test for Higher Order Thinking Skills* (*PhysTHOTS*)). Untuk itu perlu disusun instrumen penilaian kemampuan berpikir tingkat tinggi Fisika (*PhysTHOTS*) yang terdiri atas tes dan pedoman penilaian. Dengan demikian, diperlukan adanya penelitian pengembangan instrumen penilaian kemampuan berpikir tingkat tinggi fisika yang terdiri atas tes berpikir tingkat tinggi Fisika (*Physics Test for Higher Order Thinking Skills* (*PhysTHOTS*)) dan pedoman penilaian.

Sejalan dengan permasalahan tersebut, diperlakukan penelitian untuk (1) menghasilkan instrumen untuk mengukur kemampuan berpikir tingkat tinggi dalam mata pelajaran fisika, dan (2) mendapatkan karakteristik instrumen penilaian kemampuan berpikir tingkat tinggi dalam mata pelajaran fisika di SMA kelas XI yang meliputi kemampuan menganalisis, mengevaluasi, dan menciptakan

## Metode Penelitian

Penelitian ini termasuk jenis penelitian pengembangan dengan pendekatan kuantitatif. Penelitian pengembangan instrumen ini menggunakan model modifikasi Model Wilson dan Model Oriondo dan Antonio.

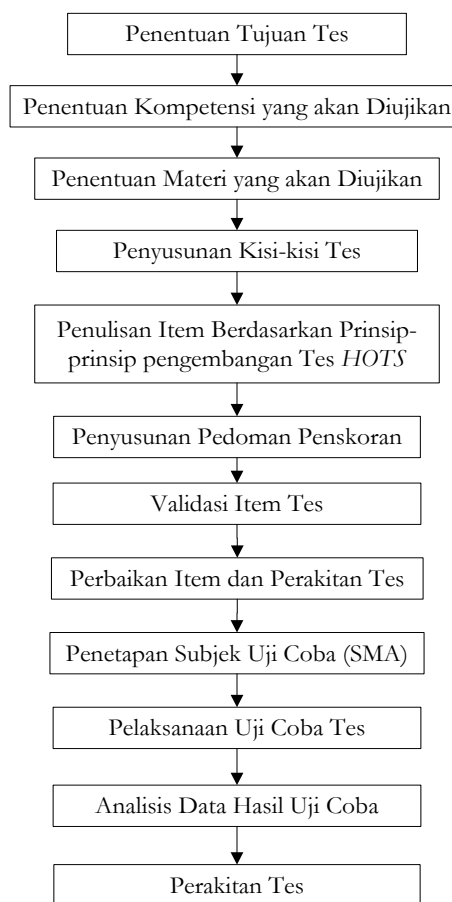
Penelitian dimulai Desember 2012 sampai dengan Mei 2013. Pengembangan awal tes yang berupa penyiapan tes, validasi, dan perakitan tes dilakukan pada bulan Desember 2012 sampai dengan Februari 2013. Uji coba dilangsungkan pada bulan Maret sampai dengan April 2013. Penelitian dilakukan di SMA di wilayah Daerah Istimewa Yogyakarta.

Subjek penelitian ini seluruh peserta didik kelas XI dari sepuluh SMAN di Daerah Istimewa Yogyakarta yang terdiri atas dua SMA dari setiap kota/kabupaten. Jumlah subjek penelitian sebanyak 1001 peserta didik.

Langkah-langkah pengembangan instrumen berupa tes menggunakan modifikasi Model Wilson dan Model Oriondo dan Antonio, yakni: (1) perancangan tes, (2) uji coba tes, dan (3) perakitan tes.

Tahap perancangan tes meliputi: (1) penentuan tujuan tes, (2) penentuan kompetensi yang diujikan, (3) penentuan materi yang diujikan, (4) penyusunan kisi-kisi tes, (5) penulisan *item* berdasarkan prinsip-prinsip pengembangan Tes *HOT*, (6) validasi *item* tes, (7) perbaikan *item* dan perakitan tes, dan (8) penyusunan pedoman penskoran. Adapun tahap uji coba tes meliputi: (1) penetapan subjek uji coba (SMA), (2) pelaksanaan uji coba, dan (3) analisis data hasil uji coba. Tahap-tahap pengembangan tes tersebut disajikan pada Gambar 1.

Berkaitan ukuran sampel, untuk analisis secara *IRT* beberapa ahli pengukuran sebaiknya 200 sampai dengan 1000 orang (Seon, 2009, p.3). Untuk analisis dengan Rasch, sampel yang digunakan antara 30 sampai 300 orang (Bond and Fox, 2007, p. 43; Keeves & Masters, 1999, pp.12-13). Reckase (2000) menyimpulkan bahwa sampel ukuran minimum yang baik untuk memperkirakan tiga parameter yang meliputi: diskriminasi (daya beda), tingkat kesulitan, dan *pseudoguessing* adalah 300 (Haladyna, 2004, p.206). Jadi, dengan model PCM 1PL peserta didik yang dijadikan subjek coba sebanyak 500 orang sudah lebih dari cukup.



Gambar 1. Langkah-langkah Pengembangan Instrumen

Dalam kegiatan uji coba ini terlebih dahulu ditentukan SMA di Provinsi DIY yang digunakan untuk uji coba berdasar ranking sekolah berdasar nilai UN Fisika Tahun 2012. Setiap kota/kabupaten dipilih dua SMA yang berada pada dua peringkat UN rendah sedang, atau tinggi. SMA yang digunakan untuk uji coba, antara lain: SMA 5 Yogyakarta, SMA 11 Yogyakarta, SMA 1 Bambanglipuro, SMA 1 Sedayu, SMA 1 Wates, SMA 1 Pengasih, SMA 1 Gamping, SMA 1 Minggir, SMA 1 Wonosari, dan SMA 1 Patuk.

Analisis data penelitian ini menggunakan *Partial Credit Model* 1 PL (PCM 1-PL) untuk pengujian *fit item* tes kemampuan berpikir tingkat tinggi untuk mata pelajaran Fisika SMA. Dasar pertimbangan yang digunakan, yang pertama bahwa PCM sebagai perluasan Rasch Model yang merupakan model 1-PL, dapat menggunakan sampel yang tidak sebesar kalau melakukan

kalibrasi data politomus menggunakan model 2-PL atau 3-PL (Keeves & Masters, 1999, pp.12-13). Kedua, bahwa karakteristik respons terhadap *item* kemampuan berpikir tingkat tinggi mengikuti PCM.

Analisis data dilakukan pada beberapa aspek, yakni: (1) kecocokan *item* instrumen, (2) reliabilitas, (3) kurva karakteristik *item* (ICC), (4) indeks kesukaran, dan (5) fungsi informasi dan SEM.

Pengujian *goodness of fit* untuk tes secara keseluruhan dan testi (*case/person*) secara keseluruhan dikembangkan Adam dan Khoo (1996, p.30) berdasarkan nilai rerata INFIT Mean of Square (*Mean INFITMNSQ*) beserta simpangan bakunya atau mengamati nilai rata INFIT *t* (*Mean INFIT t*) beserta simpangan bakunya. Jika rerata INFIT MNSQ sekitar 1,0 dan simpangan bakunya 0,0 atau rerata INFIT *t* mendekati 0,0 dan simpangan bakunya 1,0, maka keseluruhan tes *fit* dengan model.

Kecocokan butir dengan model diketahui dengan *fit item* dan testi mengikuti kaidah bahwa *Item characteristic curve* (ICC) akan mendatar (*flat*) bila besarnya INFIT MNSQ untuk *item* atau *e* lebih besar dari satuan logit >1,30 atau <0,77. Keadaan ini grafik distribusi membentuk *platokurtic curve*

dan tidak lagi membentuk *leptokurtic curve* (Keeves & Alagumalai 1999, p.36). Oleh karena itu, suatu *item* atau testi/*case/person* dinyatakan *fit* dengan model dengan batas kisaran INFIT MNSQ dari 0,77 sampai 1,30 (Adam & Khoo, 1996, pp.30&90). Dalam hal ini menggunakan kisaran nilai *t* adalah  $\pm 2$  (pembulatan  $\pm 1,96$ ) jika taraf kesalahan atau *alpha* sebesar 5% (Keeves & Alagumalai 1999, pp.34-36; Bond & Fox, 2007, p.43).

*Item* dikatakan baik jika indeks kesukaran lebih dari -2.0 atau kurang dari 2.0 (Hambleton & Swaminathan, 1985, p.36). Berdasarkan fungsi informasi dan SEM, maka dapat diketahui bahwa tes ini cocok untuk siswa dengan kemampuan ( $\theta$ ) rendah, sedang, atau tinggi.

## Hasil Penelitian dan Pembahasan

### Hasil Pengembangan Tes

Instrumen PhystHOTS terdiri atas dua perangkat (set) yang masing-masing meliputi materi: gerak, gaya, usaha dan energi, serta momentum dan impuls dari aspek menganalisis, mengevaluasi, dan menciptakan. Selanjutnya, dengan *expert judgment* PhystHOTS dinyatakan layak digunakan.

Tabel 1. Sebaran Iten Tes Kemampuan Berpikir Tingkat Tinggi Fisika Kelas XI SMA

Aspek	Subaspek	Materi Fisika Kelas XI SMA			
		Gerak	Gaya	Usaha dan Energi	Momentum dan Impuls
Meng-analisis	Membedakan	2 (1A, 9B)	2 (4A, 12B)		1 (21A, 3B)*
	Mengurutkan	1 (19A, 1B)*	2 (5A, 13B)	2 (7A, 15B)	2 (9A, 17B)
	Memberikan ciri khusus	2 (2A, 10B)	2 (6A, 14B)	1 (20A, 2B)*	
Meng-evaluasi	Mengecek		2 (3A, 11B)	2 (8A, 16B)	1 (26A, 8B *)
	Mengkritik	2 (11A, 19B)		1 (25A, 7B)*	2 (10A, 18B)
Men-ciptakan	Memunculkan Ide	1 (22A, 4B)*	4 (13A, 14A, 12B, 22B)	2 (15A, 23B)	
	Merencanakan	2 (12A, 20B)	1 (23A, 5B)*		2 (18A, 26B)
	Menghasilkan		1 (24A, 6B)*	2 (16A, 24B)	2 (17A, 25B)

Keterangan: \*) *anchor item*

PhystHOTS terdiri atas dua set yakni set I berkode A dan set II berkode B. Setiap tes meliputi materi: gerak analisis vektor, gaya dan getaran, usaha energi, momentum dan implus yang meliputi aspek dan sub

aspek menganalisis, mengevaluasi, dan menciptakan dengan sebaran sebagaimana dinyatakan Tabel 1. Kedua perangkat tes tersebut memiliki delapan *item* sebagai *anchor item*.

Hasil Uji Coba

*Kecocokan Item Instrumen (goodness fit)*

Pengujian *goodness of fit* dilakukan untuk tes secara keseluruhan ataupun tiap *item*. Pengujian *fit* tes secara keseluruhan menggunakan kaidah yang dikembangkan oleh Adam dan Khoo (1996, p.30) yakni berdasarkan nilai rerata INFIT *Mean of Square (Mean INFITMNSQ)* beserta simpangan bakunya atau mengamati nilai rata INFIT *t (Mean INFIT t)* beserta simpangan bakunya. Jika rerata INFIT MNSQ sekitar 1,0 dan simpangan bakunya 0,0 atau rerata INFIT *t* mendekati 0,0 dan simpangan bakunya 1,0, maka keseluruhan tes fit dengan model PCM 1 PL. Berdsarkan Tabel 2, nilai rerata *INFITMNSQ* 1,01 (sekitar 1) dan simpangan baku 0,02 (sekitar 0,0), maka keseluruhan tes *fit* dengan model PCM 1 PL

Tabel 2. Hasil Estimasi *Item* dan Testi HOTS Fisika menurut PCM 1-PL

No	Uraian	Estimasi untuk <i>item</i>	Estimasi untuk testi
1	Nilai rata-rata dan simpangan baku	$-0,29 \pm 0,51$	$0,19 \pm 0,01$
2	Nilai rata-rata dan simpangan baku yang sudah disesuaikan	$0,00 \pm 0,50$	$-0,11 \pm 0,00$
3	Reliabilitas	0,95	
4	Nilai rata-rata dan simpangan baku INFIT MNSQ	$1,01 \pm 0,02$	$1,01 \pm 0,07$
5	Nilai rata-rata dan simpangan baku OUTFIT MNSQ	$1,01 \pm 0,02$	$1,01 \pm 0,12$
6	Nilai rata-rata dan simpangan baku INFIT <i>t</i>	$0,84 \pm 1,47$	$0,25 \pm 1,08$
7	Nilai rata-rata dan simpangan baku OUTFIT <i>t</i>	$0,27 \pm 0,48$	$0,08 \pm 0,61$

Pengujian penetapan *fit* setiap *item* pada model mengikuti kaidah Adam dan Khoo (1996, p.30), yakni suatu *item fit* pada model jika nilai INFIT MNSQ antara 0,77 sampai 1,30. Dengan batas penerimaan *item* menggunakan INFIT MNSQ atau *fit* menurut model (antara 0,77 sampai dengan 1,30) dan menggunakan INFIT *t* dengan

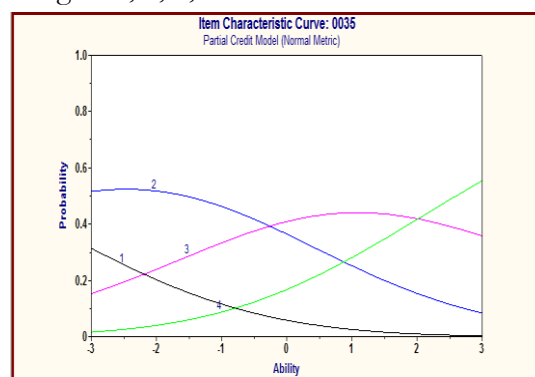
batas -2,0 sampai 2,0, maka diperoleh *item-item* yang cocok memenuhi *goodness of fit*. Nilai INFIT MNSQ PhysTHOTS antara 0,98 sampai dengan 1,05. Dengan batas penerimaan *item* menggunakan INFIT MNSQ atau *fit* menurut model (antara 0,77 sampai dengan 1,30), maka semua *item* sebanyak 44 *item* fit dengan PCM semua.

*Reliabilitas*

Berdasarkan hasil analisis diperoleh reliabilitas set instrumen (PhysTHOTS) tersebut sebesar 0,95. Nilai reliabilitas ini tergolong kategori tinggi.

*Kurva Karakteristik Item (Item Characteristic Curva, ICC)*

Karakteristik *item* ditunjukkan dengan kurva karakteritik *item* (ICC) dan indeks kesukaran. Berdasarkan analisis diperoleh kurva karakteristik *item* (ICC) sebanyak 44 buah. Pada Gambar 2 disajikan contoh ICC untuk *item* 35 (*item* 17 B), yang dapat dijelaskan bahwa: (a) skor 1 (kategori 1) sebagian besar diperoleh siswa dengan *ability* sangat rendah ( $\theta = -3$ ), (b) skor 2 (kategori 2) sebagian besar diperoleh siswa dengan *ability* rendah ( $\theta = -2$ ), (c) skor 3 (kategori 3) sebagian besar diperoleh siswa dengan abilitas tinggi ( $\theta = 1,5$ ), dan (d) skor 4 (kategori 4) sebagian besar diperoleh siswa dengan *ability* sangat tinggi ( $\theta = 3$ ). Tingkat kesulitan dari kecil ke besar berurutan kategori 1, 2, 3, dan 4.



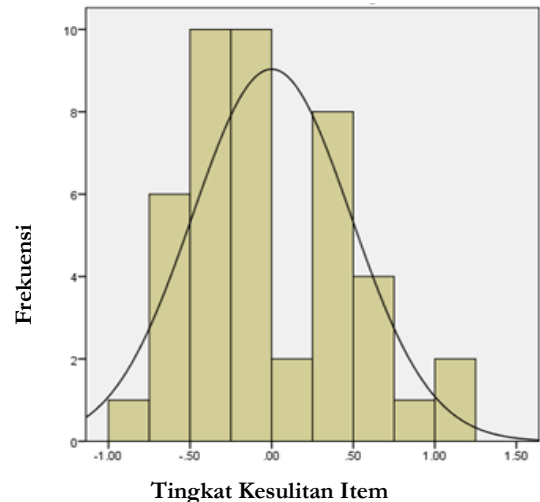
Gambar 2. Kurva Karakteristik Butir 35 (Butir 17B)

*Indeks Kesukaran (b)*

Indeks kesukaran atau tingkat kesukaran (*b*) untuk skor 0 ( $b_0$ ), skor 1 ( $b_1$ ), skor 2 ( $b_2$ ), skor 3 ( $b_3$ ), skor 4 ( $b_4$ ), dan reratanya

sebagai *difficulty*. Berdasarkan hasil analisis, *difficulty item-item* terletak antara -0,86 sampai dengan 1,06 dengan rata-rata 0 dan simpangan baku 0,42. *Item* dikatakan baik jika indeks kesukaran lebih dari -2,0 atau kurang dari 2,0 ( $-2,0 < b < 2,0$ ) (Hambleton & Swaminathan, 1985, p.36). Jadi, berdasarkan *difficulty*, semua *item* sebanyak 44 semuanya baik.

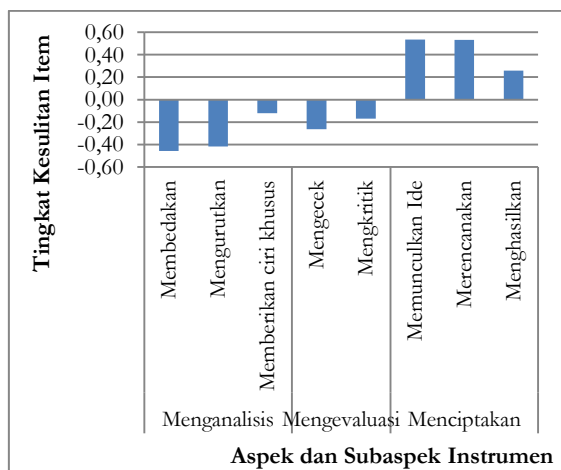
Adapun histogram distribusi indeks kesulitan dapat dilihat pada Gambar 3. Berdasarkan Gambar 3 distribusi indeks kesulitan mengikuti distribusi mendekati normal. Hal ini tidak menjadikan masalah, karena tidak ada persyaratan bahwa distribusi indeks kesulitan *item* harus normal.



Gambar 3. Distribusi Tingkat Kesulitan *Item* Tes

Tabel 3. Tingkat Kesulitan Butir setiap Subaspek

No	Aspek	Subaspek	<i>Difficulty</i>	Kesulitan Tahap			
				Kategori 1	Kategori 2	Kategori 3	Kategori 4
1	Meng-analisis	Membedakan	-0,46	-2,96	0,45	0,92	-0,16
		Mengurutkan	-0,42	-3,09	0,32	1,38	-0,21
		Memberikan ciri khusus	-0,12	-2,69	0,95	1,41	0,00
2	Meng-evaluasi	Mengecek	-0,26	-2,00	1,25	0,37	-0,31
		Meng kritik	-0,17	-2,36	1,68	0,19	-0,14
3	Men-ciptakan	Memunculkan Ide	0,54	0,99	0,32	0,72	-0,23
		Merencana kan	0,53	1,71	-0,28	0,72	-0,16
		Menghasil kan	0,26	-1,10	1,84	0,31	-0,22



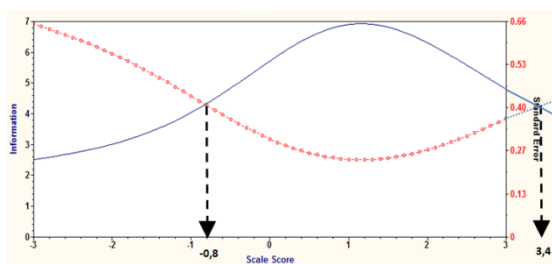
Gambar 4. Tingkat Kesulitan *Item* masing-masing Aspek dan Subaspek Instrumen

Tabel 3 menjelaskan besar tingkat kesulitan pada setiap subaspek dan aspek instrumen untuk masing-masing kategori da-

lam PCM. Berdasarkan Tabel 3 juga dapat diketahui tingkat kesulitan pada setiap subaspek dan aspek instrumen untuk setiap kategori dalam PCM. Untuk lebih jelasnya, dapat dilihat diagram distribusi tingkat kesulitan *item* menurut aspek dan subaspek instrumen yang disajikan pada Gambar 4.

Berdasarkan Gambar 4, dapat dilihat bahwa urutan tingkat kesulitan *item* setiap aspek pada tahap ujicoba berturut-turut adalah menganalisis, mengevaluasi, dan menciptakan. Hal ini berarti bahwa aspek kemampuan secara gradasi dari rendah ke tinggi secara berturut-turut adalah aspek kemampuan menganalisis, mengevaluasi, dan menciptakan. Kondisi ini sudah sesuai dengan pendapat Anderson & Krathwohl (2001, p.30).

Fungsi Informasi dan SEM



Gambar 5. Fungsi Informasi dan SEM

Hasil analisis diperoleh fungsi informasi dan *standard error measurement* (SEM). Berdasarkan fungsi informasi dan SEM yang dinyatakan Gambar 5, maka tes ini cocok untuk peserta didik dengan kemampuan berpikir tingkat tinggi Fisika ( $\theta$ ) dalam kategori tinggi, yakni  $-0,8 \leq \theta \leq 3,4$ .

Revisi Instrumen

Berdasarkan hasil uji coba, semua *item* adalah *fit*. Dua *anchor item* yang tingkat kesulitannya lebih dari satu, yakni *item* ke-22 (1,03) dan ke-23 (1,06) ditukar dengan *item* ke-13 (0,31) dan ke-12 (0,38). Penukaran *item* didasarkan pada kesesuaian sub aspek dan sub materi fisika. Instrumen *PhysTHOTS* terdiri atas dua set tes A dan B yang masing berisi 26 *item* dengan 8 *anchor item*.

Pembahasan

Reliabilitas *PhysTHOTS* cukup tinggi yakni 0,95 yang menegaskan bahwa hasil pengukuran dengan instrumen ini reliabel. Lebih dari itu, tes yang memiliki koefisien reliabilitas sekurang-kurangnya 0,90 hasil testing dengan tes tersebut dapat digunakan untuk membuat keputusan tentang individu (Suryabrata, 2002, pp.39-40). Selain itu, fungsi informasi relatif tinggi untuk kemampuan antara -0,8 sampai 3,4. Ini berarti bahwa instrumen ini memiliki kekuatan dan reliabel yang tinggi karena tersusun dari *item-item* yang memiliki fungsi informasi yang tinggi (Hambleton dan Swaminathan, 1985, p.94). Hal ini dapat terjadi karena tes ini sesuai dengan kemampuan peserta didik yang dites. Berdasarkan koefisien reliabilitas, fungsi informasi tes, dan parameter estimasi, berarti *PhysTHOTS* ini reliabel dan memiliki kestabilan yang tinggi.

Validitas isi tes telah dibuktikan dengan *expert judgment*. Validitas konstruk secara empiris dibuktikan dengan *goodness of fit* pada *partial credit model* (PCM). Berdasarkan Tabel 2 nilai rata-rata dan simpangan baku INFIT MNSQ yakni masing-masing 1,01 (sekitar 1) dan 0,02 (sekitar 0), maka tes *fit* dengan PCM 1 PL. Hal ini berarti bahwa secara empiris *PhysTHOTS* tersebut valid. Tes yang berisi pertanyaan berpikir tingkat tinggi yang valid mendorong peserta didik untuk berpikir secara mendalam tentang materi pelajaran (Barnett & Francis, 2012, p.209). Validitas tes tersebut ini didukung oleh semua *item* memiliki nilai INFIT MNSQ antara 0,98 sampai dengan 1,05 yang terletak antara batas penerimaan *item* menggunakan INFIT MNSQ atau *fit* menurut model (antara 0,77 sampai dengan 1,30) berarti semua *item* sebanyak 44 *item fit* semua. Hal ini disebabkan karena beberapa hal, antara lain: (1) *item-item* dikembangkan sesuai prosedur pengembangan *item* instrumen yang benar, (2) *item-item* dikembangkan dari indikator yang diturunkan dari aspek kemampuan berpikir tingkat tinggi dan materi fisika, (3) tes yang terdiri atas 44 *item* telah melalui uji validitas isi dengan *expert judgment*, dan (4) responden (peserta didik) yang diuji sungguh-sungguh dalam mengerjakan karena melibatkan pengawas dari guru fisika mereka. Dengan demikian, karena *PhysTHOTS* yang dikembangkan valid akan mengukur kemampuan berpikir tingkat tinggi fisika dengan hasil yang valid pula serta mendorong peserta didik untuk berpikir secara mendalam tentang materi pelajaran fisika.

Menurut Hambleton & Swaminathan (1985, p.36), tingkat kesulitan *b* untuk *item* yang baik bervariasi antara -2,00 sampai dengan 2,00. *Item* dengan tingkat kesulitan -2,00 menandakan *item* tersebut sangat mudah, sedangkan tingkat kesulitan 2,00 berarti *item* tersebut sangat sulit. Dengan demikian, ditinjau dari tingkat kesulitan *item* dan kestabilannya maka instrumen ini termasuk kategori baik.

Berdasarkan fungsi informasi dan SEM, *PhysTHOTS* akan tepat digunakan

pada peserta didik yang memiliki kemampuan dari -0,8 sampai dengan 3,4 (termasuk dalam level tinggi). Dengan demikian, instrumen ini diyakini dapat digunakan untuk menggali informasi mengenai kemampuan berpikir tingkat tinggi fisika dengan tepat pada peserta didik SMA yang memiliki abilitas tinggi. Hal ini karena peserta didik yang berabilitas tinggi memiliki pemahaman konsep yang mendalam sehingga lebih mungkin dapat menerapkan pengetahuan itu untuk memecahkan masalah baru. (Ramos, Dolipas, & Villamor, 2013, p.57).

## Simpulan dan Saran

### Simpulan

Berdasarkan uraian tersebut, simpulan yang dapat ditarik adalah sebagai berikut: (1) instrumen *PhysTHOTS* dikembangkan dalam bentuk pilihan ganda ber-alasan pada kemampuan menganalisis, mengevaluasi, dan menciptakan untuk materi fisika gerak, gaya, usaha dan energi, serta momentum dan impuls yang terdiri atas perangkat tes A dan perangkat tes B yang masing-masing terdapat 26 *item* dengan 8 *anchor item*; (2) instrumen *PhysTHOTS* telah memenuhi validitas isi dengan *expert judgment* dan telah mendapatkan bukti empiris validitas konstruk *fit* pada *Partial Credit Model (PCM)* berdasarkan data politomus empat kategori; (3) seluruh *item* pada *PhysTHOTS* dalam kriteria baik karena tingkat kesulitannya berada pada rentang antara -2,00 sampai dengan 2,00.

Reliabilitas *PhysTHOTS* telah memenuhi syarat, bahkan termasuk tinggi (koefisien reliabilitas lebih dari 0,90).

Berdasarkan fungsi informasi, *PhysTHOTS* sangat tepat digunakan untuk mengukur kemampuan berpikir tingkat tinggi fisika peserta didik yang berkemampuan dari -0,80 sampai 3,40.

### Saran

Berdasarkan hasil analisis, disarankan: (1) agar para guru menerapkan tes kemampuan berpikir tingkat tinggi fisika di SMA; (2) diadakan pelatihan penyusunan tes kemampuan berpikir tingkat tinggi fisika bagi guru; (3) dilakukan penelitian

lebih lanjut dengan menggunakan analisis data politomus menurut *generalized partial credit model (GPCM 3PL)*.

## Daftar Pustaka

- Adams, R. J. & Khoo, S. T. (1996). *Quest: The interactive test analysis system version 2.1*. Victoria: The Australian Council for Educational Research.
- Albert, L.R, Corea, D & Macadino, V. (2012). *Rhetorical Ways of Thinking Vygotskian Theory and Mathematical Learning*. New York: Springer
- Anderson, L.W., and Krathwohl, D.R. (2001). *A Taxonomy of Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Barnett, J. E and Francis, A.L. (2012). *Using higher order thinking questions to foster critical thinking: a classroom study*. Educational Psychology: An International Journal of Experimental Educational Psychology. <http://www.tandfonline.com/loi/cedp20>. Diakses tanggal 10 Desember 2012
- Bloom, B.S., et al. (1979). *Taxonomy of Educational Objectives: Handbook I Cognitive Domain*. London: Longmans Group Ltd.
- Bond, TG and Fox, CM. (2007). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences (2<sup>rd</sup> edition)*. Mahwah, NJ: Lawrence Erlbaum. Bonk.
- Brookhart, S. M. (2010). *How to Assess Higher Order Thinking Skills in Your Classroom*. Alexandria: ASCD
- BSNP. (2006). *Standar Kompetensi dan Kompetensi Dasar Mata Pelajaran Fisika Untuk SMA dan MA*. Jakarta: BSNP-Depdiknas
- Depdiknas. (2007). *Peraturan Menteri Pendidikan Nasional No 20 Tahun 2007 tentang Standar Penilaian*
- Dikpora DIY. (2012). *Hasil Nilai UN 2012 SMA/MA/SMK di DIY (27 Mei 2012)*. Yogyakarta: Dikpora DIY [http://www.pendidikan-diy.go.id/?view=v\\_berita&id\\_sub=2692](http://www.pendidikan-diy.go.id/?view=v_berita&id_sub=2692) diakses tanggal 1 Juni 2012.

- Efendi, Ridwan. (2010). *Kemampuan Fisika Siswa Indonesia dalam TIMSS*. Prosi-  
ding Seminar Nasional Fisika 2010  
ISBN: 978-979-98010-6-7
- Haladyna, T. M. (2004). *Devoping and Validating Multiple Choise Test Items*. New  
Jersey: Lawrence Erlbaum Associates,  
Inc.
- Hambleton & Swaminathan. (1985). *Item Res-  
ponse Theory Principles and Applications*.  
Boston: Kluwer Nijhoff Publishing.
- Hambleton dan Swaminathan (1991): *Fun-  
damentals of Item Response Theory*. Cali-  
fornia: SAGE Publications, Inc
- Keeves, J. P. & Alagumalai. (1999). New  
Approach to measurement. Dalam:  
Masters, G.N. & Keeves, J.P. (eds.).  
*Advances in Measurement in Educational  
Research and Assessment* (pp.23-42).  
Amsterdam: Pergamon, An imprint  
of Elsevier Science
- Keeves, J.P. & Masters, G.N. (1999). Intro-  
duction. Dalam: Masters, G.N. &  
Keeves, J.P. (eds.). *Advances in Mea-  
surement in Educational Research and  
Assessment* (pp.1-22). Amsterdam: Per-  
gamon, An imprint of Elsevier Science
- Kolen, M.J. & Brannen, R.L. (1995). *Test  
equating: Methods and practices*. New  
York: Sprinegr-Velag New York, Inc.
- Nitko, A.J & Brookhart, S. M. (2011).  
*Educational assessment of students*. (6<sup>th</sup>  
ed). Boston: Pearson Education, Inc.
- Oriondo, L.L.and Dallo-Antonio, E.M.  
(1998). *Evaluation Educational Outcomes*.  
Manila: Rex Printing Compagny, inc
- Piaget, J. (2005). *The psychology of intellegence*  
[Versi elektronik]. Taylor & Francis e-  
Library.
- Pusat Kurikulum. (2007). *Naskah Akademik  
Kajian Kebijakan Kurikulum Mata Pel-  
ajaran IPA*. Jakarta: Balitbang Depdik-  
nas
- Ramos, J. L. S., Dolipas, B. B., & Villamor  
B. B. (2013). Higher Order Thinking  
Skills and Academic Performance in  
Physics of College Students: A Re-  
gression Analysis. *International Journal  
of Innovative Interdisciplinary Research*,  
ISSN 1839-9053l , p: 48-60. Diambil  
tanggal 20 Desember 2013 dari  
[http://www.auamii.com/jiir/Vol-  
01/issue-04/5ramos.pdf](http://www.auamii.com/jiir/Vol-01/issue-04/5ramos.pdf)
- Reedal, K.E. (2010). *Jean Piaget's Cognitive  
Development Theory in Mathematics Edu-  
cation*. Department of Mathematics  
and Computer Science – Ripon Colle-  
ge. Summation, May 2010, pp. 16-20  
<http://ripon.edu/mac/s/summation>.
- Schraw, G, and Robinson, D.H. (2011).  
*Assessment of Higher Order Thinking  
Skills*. New York: Information Age  
Publishing, Inc
- Seon, Hi Sin. (2009). How to tread omitted  
respons in Rasch model based equa-  
ting [Versi elektronik] *Practical Assess-  
ment, Research & Evaluation*. ISSN 1531-  
7714, Volume 14, Number 1, p: 1-8
- Sudjana, Nana. (1990). *Penilaian Hasil Blajar  
Mengajar*. Bandung: PT Remaja Ros-  
dakarya
- Suryabrata, S. (2002). *Pengembangan alat ukur  
psikologis*. Yogyakarta: Andi Offset.
- TIMSS & PIRLS International Study Cen-  
ter. (2012). *TIMSS 2011 international  
results in science*. Boston: The TIMSS &  
PIRLS International Study Center,  
Boston College. Diambil tanggal 5  
Januari 2013, dari [http://timss.bc.edu/  
timss2011/release.html](http://timss.bc.edu/timss2011/release.html)
- Van den Berg, G. 2008. The use of assess-  
ment in the development of higher-  
order thinking skills. *Africa Education  
Review*, 1:2, 279-294. Diambil Tanggal  
24 Oktober 2012 dari [http://dx.doi.  
org/10.1080/18146620408566285](http://dx.doi.org/10.1080/18146620408566285).
- Van der Linden, W. J & Hambleton, R. K.  
(1997). *Handbook of Modern Item Res-  
ponse Theory*. New York: Springer-  
Verlag New York, Inc
- Widhiarso, Wahyu. (2010). *Model politomi  
dalam teori respons butir*. Yogyakarta:  
Psikologi UGM
- Wright, B.D. & Masters, G.N. (1982).  
*Rating scale analysis*. Chicago: Mesa  
Press.
- Wu, M., & Adams, R. (2007). *Applying the  
Rasch model to psychosocial measurement:  
A practical approach*. Melbourne: Edu-  
cational Measurement Solutions.

## PENGEMBANGAN ASESMEN HASIL BELAJAR PENJASORKES SISWA SMA PADA PERMAINAN BOLAVOLI

<sup>1)</sup>Guntur, <sup>2)</sup>Sukadiyanto, <sup>3)</sup>Djemari Mardapi

<sup>1,2)</sup>Fakultas Ilmu Keolahragaan Universitas Negeri Yogyakarta,

<sup>3)</sup>Fakultas Teknik Universitas Negeri Yogyakarta

<sup>1)</sup>guntursetiyonos@yahoo.co.id, <sup>2)</sup>sukadiyanto\_fik@yahoo.com, <sup>3)</sup>djemarimardapi@gmail.com

### Abstrak

Penelitian ini bertujuan menghasilkan instrumen asesmen yang valid dan reliabel untuk mengukur hasil belajar pendidikan jasmani olahraga dan kesehatan siswa SMA pada permainan bolavoli. Penelitian pengembangan mengadopsi penelitian pengembangan Borg and Gall dengan 10 langkah. Uji coba skala kecil dilakukan pada siswa Sekolah Laboratorium Olahraga FIK UNY yang berjumlah 24 anak dan uji coba skala besar pada siswa SMAN 1 Yogyakarta, SMAN 2 Wates, SMAN 1 Seyegan, SMAN 1 Sewon, dan SMAN 1 Tanjung Sari berjumlah 120 anak. Instrumen pengumpul data berupa pedoman observasi, sedangkan analisis data dengan menggunakan analisis diskriptif. Hasil penelitian ini adalah instrumen pengamatan hasil belajar siswa penjasorkes pada permainan bolavoli yang memiliki indikator, deskripsi, rubrik, prosedur asesmen. Validitas isi berdasarkan *expert judgement* termasuk kategori baik dan reliabilitas *interrater* dengan paket program Genova menghasilkan koefisien sebesar 0,82, dan *Cohen's Kappa* sebesar 0,79. Kesimpulan penelitian berdasarkan pendapat para guru ialah instrumen ini dapat digunakan untuk mengukur hasil belajar penjasorkes siswa SMA pada permainan bolavoli.

**Kata kunci:** *asesmen hasil belajar, pendidikan jasmani olahraga dan kesehatan, permainan bolavoli*

## DEVELOPING THE ASSESSMENT OF LEARNING OUTCOMES FOR THE STUDENT OF PHYSICAL, SPORTS, AND HEALTH EDUCATION IN VOLLEYBALL GAME FOR SENIOR HIGH SCHOOLS

<sup>1)</sup>Guntur, <sup>2)</sup>Sukadiyanto, <sup>3)</sup>Djemari Mardapi

<sup>1,2)</sup>FIK Universitas Negeri Yogyakarta, <sup>3)</sup>FT Universitas Negeri Yogyakarta

<sup>1)</sup>guntursetiyonos@yahoo.co.id, <sup>2)</sup>sukadiyanto\_fik@yahoo.com, <sup>3)</sup>djemarimardapi@gmail.com

### Abstract

This study aims to produce valid and reliable assessment instruments and to measure the learning outcomes for the students of physical, sport and health education in volleyball game for senior high schools. The research and development model chosen was the model developed by Borg & Gall, with a procedure consisting of ten stages. The field test sample consisted of the 24 students of the sports laboratory school for volleyball of the Faculty of Sports Science, Yogyakarta State University and the large-scale tests were on students of SMAN 1 Yogyakarta, Wates SMAN 2, SMAN 1 Seyegan, SMAN 1 Sewon, and SMAN 1 Tanjung Sari totaling 120 students. The instrument to collect data was observation sheet, whereas data analysis used descriptive analysis. The result of the study is an instrument for assessing the learning outcomes of physical, sports, and health education for the volleyball game that includes indicators, descriptions, and rubrics of performances, and the content validity game based on expert judgment which is good based on expert judgment; Reliability coefficient of the instrument for assessing the practice of the volleyball game by means of the Genova package program is 0.82 and that by means of Cohen's Kappa is 0.79, both satisfy the reliability requirements. Based on the the teachers' opinions, these instruments can be used to measure student learning outcomes of physical, sport and health education at volleyball game in high schools.

**Keywords:** *assessment of learning outcomes, physical sports and health education, volleyball game*

## Pendahuluan

Peranan penting pendidikan jasmani di sekolah adalah memberikan kesempatan kepada siswa untuk terlibat langsung dalam berbagai pengalaman belajar melalui aktivitas jasmani yang dipilih dan dilaksanakan dengan sistematis (Lutan, 2001, p.17). Pendidikan jasmani olahraga dan kesehatan (penjasorkes) merupakan bagian integral dari pendidikan di SMA. Dalam kurikulum kelompok mata pelajaran ini, siswa dituntut untuk dapat menguasai suatu standar kompetensi.

Materi permainan merupakan salah satu standar kompetensi yang harus dikuasai siswa dalam pembelajaran penjasorkes di sekolah menengah atas. Permainan merupakan salah satu cabang olahraga yang diharapkan mampu mengembangkan siswa sesuai dengan tujuan pendidikan yang ingin dicapai. Pada umumnya manusia, baik anak-anak maupun orang dewasa senang bermain, dari rasa senang bermain inilah mudah diketahui keadaan sebenarnya tiap orang yang bermain, sebab secara spontan berbicara, bertindak dan bertingkah laku sesuai dengan kebiasannya. Dengan demikian, permainan dapat digunakan sebagai alat untuk memacu potensi siswa yang berupa gerak, sikap dan perilaku dalam pembelajaran.

Berdasarkan karakteristik permainan ini, dalam mata pelajaran penjasorkes, khususnya permainan bolavoli digunakan oleh guru sebagai media untuk mencapai tujuan pembelajaran berbasis kompetensi. Pembelajaran tersebut menekankan pada pengembangan kemampuan siswa dalam melakukan tugas-tugas dengan unjuk kerja sehingga hasil pembelajarannya berupa penguasaan seperangkat kompetensi.

Dalam rangka proses pengumpulan bukti tentang pencapaian hasil belajar siswa penjasorkes pada permainan bolavoli melibatkan suatu ujian, pengukuran dan asesmen yang dilakukan pada tengah dan akhir pembelajaran. Pelaksanaan asesmen hasil belajar siswa pada mata pelajaran pendidikan jasmani akan berjalan dengan baik apabila ditunjang oleh para guru yang telah

memiliki pengetahuan dan pemahaman tentang tes, dan cara menyusunnya sesuai karakteristik materi (Nurhasan, 2001, p.9). Kemampuan melaksanakan suatu asesmen hasil belajar pada siswa merupakan salah satu keterampilan profesional yang harus dikuasai oleh guru. Keberhasilan dalam melaksanakan asesmen hasil belajar ini akan sangat ditentukan oleh kemampuan guru dalam mengkonstruksi alat ukur, dan menggunakan alat ukur yang telah dikonstruksi itu dengan cara yang benar, serta kemampuan menganalisis data informasi yang dihasilkan oleh alat ukur itu. Apabila keseluruhan kemampuan itu tidak dikuasai oleh guru, maka kemungkinan besar akan terjadi kesalahan dalam pengukuran hasil belajar, yang pada gilirannya akan mengakibatkan kerugian bagi siswa.

Asesmen hasil belajar siswa yang sudah dilaksanakan oleh para guru penjasorkes materi permainan bolavoli selama ini menggunakan tes keterampilan cabang olahraga (*sport skills tests*). Tes keterampilan tersebut suatu tes simulasi yang diadopsi menyerupai situasi olahraganya bertujuan untuk mengukur keterampilan para siswa dalam suatu cabang olahraga.

Jenis tes keterampilan olahraga bermain bolavoli yang sudah ada, di antaranya; *AAPHER serving accuracy test*, *Brumbach forearm pass wall-volley test*, *Kautz volleyball passing test*, *AAPHER face pass wall-volley test*, *Stanley spike test* dan *Brady wall-volley test*. Selama digunakan oleh para guru penjas sebagai alat ukur keterampilan siswa bermain bolavoli, keenam tes tersebut masih memiliki kelemahan-kelemahan di samping kelebihan yang ada.

Tes *AAPHER serving accuracy test* lebih mengarah pada pengukuran kemampuan servis, yaitu dilakukan dengan *teste* berada dalam daerah servis dan melakukan servis bola pada lapangan. Tes *Brumbach forearm pass wall-volley test* untuk mengukur kemampuan *passing* bawah, yaitu *teste* melakukan *passing* bawah dengan sasaran pada tembok. Tes *AAPHER face pass wall-volley test* untuk mengukur kemampuan *passing* atas atau umpan, yaitu *teste* melakukan *passing* atas dengan sasaran pada tembok. Tes *Brady*

*wall-volley test* untuk mengukur keterampilan bermain bolavoli, yaitu *teste* melakukan *passing* bawah dan *passing* atas dengan sasaran pada tembok. Tes *Stanley spike test* untuk mengukur kemampuan *spike* atau *smash*, yaitu *teste* berada di dalam daerah bebas dan melakukan *smash* dalam lapangan.

Kelemahan dari AAPHER *serving accuracy test* dan *Stanley spike test* adalah harus tersedia lapangan yang berukuran 9 x 18 m. Sedangkan pada Brumbach *forearm pass wall-volley test*, Kautz volleyball *passing test*, AAPHER *face pass wall-volley test* dan Brady *volley test* harus tersedia tembok sasaran berukuran 5 x 6 feet dengan tinggi 10 feet dari lantai. Selain itu, tes keterampilan olahraga bolavoli berbentuk simulasi tersebut memerlukan peralatan dan tempat yang lebih banyak karena antara butir tes yang satu dengan butir tes lainnya dilakukan secara terpisah sehingga kurang efektif dan efisien bila digunakan untuk mengukur siswa dalam jumlah banyak.

Kelemahan penggunaan tes keterampilan olahraga ini merupakan prediktor yang dianggap tidak valid dalam mengukur kemampuan siswa saat bermain sesungguhnya karena keterampilan siswa dari hari ke hari tidak tetap dan keterampilan yang biasanya diujikan pada siswa keluar dari konteks (Veal, 2002, p.88). Jenis tes tersebut banyak mengukur pada aspek hasil yaitu pada ketepatan bola dan belum mengukur pada aspek proses berupa tahapan-tahapan gerak yang baik dan benar pada suatu teknik atau cara memainkan bola, di samping itu tes ini sebenarnya cocok untuk mengukur kemampuan atlet profesional dan tidak relevan untuk mengukur kompetensi siswa di sekolah.

Model asesmen yang diharapkan dalam pendidikan jasmani mampu mengukur kompetensi siswa dalam konteks kehidupan nyata adalah penilaian unjuk kerja (Mitchell, 1999, p.19). Asesmen unjuk kerja dalam mata pelajaran pendidikan jasmani memungkinkan guru untuk melihat kompetensi siswa yang teramati dalam bermain permainan bolavoli secara nyata.

Asesmen unjuk kerja (*performance assessment*) permainan bolavoli adalah asesmen

yang meminta peserta tes untuk melakukan unjuk kerja bermain bolavoli pada situasi nyata sesuai kriteria yang diharapkan dan dapat diaplikasikan secara aktual. Pada asesmen ini ditinjau beberapa aspek dalam asesmen yaitu proses dan hasil yang dilakukan siswa dalam bermain bolavoli. Kedua aspek asesmen ini akan memberikan gambaran tentang kompetensi siswa yang sebenarnya.

Dalam asesmen proses seorang guru dapat mengamati bagaimana ketepatan pelaksanaan gerak siswa dalam tahapan persiapan saat akan memainkan bola berupa keterampilan siswa menyiapkan sikap awal tubuh, gerak awal, gerak perkenaan bola dan gerak lanjut sesuai teknik atau cara yang akan digunakan dalam memainkan bola. Pada asesmen produk seorang guru dapat melihat hasil penempatan bola siswa setelah mengalami serangkaian proses teknik atau cara memainkan bola dalam bermain bolavoli dan nilai-nilai berupa sikap kerja sama dan sportivitas siswa dalam bermain.

Permasalahan yang sering dihadapi guru pendidikan jasmani dalam melakukan asesmen unjuk kerja terletak pada validitas dan reliabilitas alat ukur yang digunakan. Penyusunan tes unjuk kerja siswa masih sangat terbatas pada pengetahuan dan pemahaman guru tentang tes berbentuk simulasi. Hasil asesmen sering dipengaruhi oleh subjektivitas guru sebagai *rater* karena dalam melakukan asesmen dilakukan sendiri tanpa melibatkan guru yang lain sebagai kolaborator.

Berdasarkan latar belakang di atas, maka perlu dikembangkan suatu instrumen asesmen hasil belajar siswa penjasorkes pada permainan bolavoli yang menghasilkan alat asesmen berguna bagi guru untuk mengumpulkan informasi pencapaian kompetensi siswa. Diharapkan instrumen ini bermanfaat secara optimal untuk meningkatkan kualitas pembelajaran pendidikan jasmani di jenjang SMA.

Mardapi (2012, p.12), berpendapat bahwa penilaian atau asesmen merupakan komponen penting dalam penyelenggaraan pendidikan. Upaya meningkatkan kualitas

pendidikan dapat ditempuh melalui peningkatan kualitas pembelajaran dan kualitas sistem penilaiannya. Keduanya saling terikat, sistem pembelajaran yang baik akan menghasilkan kualitas belajar yang baik. Kualitas pembelajaran ini dapat dilihat dari hasil penilaiannya.

Asesmen adalah proses pengumpulan data secara sistematis untuk membuat keputusan tentang seseorang (Berk, 1986, p.ix). Asesmen mencakup semua cara yang digunakan untuk mengumpulkan data tentang individu sehingga keputusannya juga terhadap individu (Mardapi, 2012, p.13). Data yang diperoleh dengan berbagai cara kemudian diolah menjadi informasi tentang individu. Jadi proses asesmen meliputi pengumpulan bukti tentang pencapaian hasil belajar siswa. Berdasarkan pendapat tersebut maka dapat disimpulkan bahwa asesmen adalah semua cara yang digunakan untuk mengumpulkan data tentang individu.

Peranan penting asesmen hasil belajar adalah untuk menentukan capaian belajar siswa (Jennifer, 2009, p.33). Menurut Els van der Werf (2006, p.17) hasil belajar adalah kemampuan yang bisa dihasilkan siswa yang melibatkan pengukuran pada ranah kognitif, afektif dan psikomotor yang secara terpadu dikaitkan dengan seluruh domain perkembangan siswa itu sendiri.

Asesmen hasil belajar siswa pada kelompok mata pelajaran penjasorkes dilakukan melalui pengamatan terhadap perubahan perilaku dan sikap untuk menilai perkembangan psikomotor dan afektif siswa (BSN-P, 2007, p.17). Asesmen yang dilakukan guru mencakup semua hasil belajar siswa yaitu kemampuan kognitif atau berpikir, kemampuan psikomotor atau kemampuan praktik, dan kemampuan afektif, namun penekanan pada setiap ranah tidak sama sehingga harus diperhatikan karakteristik mata pelajaran yang akan diukur (Mardapi, 2012, p.15). Penilaian capaian hasil belajar pada tingkat kognitif yang lebih tinggi (*higher-order thinking*), menurut Nitko dan Brookhart (2007, p.208) dibutuhkan tes (*task*) yang menuntut peserta didik untuk menggunakan pengetahuan dan keterampilan dalam situasi yang

baru (*new or novel situations*). Dengan demikian, peserta didik tidak hanya dituntut untuk memahami, tetapi sampai mampu untuk menganalisis, mengevaluasi dan berkreasi.

Penilaian penguasaan kompetensi aspek keterampilan atau psikomotor siswa di mata pelajaran pendidikan jasmani dilakukan dengan penilaian unjuk kerja (*performance assessment*) (Metzler, 2005, p.178). Dalam rangka untuk mengetahui capaian hasil pembelajaran berbasis kompetensi melibatkan penggunaan suatu sistem asesmen kompetensi. Hayton dan Wagner (1998, p.71) menyatakan *performance assessment is a technique that is likely to be used in a competency-based system because both the system and the technique have a focus on criterion activities or outcomes*.

Menurut Stiggins, R (1997, p.34) *“performance assessments call upon the examinee to demonstrate specific skills and competencies, that is, to apply the skills and knowledge they have mastered”*. Wiggins, G (1993, p.57) menekankan hal yang lebih unik lagi perlunya kinerja ditampilkan secara efektif dan kreatif. Selain itu, tugas yang diberikan dapat berupa pengulangan tugas atau masalah yang analog dengan masalah yang dihadapi. Hal yang penting dalam penilaian unjuk kerja adalah cara mengamati dan menskor kemampuan kinerja peserta didik.

Guna meminimalkan faktor subjektivitas keadilan dalam menilai kemampuan kinerja peserta didik, biasanya *rater* atau penilai jumlahnya lebih dari satu orang sehingga diharapkan hasil penilaian mereka menjadi lebih valid dan reliabel. Di samping itu, dalam pelaksanaan penilaian diperlukan suatu pedoman penilaian yang bertujuan untuk memudahkan penilai dalam menilai sehingga tingkat subjektivitas bisa ditekan.

Penilaian pada domain afektif dalam pendidikan jasmani sangat penting dilakukan karena sebagian besar aktivitasnya melibatkan interaksi dengan orang lain dan hanya melalui interaksi sosial kualitas siswa melakukan pembelajaran sepenuhnya mencapai sasaran (Vicki Worrell, 2002, p.9). Penilaian aspek afektif pada mata pelajaran penjas sebaiknya lebih ditekankan kepada internalisasi nilai-nilai sikap siswa berupa

kerja sama yang ditampilkan dalam proses mempraktikkan suatu permainan (Grineski, S. 1989, p.21). Kegiatan pendidikan jasmani penuh dengan kesempatan untuk mengajarkan keterampilan psikososial seperti kerjasama dan sportivitas (Tomme & Wendt, 2003, p.68). Menurut Vicki Worrell (2002, p.7) penilaian yang efektif pada domain ini memerlukan sasaran pengukuran dari unsur-unsur afektif salah satu unsurnya yaitu interaksi interpersonal seperti sportivitas sebagai tujuan belajar siswa dalam pendidikan jasmani.

Sportivitas adalah menjadi olahragawan (*sportman*) yaitu orang yang dapat mengambil kerugian atau kekalahan tanpa keluhan atau kemenangan tanpa sombong dan yang memperlakukan lawan-lawannya dengan keadilan, kemurahan hati, dan sopan (Keating, 2001, p.12). Penilaian pada permainan olahraga bolavoli dilakukan terutama pada aspek kemampuan siswa menampilkan sikap dalam bermain terutama pada kerja sama dan *sportivitas* ketika melaksanakan bermain yang dilakukan dengan cara, antara lain: observasi perilaku.

Mata pelajaran Pendidikan Jasmani, yang dalam kurikulum disebut secara paralel dengan istilah lain menjadi Pendidikan Jasmani, Olahraga dan Kesehatan, merupakan salah satu mata pelajaran yang disajikan di sekolah, mulai dari SD sampai dengan SMA. Menurut Jewett dan Nixon (1995, p.27), pendidikan jasmani adalah satu aspek dari proses pendidikan keseluruhan yang berkenaan dengan perkembangan dan penggunaan kemampuan gerak individu yang sukarela dan berguna serta berhubungan langsung dengan respon mental, emosional dan sosial.

Menurut Pettifor (1999, p.134), pendidikan jasmani menyediakan siswa untuk mengamalkan hidup aktif dan sehat dengan menyediakan cakupan pengalaman-pengalaman pembelajaran yang rapi dan sistematis. Artinya, pendidikan jasmani membekali ruang untuk membentuk siswa yang mantap secara seluruhnya (*whole child*) yang diarahkan pada pembangunan aspek-aspek kognitif, afektif dan psikomotor.

Penjasorkes merupakan satu-satunya mata pelajaran di sekolah yang menggunakan gerak sebagai media pembelajaran untuk mencapai tujuan pendidikan. Wuest dan Bucher (2009, p.97) menyebutkan, "*Movement is the Keystone of Physical Education and Sport.*" Jelas dinyatakan bahwa gerak merupakan kunci dari pendidikan jasmani dan olahraga. Menurut Lutan (2001, p.15) proses belajar untuk bergerak dan belajar melalui gerak merupakan dua makna yang patut dipegang dalam penjasorkes.

Pembelajaran pendidikan jasmani diajarkan dengan berbagai keterampilan gerak teknik dan strategi permainan/olahraga, internalisasi nilai *sportivitas*, dan kerja sama (Asmawi, 2006, p.134). Melalui suatu gerakan siswa dituntun untuk mengetahui cara melakukan gerakan tersebut, mengetahui kebermanfaatan gerakan tersebut dan juga mampu menunjukkan perilaku-perilaku positif selama pembelajaran yang diharapkan mampu juga diwujudkan siswa dalam kehidupannya sehari-hari.

Dalam mata pelajaran pendidikan jasmani olahraga dan kesehatan di sekolah menengah digunakan permainan bolavoli sebagai salah satu materi pembelajaran gerak. *American Sport Education Program* (2007, p.22) menyatakan permainan bolavoli adalah sebagai berikut: *Volleyball is unique in that it is a game of errors where the objective is to get the ball to hit the floor on the opponent's side of the net or force the opponent to make ballhandling error. Thus, the majority of points scored in volleyball are result of an error.*

Konsep dan prinsip bolavoli menurut Yudiana & Subroto (2010, p.36) permainan bolavoli pada awal ide dasarnya adalah permainan memantul-mantulkan bola (*to-volley*) oleh tangan atau lengan dari dua regu yang bermain di atas lapangan yang mempunyai ukuran-ukuran tertentu. Untuk masing-masing regu, lapangan di bagi dua sama besar oleh net atau tali yang dibentangkan di atas lapangan. Lapangan di bagi dua sama besar oleh net atau tali yang dibentangkan di atas lapangan dengan ukuran ketinggian tertentu. Satu orang pemain tidak boleh memantulkan bola dua kali secara berturut-

turut kecuali pembendung, dan satu regu dapat memainkan bola maksimal tiga kali sentuhan di lapangan sendiri.

Pembelajaran dasar permainan bola-voli dapat bertujuan untuk kesenangan, lebih lanjut menurut Joel (2003, p.vii), "*volleyball is a unique, exciting game that requires solid teamwork and consistent individual execution*". Keunikan bolavoli tidak seperti olahraga lain karena dalam bermain selain harus adanya unsur kerja sama tim sehingga tanpa kerja sama tim tidak akan bisa melakukan serangan dengan tepat (Kinda, 2006, p.v). Bolavoli sebagai olahraga tim memiliki karakteristik yang unik yaitu memiliki aturan bermain, menggunakan lapangan berukuran 18x9 m dan dibatasi net. Permainan bolavoli menurut jenis klasifikasi permainannya termasuk pada jenis *net games* (Hopper, 1998, p.17). *This primary rule leads to progressive principles of play that are consistency, then placement of the object and positioning in relation to opponent's target area, and finally spin and power to make it difficult for an opponent to get the object back into play* (Hopper, 1998, p.18). Permainan ini berkaitan dengan beberapa hal yaitu melakukan *rally* secara konsistensi, penempatan bola pada target lawan atau daerah lawan, dan keterampilan untuk memukul bola dengan teknik yang dipilih untuk mematikan lawan.

Menurut Crum (2006) bolavoli adalah *rallying versus playing to the ground, to hit (or touch) somebody's field (or the field of the other team) with an object (for example a ball) with respect to prevent that or to hit (or touch) somebody's field (or the of other team) that way the other person (or team) the object can't return (in my our field)*. Menurut Viera, (2004, p.4) rally adalah permainan bola menyeberangkan net antara dua tim yang saling berlawanan yang berakhir dengan perolehan angka atau pindah bola.

Viera (2004, p.121) menyatakan dalam bermain yang sesungguhnya dibutuhkan keterampilan dasar yang sistematis. Artinya, bermain bolavoli merupakan suatu urutan rangkaian permainan yang terdiri dari sebuah servis, yang kemudian dioperkan dengan menggunakan *passing* bawah atau *passing* atas

ke arah pengumpan, kemudian mengumpankan kepada penyerang, memukul bola ke lapangan lawan, pada saat bersamaan di mana lawan juga mempersiapkan pertahanan dengan menggunakan blok, dan penerima di lapangan belakang, rangkaian ini terjadi secara terus-menerus selama pertandingan. Dari pendapat di atas dapat disimpulkan bahwa permainan bolavoli adalah permainan yang aktivitasnya berusaha saling menjatuhkan bola ke lantai lawan dengan cepat dengan tujuan memaksa regu lawan untuk tidak mampu mengembalikan bola sehingga tercetak skor/*point*.

Dalam rangka pengumpulan bukti-bukti tentang pencapaian hasil belajar siswa penjasorkes pada permainan bolavoli di SMA melibatkan kegiatan pengukuran pada unjuk kerja siswa bermain bolavoli. Permainan bolavoli berorientasi pada kinerja yang melibatkan keterampilan gerak (*skill movemet*) (Bonnie, 2006, p.2). Penilaian otentik diakui sebagai cara lain untuk menilai belajar siswa dan meningkatkan pengajaran pendidikan jasmani (Hensley, 1997, pp.19-24). Penilaian otentik digambarkan oleh Veal (2002, p.93) adalah adanya hubungan antara tugas-tugas instruksional sehari-hari dan penilaian.

Menurut Wiggins (1993, p.205) tugas pertama guru harus memutuskan apa yang penting bagi siswa untuk tahu. Ketika guru telah mendefinisikan apa yang siswa harus ketahui dan lakukan, barulah guru kemudian dapat merancang sebuah penilaian otentik tidak hanya untuk mencocokkan tujuan program sesuai dengan kurikulum, tetapi juga untuk mencocokkan instruksi tugas guru.

Penilaian otentik bercirikan sebagai penilaian berbasis kinerja, artinya siswa mengambil bentuk tugas. Hasil dari tugas berupa pekerjaan siswa yang disebut sebagai produk, dan mencetak rubrik atau panduan yang dikembangkan dan dirancang untuk sesuai dengan isi tugas dan digunakan untuk menilai produk siswa. Terdapat dua komponen penting dalam asesmen otentik, yaitu tugas kinerja (*performance task*) dan rubrik performansi (*performance rubrics*).

*Task* merupakan tugas-tugas yang akan dilakukan untuk membuat asesmen kinerja, sedangkan rubrik terdiri dari daftar kriteria yang diwujudkan dengan dimensi-dimensi kerja, aspek proses atau konsep-konsep yang akan dinilai dan gradasi mutu mulai dari tingkat yang paling sempurna sampai dengan tingkat yang buruk (Zainul, 2005, p.13). Dalam pengukuran unjuk kerja diperlukan adanya reliabilitas, Johnson (2009, p.22) menyatakan “*by reliability, we mean the consistency of examinees score across such factors as occasions, task, and raters. in others words, reliability addresses whether an examinees score would be the same if she were to take the exam on a different occasion, complete different tasks, or be scored by different raters*”.

Dalam penilaian otentik digunakan metode reliabilitas antar-rater, mengharuskan penilai melakukan pengamatan untuk menilai atau mengukur beberapa aspek perilaku siswa dalam tugas. Untuk mencapai tingkat keandalan antar-rater yang tinggi perlu dirancang dan diterapkan rubrik.

Pengembangan rubrik sangat diperlukan dalam penilaian kinerja, yang digunakan sebagai dasar pengukuran. Rubrik adalah pedoman dari produk suatu penilaian. Rubrik atau kriteria merupakan panduan untuk memberi skor yang jelas dan disepakati oleh guru dan siswa (Zainul, 2005, p.9). Panduan ini menjelaskan kepada guru dan siswa tentang standar yang ada dalam unjuk kerja (Winter, 1996, p.47).

Lund (2002, p.43) menyatakan bahwa “*rubric indicate the criteria a person scoring performance-based assessment should use when doing evaluations*”. Lebih lanjut kriteria untuk penilaian unjuk kerja sering disebut *scoring criteria, scoring guidelines, rubrics, and scoring rubric* tetapi memiliki makna yang sama (Lund, 2002, p.44). Desain rubrik penilaian membutuhkan spesifikasi dari kriteria untuk menilai kualitas kinerja dan pilihan prosedur penilaian (Brenan, 2006, p.394). Kriteria dalam rubrik menggambarkan elemen-elemen penting dari unjuk kerja dan menjadi sumber kriteria yang akan dinilai.

Kriteria yang jelas merupakan esensi dalam penilaian unjuk kerja sehingga siswa

dapat dinilai secara konsisten (Arter, 1996, p.vi). Kriteria rubrik yang digunakan guru adalah *product criteria* (Guskey 1996b, p.4). Kriteria produk adalah mengenai hasil apa dari siswa.

Dalam pendidikan jasmani kriteria produk sering disebut unjuk kerja siswa yang dicontohkan dalam produk kinerja siswa dalam bermain permainan olahraga (*play to game*) (Lund, 2002, p.45). Penilaian unjuk kerja mata pelajaran penjasorkes di SMA menerapkan penilaian unjuk kerja, di mana siswa diminta untuk melaksanakan tugas bermain bolavoli.

Penilaian unjuk kerja melibatkan dua macam relevansi, yaitu sesuai dengan kompetensi dan bermakna dalam kehidupan nyata, artinya sebuah kriteria atau indikator penilaian capaian hasil belajar harus cocok dengan kompetensi yang dibelajarkan dan sekaligus bermakna atau relevan dengan kehidupan nyata. Terkait dengan hal tersebut, pada penelitian dideskripsikan: (1) bentuk instrumen asesmen hasil belajar siswa penjasorkes pada permainan bolavoli yang tepat di sekolah menengah atas; (2) validitas instrumen asesmen hasil belajar siswa; (3) reliabilitas instrumen asesmen hasil belajar siswa penjasorkes pada permainan bolavoli di sekolah menengah atas yang dikembangkan; (4) prosedur penerapan asesmen hasil belajar siswa penjasorkes pada permainan bolavoli di sekolah menengah atas yang dikembangkan; (5) profil hasil belajar siswa penjasorkes pada permainan bolavoli di sekolah menengah atas berdasarkan asesmen yang dikembangkan.

## Metode Penelitian

### Jenis Penelitian

Penelitian ini menggunakan pendekatan penelitian pengembangan (*Research and Development*), karena sesuai dengan tujuan yang hendak dicapai, yaitu pengembangan asesmen hasil belajar siswa penjasorkes khususnya pada permainan bolavoli. Wujudnya berupa instrumen asesmen yang diharapkan dapat digunakan guru penjasorkes untuk memotret hasil belajar siswa pada materi

permainan bolavoli di sekolah menengah atas.

Model penelitian pengembangan yang dipilih adalah model penelitian dan pengembangan yang dikembangkan oleh Borg and Gall (1983, p.772). Dasar peneliti menggunakan model penelitian pengembangan karena rangkaian langkah-langkah penelitian dan pengembangan dilakukan secara siklis, dan pada setiap langkah yang dilakukan selalu mengacu pada hasil langkah sebelumnya hingga pada akhirnya diperoleh suatu produk pendidikan baru.

Adapun langkah-langkah metode R & D Borg & Gall (1983, p.775) adalah; *research and information collecting, planning, develop preliminary form of product, preliminary field testing, main product revision, main field testing, operational product revision, operational field testing, final product revision, dissemination and implementation.*

Berdasarkan sepuluh langkah penelitian dan pengembangan yang dikembangkan Borg & Gall tersebut, dalam penelitian ini pada proses pelaksanaannya melakukan adaptasi yang mengacu pada model pendekatan tersebut. Adapun adaptasi penelitian pengembangan ini secara garis besar terdiri dari empat tahap kegiatan pokok, yakni: (1) studi pendahuluan; (2) pengembangan produk; (3) uji coba dan revisi produk; dan (4) diseminasi produk hasil pengembangan.

#### Waktu dan Tempat Penelitian

Uji coba instrumen penilaian dilaksanakan pada bulan Januari 2013 kepada para siswa Sekolah Laboratorium Olahraga (SELABORA) FIK UNY yang melibatkan tiga puluh satu orang, meliputi subjek uji coba keseluruhan berjumlah dua puluh empat orang terdiri dari dua belas orang siswa putera dan dua belas orang siswa puteri yang berasal dari SMA Negeri di DIY, tiga orang guru sebagai *rater*, dua orang sebagai wasit, dan dua orang sebagai pengambil gambar *video*.

#### Subjek Penelitian

Sumber informasi dalam penelitian ini adalah semua siswa yang menempuh mata

pelajaran penjasorkes pada Sekolah Menengah Atas di DIY. Populasi siswa SMA yang mengikuti mata pelajaran penjasorkes di DIY tersebar di empat kabupaten dan satu kotamadya, yaitu Kab. Sleman, Kab. Bantul, Kab.Kulon Progo, Kab. Gunung Kidul dan kota Yogyakarta. Peneliti mengambil kelas XI dari SMA negeri bukan dari SMA swasta, karena SMA Negeri dapat dijadikan tolak ukur dalam proses pembelajaran penjasorkes.

Sampel sumber informasi dalam penelitian ini terdiri atas dua elemen yaitu: (1) siswa kelas XI; dan (2) penilai atau *rater* dalam bidang permainan bolavoli yaitu dari guru bidang studi penjasorkes. Penentuan sampel dalam penelitian ini diambil secara *purposive sampling* dengan kriteria; (1) sekolah SMA Negeri yang menyelenggarakan kelas khusus yaitu kelas olahraga, (2) sekolah SMA Negeri yang minimal sudah merintis kelas olahraga. Sehingga diperoleh sampel siswa dari Kab.Sleman diambil dari SMAN 1 Seyegan, Kab.Kulon Progo diambil SMAN 2 Wates, Kab.Bantul diambil dari SMAN I Sewon, Kab.Gunung Kidul diambil SMAN 1 Tanjung Sari, dan Kota Madya Yogyakarta diambil SMAN 1 Yogyakarta. Jumlah siswa yang dijadikan sampel dari empat kabupaten, dan satu kota madya adalah 120 siswa yang terdiri 60 siswa putera dan 60 siswa puteri.

#### Instrumen dan Teknik Pengumpul Data

Asesmen hasil belajar siswa penjasorkes pada permainan bolavoli dilakukan dengan instrumen pengamatan (observasi) pada unjuk kerja siswa. Teknik pengumpulan data asesmen hasil belajar siswa penjasorkes pada permainan bolavoli di SMA menggunakan tugas unjuk kerja siswa bermain bolavoli 6 lawan 6 dibatasi tiga putaran posisi *server* pertama yang diamati menggunakan pedoman observasi dilengkapi dengan lembar rubrik.

Pedoman pengamatan dalam asesmen hasil belajar siswa penjasorkes pada permainan bolavoli dipergunakan untuk mengamati dan menilai setiap siswa peserta tes unjuk kerja dengan menggunakan rubrik

berbentuk skala penilaian (*rating scale*) beserta bobot penilaiannya. Penyusunan konstruk instrumen pengamatan dalam bentuk skala penilaian berdasarkan materi yang mencerminkan keterampilan yang akan diukur. Selanjutnya, ditentukan skala penilaian untuk setiap materi. Dalam penelitian ini digunakan empat skala penilaian, yaitu satu sampai empat (1 – 4). Skala tersebut dibagi menjadi dua bagian, yaitu kompeten dan tidak kompeten.

Penetapan standar kompetensi pada asesmen hasil belajar siswa penjasorkes pada permainan bolavoli berupa unjuk kerja siswa bermain bolavoli 6 lawan 6 mencakup kemampuan dalam persiapan saat akan memainkan bola, proses teknik memainkan bola, hasil penempatan bola dan sikap siswa dalam bermain. Apabila siswa memperoleh nilai  $\geq 70$  KKM maka dikategorikan kompeten. Apabila siswa memperoleh nilai  $< 70$  KKM maka dikategorikan tidak kompeten dan harus menempuh remedi pada bagian yang tidak kompeten.

Penetapan standar kompetensi mengacu pada dua pendekatan metode *Bookmark* dan kriteria ketuntasan minimal (KKM). Metode *Bookmark* berkenaan dengan penetapan tingkat kesulitan (Hambleton & Pitoniak, dalam Brennan, 2006, pp.442-444). Dalam teori respon butir, model logistik satu parameter adalah salah satu model *item respons test* (IRT) yang mengandung satu karakteristik *item*, yaitu tingkat kesulitan. Dalam metode *Bookmark* digunakan sebuah respon probabilitas 0,67, yang menunjukkan titik skor yang ditetapkan panelis sebagai probabilitas respon.

Berdasarkan kedua sumber tersebut di atas dan kesepakatan antara guru penjasorkes, peneliti menetapkan KKM sebesar 70 sebagai pedoman penetapan standar kompeten pada asesmen hasil belajar siswa penjasorkes pada permainan bolavoli. Nilai KKM sebesar 70 yang dicapai siswa penjasorkes pada permainan bolavoli menunjukkan bahwa siswa mampu bermain bolavoli (mulai dari menyiapkan sikap awal tubuh yang sesuai dengan jenis teknik atau cara memainkan bola yang dipilih sampai hasil

penempatan bola ke lapangan lawan beserta kemampuan kerjasama dan sportivitas).

#### Teknik dan Analisis Data

Dalam rangka untuk mengetahui apakah instrumen penelitian ini sudah memenuhi kriteria sebagai instrumen yang baik, diperlukan analisis. Sejauhmana instrumen pengamatan dapat mengukur hasil belajar siswa penjasorkes pada permainan bolavoli, untuk itu diperlukan adanya validasi. Konstruk instrumen pengamatan yang digunakan dalam asesmen hasil belajar siswa penjasorkes pada permainan bolavoli ini disesuaikan dengan aspek-aspek yang diperlukan dalam unjuk kerja nyata siswa bermain bolavoli 6 lawan 6 namun sebelum instrumen ini digunakan harus diuji validitas.

Penelitian ini berorientasi pada pencapaian hasil belajar sehingga penekannya lebih kepada validitas isi, karena berhubungan dengan kemampuan instrumen untuk menggambarkan secara tepat domain perilaku yang dinilai. Langkah validitas isi dijabarkan ke dalam aspek yang terinci dan dideskripsikan indikatornya.

Proses validasi instrumen pengamatan unjuk kerja siswa bermain bolavoli 6 lawan 6 berdasarkan *expert judgment* melalui *Focus Group Discussion* (FGD). Forum ini melibatkan pendapat para pakar, yaitu: guru bidang penjasorkes SMA, dosen mata kuliah permainan bolavoli jurusan POR FIK UNY, praktisi dalam bidang penjasorkes, dan pakar pengukuran.

Penentuan koefisien reliabilitas instrumen asesmen hasil belajar siswa penjasorkes pada permainan bolavoli dilakukan dengan menggunakan koefisien Genova (*Generalizability of variants*). Program Genova mengacu pada teori *Generalizability* yang dikembangkan oleh Crick dan Brennan pada tahun 1983, yang terdiri dari *Generalizability Study* (*G study*) dan *Decision Study* (*D study*) dengan komponen variansinya adalah *person*, *rater*, dan *item*. *G study* digunakan untuk mengestimasi komponen variansi kesalahan yang diakibatkan oleh berbagai sumber variansi. *D study* digunakan untuk menge-

tahui koefisien reliabilitas (Retnowati, 2009, p.184).

Besarnya koefisien keandalan merupakan rasio antara varians sistematis atau varians sebenarnya dan varians total. Penelitian ini menggunakan GENOVA yang komponen variansnya adalah *person* (P), *rater* (I), *item* (I), interaksi *person* dan *rater*, dan kesalahan. *G study*nya menggunakan rancangan bersarang (*nested design*) dan *D-study*-nya juga menggunakan rancangan bersarang (*nested design*).

Penelitian ini menggunakan satu *facet* P x (I:I) *G-study* yang bersarang untuk mengestimasi varians komponen, varians kesalahan, generalizability dan koefisien phi untuk *one-facet, nested, I:I D-study*. Varians komponen yang berbaur pada rancangan bersarang (P,T:I,E) adalah jumlah varians komponen dalam *G study* bersarang yang dapat ditulis sebagai berikut.

$$\sigma^2_{p,r:i,e} = \sigma^2_p + \sigma^2_{r i e}$$

Keterangan:

p = person

r = guru/*rater*

i = *item*

r:i = *rater* bersarang pada *item*

e = kesalahan

Setelah varians komponen diperoleh, termasuk varians kesalahan, maka dapat diestimasi varians sebenarnya (*true variance*). Selanjutnya, dapat diestimasi besarnya indeks keandalan hasil pengukuran, yaitu rasio varians sebenarnya terhadap varians keseluruhan komponen. Estimasi varians setiap komponen dan besarnya indeks keandalan hasil pengukuran dengan instrumen yang dikembangkan peneliti menggunakan paket program GENOVA.

Rancangan yang digunakan untuk *G-study* adalah P X (I:I), yaitu *item* bersarang pada *rater*, penilai dalam menilai anak bermain permainan bolavoli berinteraksi dengan anak yang bersarang pada *item*. Cara penilai atau *rater* (I) dalam menilai siswa bermain permainan bolavoli 6 lawan 6 (P) tergantung pada pendapat penilai terhadap *item* yang dinilai, sehingga dikatakan *rater* bersarang pada *item*.

Rancangan P X (I:I) ini berdasarkan analisis varians efek random memiliki efek utama: P, T, I:I dan efek interaksinya adalah PI, PT bersarang pada I. Jadi, ada varians *person*, varians *rater*, dan varians penilai bersarang pada I untuk efek utama, sedang untuk efek interaksinya adalah varians *person item*, varians *rater* yang bersarang pada *item*.

Besarnya varians R bersarang pada I dapat ditulis sebagai berikut:

$$\sigma^2(\mathbf{T: I}) = \sigma^2(\mathbf{T, TI}) = \sigma^2(\mathbf{T}) + \sigma^2(\mathbf{TI})$$

Besarnya koefisien keandalan instrumen penilaian adalah:

$$E\rho^2 = \frac{\sigma^2(\mathbf{P})}{\sigma^2(\mathbf{P}) + \sigma^2(\delta)}$$

$E\rho^2$  adalah nilai harapan koefisien keandalan instrumen,  $\sigma^2(\mathbf{P})$  adalah varians *person* (siswa)  $\sigma^2(\delta)$  adalah varians kesalahan. Varians kesalahan terdiri atas varians *rater*, varians *item*, dan varians interaksi *rater item*. Besarnya varians ini diestimasi dengan menggunakan teknik analisis varians rancangan efek random. Untuk melihat reliabilitas dari kriteria instrumen penilaian permainan bolavoli hasil uji coba, digunakan analisis koefisien *interrater*.

Koefisien *interrater* adalah salah satu sarana untuk melihat tingkat konsistensi atau keajegan antar-*rater* dalam memberikan *rating* terhadap unjuk kerja siswa dalam bermain bolavoli. Untuk keperluan ini, digunakan koefisien Cohen's Kappa. Ada 3 (tiga) orang *rater* yang memberikan *rating* pada penilaian unjuk kerja bermain bolavoli 6 lawan 6 pada siswa putera dan puteri.

Pada penilaian persiapan saat akan memainkan bola, ada 5 (lima) *item* yang menjadi objek penilaian, pada penilaian proses teknik memainkan bola ada 15 (lima belas) *item*, penilaian hasil penempatan bola ada 7 (tujuh) *item* dan penilaian kemampuan menampilkan sikap saat bermain siswa ada 2 (dua) *item*. Selanjutnya nilai koefisien  $\kappa$  yang dihasilkan dibandingkan dengan kriteria minimal yang diperkenankan, yaitu 0,70 (Linn, 1990, p.143). Selanjutnya, asesmen hasil belajar siswa penjasorkes pada permainan bolavoli dianalisis dengan menggunakan teknik analisis deskriptif, yaitu

mendesripsikan penilaian yang dilakukan selama siswa melakukan tes unjuk kerja bermain bolavoli 6 lawan 6.

### Hasil dan Pembahasan Penelitian

#### Data Pendahuluan

##### Hasil Analisis Kebutuhan

Dalam analisis kebutuhan lapangan peneliti mewancarai 22 orang guru mata pelajaran penjasorkes SMA yang berada di 5 wilayah kabupaten DIY. Ada 5 *item* yang diidentifikasi dalam studi awal ini, yaitu kurikulum, tujuan hasil belajar penjasorkes pada permainan, penggunaan perangkat penilaian, metodepenilaian, penggunaan instrumen.

Hasil wawancara yang dilakukan oleh peneliti berkaitan dengan masing-masing *item* tersebut dideskripsikan sebagai berikut.

Pengalaman guru penjasorkes dalam menentukan tujuan pembelajaran penjasorkes adalah berdasarkan kurikulum tingkat satuan pendidikan (KTSP) dan menggunakan gabungan KTSP dengan kurikulum sebelumnya. KTSP yang dikembangkan berdasarkan Standar Kompetensi dan Kompetensi Dasar Permen No 2 Tahun 2006 tentang standar isi pendidikan merupakan kurikulum terbaru yang digunakan pada jenjang sekolah dasar sampai sekolah menengah.

Tabel 1. Penggunaan Kurikulum

Urut- an	Tanggapan	Jumlah Respon	%
1	Pengalaman guru dalam menentukan tujuan pembelajaran siswa berdasarkan kurikulum yang berlaku yaitu KTSP	17	77,28
2	Pengalaman guru dalam menentukan tujuan pembelajaran siswa berdasarkan KTSP dan menggunakan gabungan dengan kurikulum sebelumnya	5	22,72

N=22

Kurikulum yang digunakan sebelumnya adalah Kurikulum 2004, yaitu kurikulum berbasis kompetensi (KBK). Namun, dapat disimpulkan KTSP pada dasarnya sama dengan kurikulum 2004, perbedaannya hanya

pada standar kompetensi yang ingin dicapai dan cakupannya.

Terkait dengan tujuan dan hasil belajar, pada kenyataannya prioritas hasil belajar pada penjasorkes pada permainan bolavoli di kalangan para guru sangat beragam. Keberagaman tersebut secara rinci dapat dilihat melalui jumlah respon, persentase, dan prioritas pilihan sebagaimana terangkum dalam Tabel 2.

Tabel 2. Tujuan dan Hasil Belajar Permainan Bolavoli pada Penjasorkes

Urut- an	Aspek	Jumlah Respon	%
1	Penguasaan siswa pada aspek proses teknik atau cara memainkan bola yang diperoleh melalui berbagai pengalaman dalam bermain bolavoli (menerima bola, mengumpan, memukul, membendung, dan lain-lain) membuat senang siswa.	9	40,90
2	Bermain bolavoli dan keterkaitannya dengan aspek kerjasama, sportivitas, menaati aturan main, saling menghormati.	5	22,72
3	Aspek keterampilan siswa pada hasil penempatan bola.	8	36,36

N=22

Dari data yang diperoleh menunjukkan bahwa sebagian guru masih berpendapat bahwa keterampilan siswa dalam memainkan bola atau penguasaan konsep dalam keterampilan bermain menjadi prioritas pertama yang akan dicapai dalam pembelajaran penjasorkes materi permainan bolavoli. Bahkan, di antaranya masih berpendapat pula bahwa bermain bolavoli dan keterkaitannya dengan berbagai aspek nilai-nilai yang terkandung namun tampak seperti kerja sama, sportivitas sebagai prioritas utama.

Meskipun demikian, berdasarkan tanggapan para guru dapat disimpulkan bahwa penguasaan konsep gerak siswa meliputi menerima bola, mengumpan, memukul bola dan membendung bola yang dilakukan melalui berbagai pengalaman teknik atau cara memainkan bola saat bermain ternyata tetap menjadi prioritas yang paling banyak dipilih

oleh para guru. Selanjutnya, hal itu diikuti oleh prioritas guru dalam mencapai hasil belajar yang berupa unsur kerja sama siswa, *sportivitas* siswa dalam bermain bolavoli.

Keberhasilan pencapaian kompetensi siswa akan sangat ditentukan oleh pengalaman dan pengetahuan guru dalam mengembangkan, dan menggunakan alat ukur yang telah dikonstruksi itu dengan cara yang benar, serta kemampuan menganalisis informasi yang dihasilkan oleh alat ukur itu. Berkaitan dengan pengalaman guru dalam pengembangan perangkat instrumen yang digunakan dalam melakukan penilaian praktik bolavoli selama ini dapat dilihat melalui jumlah respon, persentase, dan prioritas pilihan sebagaimana terangkum dalam Tabel 3.

Tabel 3. Penggunaan Perangkat Penilaian

Urutan	Perangkat Penilaian	Jumlah Respon	%
1	Kisi dan tugas praktik	18	81,81
2	Lembar pengamatan	4	18,18
3	Rubrik penilaian	0	0
4	Prosedur Penilaian	0	0

N=22

Dari data yang diperoleh menunjukkan bahwa sebagian besar guru menyusun kisi dan suatu tes praktik sebelum dilakukan penilaian. Lembar pengamatan, rubrik penilaian dan prosedur penilaian tidak dibuat karena tes praktik yang digunakan oleh guru diadopsi dari tes-tes simulasi yang sudah memiliki perangkat penilaian berupa prosedur penilaian.

Meskipun demikian, berdasarkan tanggapan para guru dapat disimpulkan bahwa selama ini penguasaan guru paling utama hanya menyusun kisi dan tes praktik. Hal ini disebabkan karena pemahaman tentang pengembangan rubrik penilaian dan prosedur penilaian belum banyak dimengerti.

Hasil wawancara pada guru terkait dengan metode penilaian untuk mengukur kompetensi siswa pada materi permainan bolavoli yang digunakan oleh guru antara lain: tes praktik, pengamatan, tes tertulis, portofolio, penugasaan individu atau kelompok, tes lisan, dan jurnal. Berdasarkan pengalamannya, pada kenyataannya setiap guru

memiliki prioritas yang berbeda terhadap ketujuh metode penilaian tersebut. Tanggapan para guru terhadap metode penilaian terangkum dalam Tabel 4.

Tabel 4. Metode Penilaian

Urutan	Metode	Jumlah Respon	%
1	Tes praktik (tes kinerja)	15	68,18
2	Pengamatan (observasi)	-	-
3	Tes tertulis	7	31,82
4	Portofolio	-	0
5	Penugasaan individu atau kelompok	-	0
6	Tes lisan	-	0
7	Jurnal	-	0

N=22

Berdasarkan data hasil wawancara dengan guru dapat disimpulkan bahwa, pada umumnya para guru menempatkan metode penilaian permainan bolavoli yang mencakup metode; tes praktik, dan tes tertulis sebagai prioritas pertama yang perlu diperhatikan. Metode pengamatan, tes portofolio, penugasaan individu, tes lisan dan jurnal tidak sama sekali digunakan oleh guru.

Berdasarkan karakteristik yang dinilai dan instrumen yang digunakan guru dalam melakukan penilaian praktik bolavoli guru menggunakan mengadopsi tes keterampilan bolavoli dari berbagai macam tes. Informasi tersebut dapat disajikan pada Tabel 5.

Tabel 5. Penggunaan Instrumen

Urutan	Instrumen	Jumlah Respon	%
1	Kemampuan siswa melakukan teknik servis menggunakan <i>serving accuracy test</i> dari AHPPER	5	22,72
2	Kemampuan siswa melakukan <i>passing</i> digunakan Brumbach <i>forearm pass wall-volley test</i> dan Kautz volleyball passing test	5	22,72
3	Kemampuan siswa melakukan <i>spike</i> menggunakan Stanley <i>spike test</i>	5	22,72
4	Kemampuan siswa bermain bolavoli dengan menggunakan <i>brady volley test</i>	7	31,81

N=22

Dari data yang diperoleh menunjukkan bahwa sebagian besar guru menggunakan tes keterampilan tersebut untuk menilai keterampilan siswa pada masing-masing teknik atau cara memainkan bola dalam bolavoli. Pelaksanaannya berupa simulasi secara terpisah pisah antar-*item* butir sehingga dalam melakukan penilaian, guru belum menggunakan perangkat penilaian sebagai satu sistem penilaian yang menyatu.

#### Data Pengembangan Produk

Pengembangan ini bertujuan sebagai tindak lanjut dari hasil analisis kebutuhan guru penjasorkes yang membutuhkan instrumen yang bisa mengakses hasil belajar penjasorkes pada permainan bolavoli di SMA. Langkah yang ditempuh peneliti dalam tahap pengembangan ini antara lain mencakup kegiatan perencanaan dan perancangan produk, dan kegiatan *focus group discussion* (FGD) sehingga pada akhir kegiatan ini dihasilkan produk berupa instrumen sistem asesmen hasil belajar materi permainan bolavoli pada mata pelajaran penjasorkes di SMA.

#### Data Uji Coba Produk

Dalam tahap ini dilakukan uji coba lapangan dan sekaligus revisi produk produk instrumen asesmen hasil belajar siswa penjasorkes pada permainan bolavoli di SMA. Secara berturut-turut dilakukan dua kegiatan pokok yaitu pra-ujicoba dan uji coba instrumen yang dikembangkan.

#### Data Validitas Instrumen

Validitas adalah suatu ukuran untuk menunjukkan sejauhmana suatu alat pengukur itu dapat mengukur apa yang hendak diukur. Instrumen asesmen hasil belajar siswa penjasorkes pada permainan bolavoli ini menggunakan validitas isi.

Validitas isi dimaksudkan untuk mengukur sejauh mana butir-butir instrumen yang digunakan mencakup keseluruhan isi objek yang diukur. Validitas isi instrumen penelitian ini dilakukan dengan menggunakan analisis rasional yaitu apakah butir pada kisi-kisi instrumen menggambarkan indikator dari variabel yang diukur dari unjuk kerja dalam penelitian. Adapun langkah-langkah yang di-

lakukan dalam validitas isi adalah melaksanakan FGD tahap II dengan para pakar ahli untuk memeriksa isi instrumen secara sistematis untuk mengevaluasi relevansi dengan variabel yang telah ditentukan.

Langkah ini telah dilakukan pada tahap validasi untuk mereview produk awal berupa perangkat instrumen. Berdasarkan hasil dari FGD tahap II menurut pendapat para ahli tentang kisi-kisi unjuk kerja siswa bermain bolavoli yang dikembangkan sudah mencerminkan materi yang terdapat dalam kurikulum KTSP SMA khususnya kelas XI/2 dan profil unjuk kerja siswa. Butir-butir keterampilan dan kemampuan yang disusun pada pedoman pengamatan sudah mencerminkan penguasaan kompetensi siswa sehingga validitas instrumen penilaian pengamatan unjuk kerja bermain bolavoli yang dikembangkan termasuk kategori baik.

#### Data Reliabilitas Instrumen

Bagian ini mendeskripsikan tentang hasil uji coba penggunaan perangkat instrumen penilaian yang diujicobakan kepada tiga orang guru sebagai penilai terhadap unjuk kerja siswa bermain permainan bolavoli. Indikator yang digunakan sebagai acuan untuk melakukan *rating* oleh para penilai telah diperoleh dari hasil pengembangan pada tahap sebelumnya.

Pedoman pengamatan tugas unjuk kerja siswa bermain bolavoli 6 lawan 6 ini terdiri atas empat indikator yakni: (1) penilaian keterampilan persiapan siswa saat akan memainkan bola; (2) penilaian keterampilan proses teknik memainkan bola; dan (3) penilaian keterampilan hasil penempatan bola; dan (4) penilaian kemampuan siswa menampilkan sikap dalam bermain. Hasil analisis *G study* digunakan untuk mengetahui koefisien reliabilitas alat penilaian yang dikembangkan serta estimasi komponen variansi kesalahan yang diakibatkan oleh berbagai sumber variansi, dalam pengembangan ini yakni sumber variansi siswa (P), penilai (I) dan *item* (I).

Setelah koefisien G dapat diketahui, maka pada tahapan analisis lanjut (analisis *D study*) akan didapatkan informasi tentang keputusan seberapa jauh penggunaan instrumen yang telah diuji memiliki keberlakuan

pada faset yang lebih luas terutama menyangkut kesamaan kondisi pengukuran, dan dapat diterimanya kondisi faset tersebut bagi penilai yang lain.

Hasil *G study* untuk mengetahui tingkat kebermaknaan penggunaan instrumen pengamatan pada unjuk kerja siswa bermain bolavoli dari uji coba di lapangan dapat dirangkum pada Tabel 6.

Tabel 6. Rangkuman Hasil Koefisien *G Study* pada Unjuk Kerja Bermain Bolavoli

No	Komponen	Uji Faset siswa	Jumlah Item	Koefisien Genova	Rerata Koefisien G
1	Persiapan saat akan memainkan Bola	Pa	5	0,90*	0,90*
		Pi	5	0,91*	
2	Proses teknik memainkan bola	Pa	15	0,86*	0,84*
		Pi	15	0,83*	
3	Hasil penempatan bola	Pa	7	0,75*	0,73*
		Pi	7	0,71*	
4	Sikap dalam bermain	Pa	2	0,79*	0,83*
		Pi	2	0,87*	
5	Semua komponen	Pa	29	0,78*	0,82*
		Pi	29	0,83*	

\*) memenuhi syarat menurut kriteria standar minimal 0,70 (Linn,1995, p.106)

Koefisien *G* dari komponen-komponen unjuk kerja bermain bolavoli menunjukkan bahwa secara keseluruhan pengembangan instrumen asesmen hasil belajar siswa penjasorkes pada permainan bolavoli dapat diterima untuk digunakan melakukan penilaian pada faset yang lebih luas. Dengan kata lain telah memenuhi untuk kepentingan faset pengukuran yang berkaitan dengan objek pengukuran pada unjuk kerja siswa bermain bolavoli yakni ditunjukkan oleh indeks koefisien *G* sebesar 0,82. Koefisien *G study* yang diperoleh menunjukkan bahwa nilai koefisien *G Study* lebih besar dari kriteria yang ditentukan, yaitu 70.

Dengan demikian, bentuk instrumen pengamatan unjuk kerja siswa bermain bolavoli dapat digunakan sebagai acuan guru dalam melakukan penilaian pada faset yang lebih luas. Namun, jika dilihat dari varian kesalahan untuk masing-masing komponen

penilaian pada analisis *Genova* untuk unjuk kerja siswa bermain bolavoli masih memerlukan penyempurnaan dalam hal administrasi penyelenggaraan yakni harus meningkatkan keterampilan dan pengalaman agar diperoleh hasil pengukuran yang konsisten.

Konfirmasi data hasil uji coba dari hasil Anava, berikut ini disajikan hasil analisis koefisien *interrater*. Koefisien *interrater* merupakan salah satu sarana untuk melihat tingkat konsistensi atau keajegan antar penilai dalam memberikan *rating* terhadap unjuk kerja siswa bermain bolavoli. Untuk keperluan ini, peneliti menggunakan koefisien *Cohen's Kappa*. Rangkuman koefisien dapat ditunjukkan pada Tabel 7.

Tabel 7. Rangkuman Hasil Koefisien *K* (*Kappa*)

No	Komponen Bermain Bolavoli	Sasaran Uji	Koefisien Kappa	Rerat Koefisien (Kappa)
1.	Keterampilan Persiapan Saat akan Memainkan Bola	Pa	0,78	0,79
		Pi	0,80	
2.	Keterampilan Proses Teknik Memainkan Bola	Pa	0,79	0,78
		Pi	0,77	
3.	Keterampilan Hasil Penempatan Bola	Pa	0,80	0,80
		Pi	0,80	
4.	Kemampuan Menampilkan Sikap Dalam Bermain Siswa	Pa	0,81	0,79
		Pi	0,78	
5	Semua komponen	Pa	0,79	0,79
		Pi	0,78	

Tabel 7 dapat memberikan gambaran bahwa hasil analisis kesepakatan dan kesepahaman *rater* terhadap konstruk instrumen pada empat komponen penilaian siswa dalam bermain dengan teknik *Kappa* pada siswa putera diperoleh koefisien sebesar 0,79. Berdasarkan besarnya koefisien reliabilitas antar-*rater* yang diperoleh hasilnya memenuhi syarat reliabilitas.

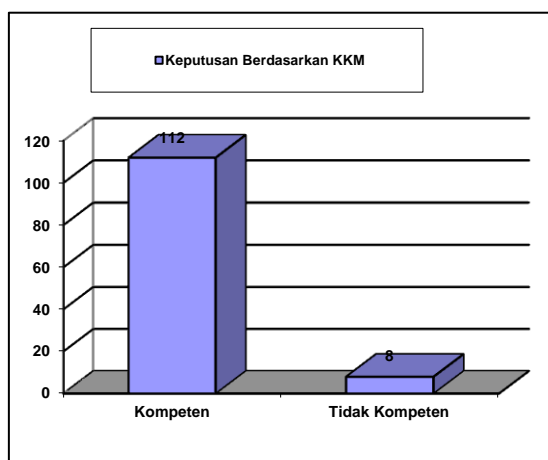
#### Data Penerapan Prosedur Asesmen

Prosedur pelaksanaan asesmen hasil belajar siswa penjasorkes pada permainan bolavoli yaitu alur atau langkah-langkah yang dilakukan dalam melakukan penilaian mencakup (1) kriteria *rater* atau penilai; (2) in-

strumen penilaian yang digunakan meliputi: (a) pedoman pengamatan mencakup indikator dan deskripsi unjuk kerja; (b) tugas unjuk kerja bermain bolavoli 6 lawan 6 dibatasi tiga putaran *server* pertama; (c) rubrik adalah *rubric* sebagai suatu alat penskoran yang terdiri dari daftar seperangkat kriteria beserta bobot penilaiannya dan skala penskoran 1-4 yang digunakan untuk mengamati hasil unjuk kerja siswa pada pedoman pengamatan atau observasi; (3) interpretasi data hasil unjuk kerja siswa yaitu informasi yang menyatakan tentang penguasaan kompetensi. Hasil wawancara dan pendapat para guru penjasorkes berkaitan dengan penerapan penggunaan prosedur pelaksanaan asesmen hasil belajar siswa penjasorkes pada permainan bolavoli menyatakan dapat dengan mudah digunakan untuk mengukur kemampuan dan keterampilan siswa dalam permainan bolavoli.

*Data Profil Hasil Belajar Siswa Penjasorkes pada Permainan Bolavoli di SMA*

Profil hasil belajar siswa penjasorkes pada permainan bolavoli ini berdasarkan pada data hasil penilaian pengamatan pada tes unjuk kerja siswa bermain bolavoli 6 lawan 6 sehingga pada bagian ini akan dideskripsikan profil unjuk kerja siswa bermain pada empat aspek yaitu: 1) keterampilan persiapan saat akan memainkan bola; 2) keterampilan proses teknik memainkan bola; 3) keterampilan hasil penempatan bola; dan (4) kemampuan menampilkan sikap dalam bermain.



Gambar 1. Profil Hasil Belajar Siswa Penjasorkes Pada Permainan Bolavoli SMA

Gambar 1 menunjukkan bahwa prestasi yang dicapai dari 120 siswa dalam unjuk kerja bermain bolavoli pada mata pelajaran penjasorkes berdasarkan perolehan skor adalah 112 orang siswa memperoleh skor  $\geq 70$  masuk dalam kategori kompeten dan 8 siswa memperoleh skor  $< 70$  masuk dalam kategori tidak kompeten.

Berdasarkan pada hasil analisis deskriptif dapat disimpulkan bahwa asesmen hasil belajar siswa penjasorkes pada permainan bolavoli di SMA DIY menunjukkan 120 orang siswa dinyatakan kompeten sebanyak 93,3% atau sebanyak 112 orang siswa. Siswa yang dinyatakan tidak kompeten sebesar 6,7% atau sebanyak 8 orang siswa

**Simpulan**

Berdasarkan hasil analisis data dan pembahasan, maka dapat diajukan simpulan sebagai berikut: (1) bentuk instrumen asesmen hasil belajar siswa penjasorkes pada permainan bolavoli di SMA adalah pedoman pengamatan (*observation*) yang mencakup: (a) indikator; (b) deskripsi indikator; (c) rubrik dan; (d) prosedur asesmen; (2) validitas isi instrumen asesmen hasil belajar siswa penjasorkes pada permainan bolavoli berdasarkan *expert judgement* termasuk kategori baik; (3) reliabilitas instrumen asesmen hasil belajar siswa penjasorkes pada permainan bolavoli dianalisis dengan menggunakan paket program *Genova* menghasilkan koefisien *inter-rater* sebesar 0,82, dan *Cohen's Kappa* menghasilkan koefisien sebesar 0,79, keduanya memenuhi syarat reliabilitas; (4) prosedur pelaksanaan asesmen hasil belajar siswa penjasorkes pada permainan bolavoli berdasarkan pendapat para guru penjasorkes dapat digunakan; (5) profil hasil belajar siswa penjasorkes pada permainan bolavoli di Sekolah Menengah Atas menunjukkan bahwa 112 siswa (93%) dinyatakan kompeten dan sebanyak 8 siswa (6,7%) dinyatakan tidak kompeten.

**Saran**

Berdasarkan hasil penelitian ini disarankan sebagai berikut: (1) untuk mendapatkan informasi hasil belajar siswa penjasorkes pada permainan bolavoli yang otentik,

instrumen asesmen hasil belajar ini dapat diterapkan oleh para guru dalam pembelajaran penjasorkes di SMA; (2) untuk mengembangkan instrumen asesmen pembelajaran penjaskesorkes lebih lanjut, pihak sekolah dan atau guru dapat mengembangkan instrumen asesmen yang lain berdasarkan kebutuhan dan kompetensi dasar yang ada; (3) penilai dalam melakukan asesmen pada mata pelajaran penjasorkes harus memiliki bukti *inter-rater* yang baik.

### Daftar Pustaka

- Asmawi, Moch. (2006). Dimensi pembelajaran keterampilan gerak dalam pendidikan jasmani. *Jurnal Majalah Olahraga*. Yogyakarta: Fakultas Ilmu Keolahragaan Universitas Negeri Yogyakarta. Edisi Agustus.
- American Sport Education Program. (2007). *Coaching youth volleyball*. Human Kinetics: USA.
- Arter, J. (1996). *Performance criteria: the heart of the matter*. Ed. R. E. Blum & J. A. Arter. A handbook for student performance assessment. Alexandria, VA: Assosiation for Supervision and Curriculum Development.
- BSNP. (2007). *Panduan penilaian kelompok mata pelajaran jasmani olahraga dan kesehatan*. Jakarta: Depdiknas
- Borg & Gall. (1983). *Educational research: an introduction*. New York: Longman.
- Barbara L. Viera. (2004). *Step to success: Bola voli tingkat pemula*. Terjemahan oleh PT Raja Grafindo Persada. Jakarta.
- Berk, R.A. (1986). *Performance assessment*. Baltimore: The John Hopkins University Press.
- Bonnie Kenny & Cindy Gregory. (2006) *Volleyball step to success: Sportinstruction series*. USA: Human Kinetics.
- Brennan, R.L. (2006). *Educational measurement*. Westport: Praeger.
- Crum, Bart. (2006). Substantial view of the body, *Paper, In Service Training On Didactic Of Sport Games*. Seminar. Bandung: Universitas Pendidikan Indonesia.
- Depdiknas. (2003). *Pedoman kebusus pengembangan silabus dan penilaian mata pelajaran pendidikan jasmani*. Jakarta: Ditjendikdasmen.
- Els van der Werf. (2006). *An innovation perspective on internationalisation of higher education: The critical phase*. *Journal of Studies in International Education*, 3(1), 3-14.
- Grineski, S. (1989). Children, Games, and prosocial behavior: Insights and connections. *Journal of physical education, recreation and dance*, 60 (8), 20-25.
- Guskey, T. (1996b). *Alternative ways to document and communicate student learning*. Alexandria, VA: Association for supervision and curriculum development.
- Hayton, G. & Wagner, Z. (1998). Performance assessment in vocational education and training. *Australian and New Zealand Journal of Vocational Education Research*, vol 6, no. 1, pp.69–85.
- Hensley, L.D. (1997). Authentic skill assessment in physical education. *Journal of physical education, recreation and dance*, vol 7 pp.19-24.
- Hopper, T. (1998). *Teaching games centred games using progressive principles of play*. *Canadian Association for Health, Physical Education, Recreation and Dance Journal*, 64 (3), 4-7.
- Jennifer Beasley. (2009). *Assessing student learning: A common sense guide*. Bolton, MA: Anker Publishing
- Jewett, dan A. Nixon, J. (1995). *An Introduction to physical education* Philadelphia: Saunders College.
- Joel, Dearing. (2003). *Sport fundamentals series volley ball fundamentals a better way to learn the basic*. USA: Human Kinetics.
- Johnson, R.L., Penny, J.A., & Gordon, B. (2009). *Assessing performance: de-signing*,

- scoring, and validating per-formance task. London: The Guilford Pres.
- Kinda S. Lenberg. (2006). *Volleyball skills & drills*. American volleyball Coaches Association. USA: Human Kinetics.
- Keating, J. (2001). *Sportsmanship as a moral category*. Champaign: Human Kinetics.
- Linn, R. L., & Gronlund, N. E. (1995). *Measurement and assessment in teaching*. Seventh Edition. Englewood Cliff: Merrill, an imprint of Prentice Hall.
- Lund, J.L & Mary Fortman Kirk. (2002). *Performance based assessment for middle and high school physical education*. USA: Human Kinetics.
- Lutan, R. (2001). *Mengajar pendidikan jasmani pendekatan pendidikan gerak di Sekolah Dasar*. Jakarta: Direktorat Jendral Olahraga: Depdiknas.
- Mardapi, Djemari. (2012). *Pengukuran penilaian dan evaluasi pendidikan*. Yogyakarta: Nuha Litera.
- Metzler, M. W. (2005). *Instructional Models for Physical Education (2)* Arizona: Holcomb Hathaway Publishers.
- Mitchell, S. A. (1999). *Teaching games for understanding: Theory, research and practice*. Champaign. IL: Human Kinetics
- Nitko, J., Anton & Brookhart, Susan. M. (2007). *Educational assessment for students*. New Jersey: Merrill Prentice Hall.
- Nurhasan. (2001). *Tes dan pengukuran dalam pendidikan jasmani: prinsip-prinsip dan penerapannya*: Ditjen Pendidikan dasar dan Menengah.
- Retnowati, Trie Hartiti. (2009). *Pengembangan instrument penilaian seni lukis anak di Sekolah Dasar*. Disertasi. Tidak dipublikasikan Yogyakarta: Universitas Negeri Yogyakarta.
- Stiggins, R. (1997). *The design and development of performance assessments. Educational Measurement: Issues and Practice* 2nded. Upper Saddle River, NJ: Prentice Hall.
- Tomme, P. M & Wendt, J. C. (2003). Affective: Psycho-social of physical education. *Journal of Physical Education, Recreation & Dance*; Oct 2003;54, 8; Research Library pg.66.
- Veal, M. (2002). The role of assessment in secondary physical education: A pedagogical view. *Journal of Physical Education, Recreation & Dance*, 63(7), 88-92.
- Vicki Worrell. (2002). Assessing the cognitive and affective progress of children. *Journal of Physical Education, Recreation & Dance*; Sep; 2002; 73, 7; Research Library pg.29.
- Yudiana, Y. & Subroto, T. (2010). *Permainan bolavoli*. Bandung: Fakultas Pendidikan Olahraga dan Kesehatan Universitas Pendidikan Indonesia.
- Wiggins, G. (1991). Standards, not standardization Evoking quality student work. *Journal of Physical Education, Recreation & Dance*; April 1991; 48(5) Research Library pg.18-25.
- Wiggins, G. (1991). Standards, not standardization Evoking quality student work. *Journal of Physical Education, Recreation & Dance*; April 1991; 48(5) Research Library pg.18-25.
- Wiggins, G. (1993). *Assessing student performance*. San Francisco: Jossey Bass Publishers.
- Wuest, D. & Bucher, C. (2009). *Foundations of physical education, Exercise science and sport (16<sup>th</sup>.Ed.)*. New York: McGraw-Hill.
- Zainul, Asmawi. (2005). *Alternative Assessment*. Jakarta: Dirjen Dikti.

## PENGEMBANGAN MODEL EVALUASI PROGRAM LAYANAN PENDIDIKAN ANAK USIA DINI (PAUD)

<sup>1)</sup>I Wayan Gunartha, <sup>2)</sup>Badrun Kartowagiran, <sup>3)</sup>Siti Partini Suardiman

<sup>1)</sup>IKIP PGRI Bali, <sup>2,3)</sup>Universitas Negeri Yogyakarta

<sup>1)</sup>[w.gunartha@yahoo.com](mailto:w.gunartha@yahoo.com), <sup>2)</sup>[badrunkw@yahoo.com](mailto:badrunkw@yahoo.com)

### Abstrak

Penelitian ini bertujuan: (1) mengembangkan model evaluasi program layanan PAUD (TK kelompok B) dan (2) menilai keefektifan model yang dikembangkan. Penelitian ini merupakan *research and development*. Subjek coba adalah guru dan siswa TK kelompok B. Instrumen pengumpul data adalah kuesioner dan lembar observasi. Uji coba dilakukan 3 tahap dengan jumlah subjek meningkat setiap tahapnya. Analisis validitas kuesioner menggunakan *exploratory factor analysis* dan *confirmatory factor analysis (CFA)*. Estimasi reliabilitas kuesioner tingkat perkembangan anak dan *life skills* menggunakan *CFA* dengan menghitung *composite reliability*. Estimasi reliabilitas kuesioner kompetensi guru dan fasilitas menggunakan teknik *Alpha*. Estimasi reliabilitas lembar pengamatan dengan dua orang pengamat menggunakan teknik *Kappa*. Hasil penelitian berupa model evaluasi program layanan PAUD (Model IPPO), yang terdiri atas prosedur, panduan evaluasi, dan instrumen-instrumennya yang cukup baik. Berdasarkan uji coba, semua instrumen memiliki kecocokan model yang baik (*fit*), validitas konstruk, dan reliabilitas yang memenuhi persyaratan akademik. Menurut penilaian para praktisi dan pengguna model, model IPPO sangat efektif untuk diterapkan.

**Kata kunci:** *pengembangan, model evaluasi, pendidikan anak usia dini*

## DEVELOPING A PROGRAM EVALUATION MODEL OF EARLY CHILDHOOD EDUCATION (ECE) SERVICE

<sup>1)</sup>I Wayan Gunartha, <sup>2)</sup>Badrun Kartowagiran, <sup>3)</sup>Siti Partini Suardiman

<sup>1)</sup>IKIP PGRI Bali, <sup>2,3)</sup>Universitas Negeri Yogyakarta

<sup>1)</sup>[w.gunartha@yahoo.com](mailto:w.gunartha@yahoo.com), <sup>2)</sup>[badrunkw@yahoo.com](mailto:badrunkw@yahoo.com)

### Abstract

The study was to: (1) develop an evaluation model of early childhood education (Group-B Kindergarten); and (2) evaluate the effectiveness of the model being developed. The study is a research and development one. The subject of the research was the teachers and the students of Group-B Kindergarten. The data gathering instruments were questionnaires and observational sheet. The experiment was performed in 3 stages with increasing number of participants for each stage. To analyze the questionnaire's validity, the researchers employed exploratory factor analysis and confirmatory factor analysis (CFA). To estimate the reliability of the questionnaire of the children's developmental stage and life skills, the researchers employed CFA by calculating the composite reliability. To estimate the reliability of the questionnaire's of the teacher competence and the facility, the researchers employed Alpha technique. For the estimation of reliability toward the observational sheet with two observers the researchers employed Kappa technique. The result of the research was a model of evaluation of the early childhood education (IPPO Program), consisting of procedures, evaluation guideline and the well-designed instruments. Based on the experiment, all of the instruments have had a goodness of fit (*fit*), construct validity and reliability that meet all of the academic requirements. According to the evaluation from the practitioners and the model users, the IPPO model is very effective to be implemented.

**Keyword:** *development, evaluation model, early childhood education.*

## Pendahuluan

Kualitas sumber daya manusia (SDM) Indonesia oleh banyak kalangan dinilai masih rendah. Hal itu disebabkan oleh rendahnya mutu pendidikan di berbagai jenis dan jenjang pendidikan (Hadis dan Nurhayati B., 2010, p.2). Indikator rendahnya kualitas pendidikan tersebut antara lain: *Human Development Index (HDI)* Indonesia dari tahun ke tahun masih rendah (Kunandar, (2007, p.1; Janawi, 2011, pp.8-9). Selain itu, juga dapat dilihat dari mutu akademik antarbangsa melalui *Programme for International Student Assessment (PISA)*, siswa Indonesia masih pada peringkat bawah (Hayat dan Suhendra Yusuf 2010, p.12). Oleh karena itu, peningkatan kualitas SDM harus dilakukan dengan peningkatan mutu pendidikan.

Menurut Aqib (2011, p.16), salah satu faktor penyebab rendahnya mutu pendidikan adalah *raw input*-nya, yaitu calon siswa karena selama ini (sebelum tahun 2001) perhatian kita terhadap pendidikan bagi anak usia dini masih sangat minim. Hal ini juga dikemukakan oleh Anwar & Arsyad Ahmad (2007, p.1). Untuk itu, salah satu cara peningkatan kualitas SDM adalah dengan meningkatkan kualitas pendidikan anak usia dini (PAUD) karena usia dini merupakan periode awal yang paling penting dan mendasar sepanjang rentang pertumbuhan serta perkembangan kehidupan manusia (*golden ages*). Masa ini oleh para ahli sering disebut masa peka, masa kritis, dan lain-lain.

Konsep tersebut di atas didasarkan fakta yang ditemukan oleh ahli-ahli *neurologi*. Menurut Woolfolk (2007, p.23), kira-kira satu bulan setelah terjadinya pembuahan, perkembangan otak sudah mulai. Sel-sel neuron muncul dengan kecepatan luar biasa, yakni 50.000 sampai 1.000.000 per detik selama kira-kira tiga bulan berikutnya. Ketika lahir, kita telah memiliki sekitar 100-200 miliar neuron yang terus bertambah dan saling berhubungan kalau mendapat stimulasi dari lingkungan. Hal yang hampir sama juga dikatakan oleh Berk (2007, p.121), serta menambahkan bahwa kerumitan sambung-

an antarneuron ini akan menentukan tingkat kecerdasan anak. Sambungan antarneuron (sinaps) yang jarang digunakan akan mati, sedangkan yang sering digunakan akan semakin kuat dan permanen. Untuk itu, diperlukan stimulasi yang tepat dari lingkungan.

Konsep di atas menekankan bahwa betapa pentingnya stimulasi untuk mengoptimalkan perkembangan otak untuk meningkatkan kecerdasan anak di masa dewasa. Dengan demikian, pendidikan di usia dini diharapkan dapat memberikan kontribusi secara signifikan terhadap upaya peningkatan kualitas SDM, yang pada gilirannya akan menjadikan bangsa kita sebagai bangsa yang berkualitas tinggi dan penuh daya saing di masa yang akan datang. Hal ini dibuktikan secara empiris, oleh banyak ahli seperti Ashiabi (2007, pp.205-206); Samuelsson (2011, p.109); Mann & Reynolds (2006, p.153); Beard & Sugai (2004, p.408), yang menegaskan pendidikan anak usia dini memegang peranan penting bagi kehidupan anak.

Mengingat pentingnya pendidikan anak pada usia dini, maka perhatian pemerintah untuk mengembangkan PAUD semakin besar. Selain itu, masyarakat juga menunjukkan kepedulian terhadap masalah pendidikan, anak usia dini dengan menyediakan berbagai layanan sesuai dengan kondisi dan kemampuan masing-masing. Hal seperti ini terjadi di mana-mana, termasuk di Bali, khususnya di Kabupaten Badung. Pada tahun 2009, diterbitkan Permendiknas No. 58 Tahun 2009 tentang Standar Pendidikan Anak Usia Dini. Penyusunan standar ini dimaksudkan untuk memberikan pelayanan yang berkualitas sesuai dengan kebutuhan pertumbuhan dan perkembangan anak.

Walaupun demikian, masih banyak permasalahan yang ada di lapangan termasuk di Kabupaten Badung, Provinsi Bali. Masih banyak anak yang belum memperoleh layanan PAUD, terutama kelompok bermain (KB). Hal tersebut sesuai dengan apa yang diungkapkan oleh Direktur Jenderal Pendidikan Anak Usia Dini, Nonformal, dan Informal (PAUDNI) bahwa dari 28, 8 juta anak usia 0-6 tahun pada akhir tahun 2009,

yang memperoleh layanan PAUD baru sekitar 53,7% karena terbatasnya lembaga PAUD yang ada (Kemendiknas, 2011, p.iii).

Penelitian Hiryanto, dkk. (2004) tentang pemetaan tingkat pencapaian mutu program PAUD di Provinsi DIY menunjukkan bahwa dilihat dari kesesuaian pedoman penyelenggaraan pendidikan anak usia dini dengan kondisi riil penyelenggaraan berdasarkan 10 patokan program dikmas, ditemukan hal-hal sebagai berikut. Adanya variasi dalam implementasi penyelenggaraan pendidikan. Adanya pengelompokan usia yang tidak sesuai pedoman karena keterbatasan sarana dan prasarana dan tenaga pendidik. Masih ada pendidik yang belum memperoleh pelatihan. Rasio jumlah pendidik dan peserta didik belum ideal.

Selain hal-hal di atas, ketersediaan alat permainan edukatif (APE) serta sarana dan prasarana juga bervariasi, ada yang sama sekali minim dengan sarana, yang penting program berjalan. Ada juga yang lengkap, terutama PAUD yang diselenggarakan yayasan. Akan tetapi, jika dilihat rasio alat yang dimiliki dengan jumlah peserta didik, peralatan yang ada belum memadai. Oleh karena itu, anak biasanya berebut dalam menggunakan alat permainan dan anak yang kurang berani biasanya terus mengalah.

Penelitian yang dilakukan oleh Her-mawati (2007) di TPA Beringharjo Yogyakarta menemukan dua kelemahan dari variabel *input*, yaitu kualifikasi pendidikan pengajar dan pengasuh yang tidak relevan dengan bidang tugas. Pada variabel proses, permasalahannya adalah belum terukurnya kegiatan pendampingan oleh pengasuh. Hal ini terkait dengan pendidikan pengasuh yang mayoritas rendah. Selain itu, belum dilakukannya pendampingan secara teratur oleh penyelenggara serta akses masyarakat terhadap TPA Beringharjo terbatas karena terbatasnya daya tampung.

Berdasarkan survai awal yang dilakukan pada bulan Agustus 2011 pada lembaga PAUD di Kabupaten Badung, Provinsi Bali, juga ditemukan masih ada masalah yang terkait dengan implementasi layanan PAUD. Masalah tersebut antara lain: kualitas dan

kuantitas guru PAUD masih relatif rendah karena sebagian besar guru berpendidikan D-II PGTK dan jumlah pendidik rata-rata 3-4 orang. Kesejahteraan mereka juga belum diperhatikan. Sarana dan prasarana belajar masih terbatas, terutama alat permainan edukatif di dalam ruangan. Selain jenisnya terbatas, juga jumlah untuk tiap jenis tidak cukup untuk semua anak, kecuali di PAUD yang dikelola yayasan yang sudah besar.

Jumlah anak yang belum mengikuti PAUD masih cukup besar, terutama anak usia di bawah 4 tahun karena jumlah kelompok bermain terbatas. Di enam kecamatan, ada 65 kelompok bermain. Demikian juga taman penitipan anak (TPA), hanya ada delapan buah, dan satuan PAUD sejenis (SPS) berjumlah empat buah. Jumlah PAUD antara di pedesaan dan di perkotaan belum merata, terutama kelompok bermain (KB) karena di kota lebih banyak tersedia layanan PAUD (TPA, KB, TK), sedangkan di desa-desa pada umumnya hanya ada TK, bahkan ada yang hanya memiliki TK kelompok B saja.

Dari segi proses, TK sudah diajar membaca, menulis, dan berhitung karena menurut pendidik, jika hal itu tidak diajarkan, maka lembaganya tidak mendapat siswa. Kegiatan bermain juga kurang optimal karena alat permainan biasanya terbatas. Hasil pencapaian perkembangan belum pernah dievaluasi oleh pihak dinas pendidikan sehingga belum diketahui secara pasti. Demikian juga, informasi anak di rumah tidak pernah dilaporkan oleh orang tua kepada guru. Hal ini penting sebagai bentuk kerja sama antara keluarga atau masyarakat dengan lembaga dalam memantau perkembangan anak.

Agar layanan pendidikan anak usia dini (PAUD) berkualitas, sesuai dengan standar yang telah ada, maka perlu ada evaluasi program layanan PAUD. Menurut Nugraha (2010, p.3), layanan pendidikan yang berkualitas adalah layanan yang secara terus-menerus dievaluasi dan hasilnya ditindaklanjuti secara tepat. Hal seperti itu juga dikemukakan oleh Mardapi (2012, p.12), bahwa peningkatan kualitas pendidikan dapat di-

tempuh melalui peningkatan kualitas pembelajaran, melalui peningkatan sistem evaluasi. Dengan demikian, untuk memperbaiki kualitas pendidikan di TK, perlu adanya model evaluasi program layanan PAUD yang dapat digunakan untuk mengevaluasi program layanan PAUD secara terus menerus.

Selama ini, evaluasi terhadap program layanan PAUD secara internal belum dilakukan secara menyeluruh. Demikian juga, di Bali. Bahkan, di Kabupaten Badung, Provinsi Bali, berdasarkan survei awal yang telah dilakukan pada bulan Agustus 2011, pihak Dinas Pendidikan belum pernah melakukan evaluasi terhadap program layanan PAUD yang ada. Penentuan kualitas lembaga PAUD sering didasarkan pada seringnya PAUD mengikuti lomba dan banyaknya peserta didik. Hal ini disebabkan oleh belum ada model evaluasi program yang komprehensif termasuk format instrumen maupun panduan yang baku yang dapat diterapkan oleh Kepala TK sebagai evaluasi internal maupun oleh pihak Dinas Pendidikan.

Selama ini, evaluasi program yang telah dilakukan adalah akreditasi, akan tetapi, belum semua TK pernah terakreditasi. Selain itu, bila dicermati lampiran 1 Peraturan Menteri Pendidikan Nasional Nomor 52 Tahun 2009, tentang Instrumen Akreditasi TK/RA, tampak masih ada beberapa kelemahan, seperti butir instrumennya terbatas, indikator yang digunakan tidak jelas. Oleh karena itu, perlu ada model yang lain untuk melengkapinya. Untuk keperluan itulah, dipandang penting untuk mengembangkan model evaluasi program layanan PAUD yang komprehensif dan akurat.

Layanan pendidikan sebagai suatu sistem atau suatu program, terdiri atas komponen-komponen yang saling terkait dan saling mempengaruhi satu sama lain. Komponen tersebut adalah input, proses, dan produk. Dalam mengevaluasi program layanan PAUD, ketiga komponen inilah yang harus dievaluasi. Selain itu, menurut Sujiono (2009, p.42), urgensi pendidikan anak usia dini berdasarkan tinjauan didaktis psikologis adalah untuk mengembangkan

berbagai aspek kecerdasan yang merupakan potensi bawaan. Kecerdasan yang dimiliki oleh seorang anak hanya akan berarti apabila dapat diterapkan dalam kehidupan sehari-hari, yang dikenal dengan istilah kecakapan hidup (*life skill*). Dalam model ini, *life skill* disebut *outcome* (Kaufman & Thomas, 1980, p.42).

Berdasarkan uraian tersebut, model yang akan dikembangkan dalam penelitian ini adalah model evaluasi program layanan PAUD, yang meliputi: (1) evaluasi input mencakup komponen: evaluasi kelengkapan sarana belajar/bermain, serta kualifikasi pendidikan dan kompetensi guru; (2) evaluasi proses, mencakup perencanaan pembelajaran, pelaksanaan pembelajaran, dan proses penilaian; (3) evaluasi produk, yaitu tingkat pencapaian perkembangan anak, yang mencakup perkembangan moral keagamaan, sosial-emosional, bahasa, kognitif, dan fisik-motorik; dan (4) evaluasi *outcome*, yaitu *life skills* anak TK. Model ini diberi nama Model IPPO, singkatan dari Model Input-Proses-Produk-Outcome.

Berdasarkan latar belakang tersebut, penelitian ini dilakukan dengan tujuan: (1) membangun sebuah model evaluasi program yang khusus digunakan untuk mengevaluasi program layanan pendidikan anak usia dini, khususnya taman kanak-kanak kelompok B sehingga dapat memberikan informasi yang lengkap dan akurat bagi pengelola program dan menilai keefektifan model evaluasi program layanan PAUD yang dibangun; (2) mengembangkan model evaluasi program layanan PAUD (TK kelompok B) dan; (3) menilai keefektifan model yang dikembangkan. Penelitian ini merupakan *research and development*.

Produk yang dihasilkan dalam penelitian ini adalah sebuah model evaluasi program layanan PAUD, khususnya taman kanak-kanak kelompok B. Model evaluasi program layanan PAUD tersebut terdiri atas prosedur evaluasi yang berupa langkah-langkah umum evaluasi; instrumen yang akan digunakan untuk mengukur input, proses, produk, dan *outcome (life skills)*; panduan yang digunakan dalam melakukan evaluasi; dan

bukti keefektifan model yang dikembangkan.

## Metode Penelitian

### Model Pengembangan

Penelitian ini merupakan penelitian dan pengembangan yang bertujuan menghasilkan produk berupa model evaluasi program layanan pendidikan anak usia dini (PAUD). Berdasarkan kajian terhadap model penelitian pengembangan dari literatur yang ada, dipilih model penelitian pengembangan (R & D) dari Borg & Gall (1983, p.775), dengan 10 langkah pengembangan yang disederhanakan menjadi 4 langkah, yaitu: (1) investigasi awal, (2) tahap disain, (3) uji coba, evaluasi, dan revisi, dan (4) implementasi.

Pada tahap awal, kegiatan yang dilakukan adalah survai pendahuluan, meninjau teori model-model evaluasi, pendidikan anak usia dini, serta mengkaji hasil-hasil penelitian yang telah dilakukan. Pada tahap desain, dirancang model evaluasi program layanan pendidikan anak usia dini, yang terdiri atas prosedur evaluasi, instrumen-instrumen, panduan, dan desain uji cobanya. Pada tahap uji coba, evaluasi, dan revisi, dilakukan uji coba terhadap model yang telah dirancang di taman kanak-kanak. Data hasil uji coba kemudian dianalisis. Apabila berdasarkan hasil analisis model evaluasi beserta instrumen ternyata belum baik, maka dilakukan direvisi dan diuji coba lagi sampai diperoleh *prototype* akhir yang memenuhi syarat *fit model* (*prototype* yang baik). Uji coba dilakukan tiga tahap. Pada tahap implementasi, prosedur evaluasi, instrumen, beserta panduan yang telah diujicobakan dan sudah baik selanjutnya diimplementasikan.

### Prosedur Pengembangan

Langkah-langkah yang ditempuh dalam pengembangan model evaluasi program layanan pendidikan anak usia dini adalah sebagai berikut.

#### *Penyusunan Desain dan Perangkat Model Evaluasi*

Pada tahap ini, disusun model evaluasi program layanan PAUD, yang terdiri atas prosedur, instrumen, dan panduan

evaluasinya. Model ini mengadopsi tiga komponen model CIPP, yaitu: *input*, *process*, dan *product evaluation* dari Stufflebeam. Tiga komponen CIPP ditambahkan dengan satu komponen, yaitu evaluasi *life skills*, yang dalam model ini disebut *outcome*. Dengan demikian, model evaluasi program layanan PAUD memiliki empat komponen utama, yaitu: input, proses, produk, dan *outcome* sehingga disebut model IPPO.

Evaluasi input meliputi aspek: sarana/fasilitas layanan, kompetensi guru, dan kualifikasi guru. Evaluasi proses meliputi aspek: perencanaan layanan, pelaksanaan layanan, dan proses penilaian. Evaluasi produk meliputi lima tingkat pencapaian perkembangan anak, yaitu: perkembangan moral-keagamaan, sosial-emosional, bahasa, kognitif, dan fisik-motorik. Evaluasi *outcome* yaitu *life skills* anak TK. Berdasarkan komponen tersebut, disusun instrumen-instrumennya, prosedur, serta panduan penggunaannya. Semua instrumen berbentuk skala *Likert* dengan 5 pilihan. Desain prosedur yang disusun, instrumen, dan panduan tersebut merupakan draf awal dari model yang dikembangkan.

#### *Validasi Pakar (Expert Judgement)*

Untuk menyempurnakan draf model dan memeriksa validitas isi instrumen, setelah prosedur evaluasi beserta instrumen dan panduannya disusun, dilanjutkan dengan validasi oleh para ahli yaitu akademisi atau dosen dan praktisi (guru TK), serta pengguna model (kepala/wakil kepala TK). Proses validasi pakar menggunakan model *FGD* (*focus group discussion*). Pelaksanaan *FGD* dilakukan dua tahap, yaitu tahap pertama *FGD* (*focus group discussion*) dengan 10 orang akademisi (dosen) dari Program pascasarjana UNY. Setelah model direvisi sesuai masukan para akademisi (dosen) dilanjutkan dengan *FDG* dan uji keterbacaan dengan 3 orang kepala TK dan 17 orang guru TK. Setelah uji keterbacaan dilanjutkan penilaian terhadap model.

#### *Uji coba*

Draf awal instrumen yang sudah direvisi berdasarkan masukan yang diper-

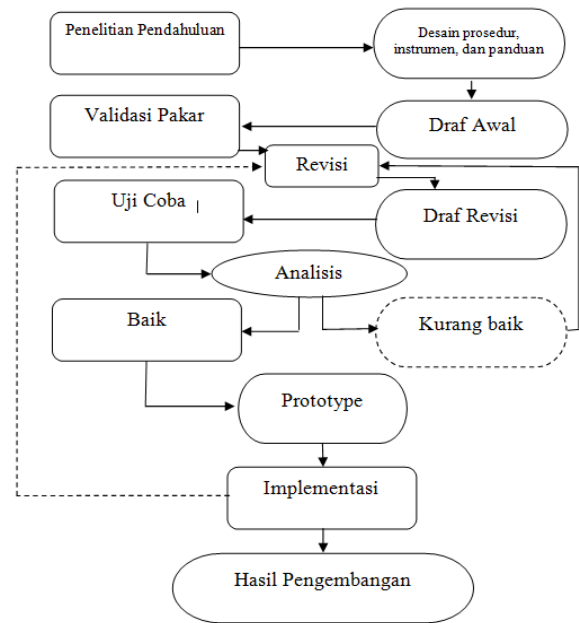
oleh dalam *FGD*, diujicobakan di taman kanak-kanak untuk mengetahui cocok/sesuai atau tidaknya model, validitas konstruk, dan reliabilitasnya. Uji coba instrumen dilakukan tiga tahap, yaitu tahap pertama, kedua, dan ketiga dengan jumlah subjek coba yang semakin meningkat. Jumlah TK berturut-turut: 10, 13, dan 18. Uji coba instrumen input menggunakan subjek coba berturut-turut 32, 60, dan 85 orang guru. Uji coba instrumen produk dan *outcome* menggunakan subjek coba berturut-turut 160, 260, dan 360 anak TK. Uji coba instrumen proses (lembar observasi) menggunakan subjek coba tiga, empat, dan lima orang guru berturut-turut.

*Analisis Data*

Data tentang kejelasan prosedur, kekomprehensifan komponen model, kejelasan instrumen, dan kejelasan panduan evaluasi dianalisis secara deskriptif. Data produk (tingkat perkembangan anak) dan data *outcome (life skills)* dianalisis dengan *Confirmatory Factor Analysis (CFA)* untuk mengetahui validitas dan reliabilitasnya, dengan program *Lisrel 8.8*. Data tentang input (kompetensi guru dan kelengkapan sarana), yang diperoleh melalui kuesioner dianalisis dengan *exploratory factor analysis (EFA)*, dengan program *SPSS 17.00 for Windows* untuk mengetahui validitas konstruksinya karena jumlah subjek tidak memenuhi syarat dianalisis dengan *CFA*. *Alpha Cronbach* digunakan untuk mengetahui reliabilitasnya. Data proses pembelajaran yang diperoleh melalui pengamatan dianalisis dengan *Kappa inter-rater reliability* untuk mengetahui reliabilitasnya.

*Implementasi*

Hasil akhir model yang telah dianalisis, yang merupakan *prototype* yang baik, diimplementasikan di beberapa 18 taman kanak-kanak (TK). Bila digambarkan dengan bagan, seluruh proses pengembangan model evaluasi program layanan pendidikan anak usia dini disajikan pada Gambar 1.



Gambar 1. Diagram Alur Prosedur Pengembangan Model IPPO

- : Proses kegiatan
- : Hasil Kegiatan
- : Hasil Analisis
- : Urutan Berikutnya
- : Peninjauan kembali untuk perbaikan

Desain Uji coba Produk

*Desain Uji coba*

Uji coba dimaksudkan untuk memperoleh data secara lengkap yang digunakan sebagai bahan revisi produk yang dihasilkan. Aspek yang diujicobakan dalam tahap uji-coba model ini adalah draf model evaluasi, yang terdiri dari prosedur evaluasi, instrumen, dan panduan evaluasi.

*Subjek Coba*

Subjek uji coba dalam penelitian ini adalah guru dan siswa TK kelompok B. Jumlah subjek uji coba meningkat dari tahap pertama, kedua, dan ketiga. Jumlah guru dalam uji coba adalah 32, 60, dan 85 orang berturut-turut. Jumlah siswa dalam uji coba adalah 160, 260, dan 360 orang berturut-turut. Jumlah guru untuk uji coba instrumen proses (lembar observasi) adalah tiga, empat, dan lima orang berturut-turut.

*Teknik dan Instrumen Pengumpul Data*

Dalam penelitian ini, instrumen pengumpul data terdiri atas: (a) angket/kuesioner, (b) lembar observasi, dan (c) dokumentasi. Angket digunakan untuk menjangkau data untuk variabel input, produk, dan *outcome*, serta kejelasan instrumen, prosedur, dan panduan. Lembar observasi digunakan untuk pengambilan data tentang proses layanan PAUD. Dokumentasi digunakan dalam pengumpulan data tentang perencanaan. Semua angket berbentuk *Skala Likert* dengan 5 pilihan.

*Teknik Analisis Data*

Teknik analisis data dalam penelitian ini menggunakan analisis data secara kuantitatif dan deskriptif kualitatif. Instrumen pengumpul data tentang *input* (fasilitas layanan dan kompetensi guru PAUD) menggunakan teknik *Exploratory Factor Analysis (EFA)* untuk mengetahui validitas konstruksinya. Reliabilitasnya dianalisis dengan teknik *internal consistency*, yaitu *Alpha Cronbach* untuk setiap dimensi dan selanjutnya dihitung reliabilitas kompositnya (Azwar, 2012, p.84).

Data tentang produk (perkembangan anak) dan *outcome (life skills)*, pemeriksaan validitas dan reliabilitas dilakukan dengan *Confirmatory Faktor Analysis (CFA)*. Penen-

tuan *goodness of fit* menggunakan beberapa indikator, yaitu: (a) nilai *chi-square* dengan *p-value*  $\geq 0,05$ , (b) *RMSEA*  $\leq 0,08$ , dan *GFI*  $\geq 0,9$  (Latan, 2012, p.53; Gozali & Fuad, 2008, pp.29-31). Reliabilitas konstruk dihitung dengan memperhatikan *construct reliability (CR)* berdasarkan *lamda* ( $\lambda$ ) masing-masing indikator, dan *error variance* ( $\delta$ ) indikator. Data tentang proses, yaitu perencanaan, pelaksanaan, dan pelaksanaan evaluasi, yang dikumpulkan dengan lembar observasi dianalisis dengan *inter-rater* untuk menganalisis reliabilitasnya. Teknik yang digunakan adalah *Cohen's kappa (K)* karena pengamat berjumlah dua orang. Analisis menggunakan program *SPSS for Windows 17.00*.

Pada analisis deskriptif-kualitatif, data kuantitatif yang diperoleh melalui instrumen penilaian terhadap prosedur, instrumen, dan panduan, dan keefektifan model dihitung skor reratanya, kemudian dikonversikan ke data kualitatif dengan skala 5 dan akhirnya dimaknai secara kualitatif. Hasil analisis kualitatif tersebut digunakan sebagai dasar menentukan model yang dikembangkan sudah baik atau tidak. Konversi data kuantitatif ke dalam data kualitatif dengan skala 5 menggunakan aturan yang merupakan modifikasi dari aturan yang dikembangkan oleh Sudijono (2011, p.329) seperti pada Tabel 1.

Tabel 1. Penilaian terhadap Prosedur, Instrumen, dan Pandua

Rerata Skor	Kualifikasi	Kesimpulan
> 4,2	Sangat baik	Dapat dijadikan contoh
> 3,4 – 4,2	Baik	Dapat digunakan tanpa perbaikan
> 2,6 – 3,4	Cukup Baik	Dapat digunakan dengan sedikit perbaikan
> 1,8 – 2,6	Kurang Baik	Dapat digunakan dengan banyak perbaikan
$\leq 1,8$	Tidak Baik	Belum dapat digunakan

**Hasil Penelitian dan Pembahasan**

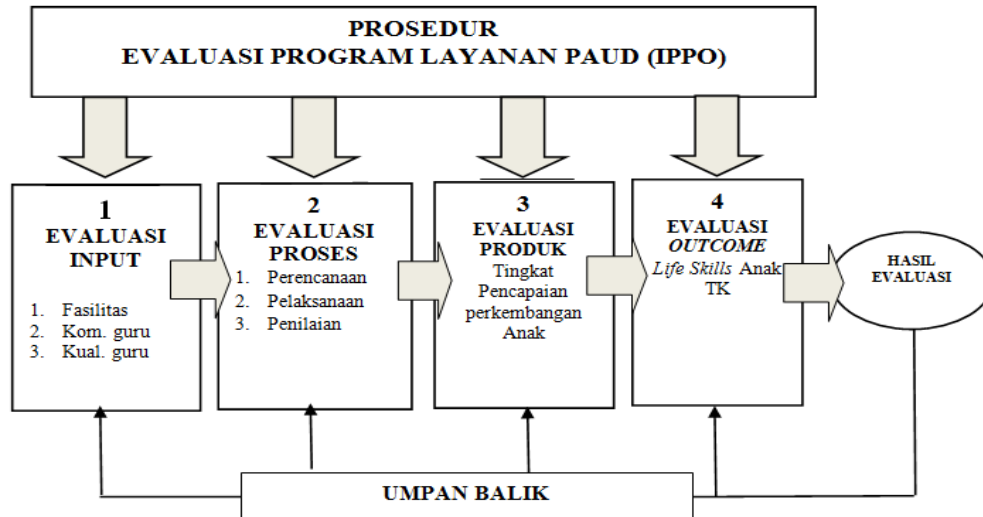
Prosedur Evaluasi Program Model IPPO

Model evaluasi program layanan PAUD (Model IPPO) terdiri atas empat komponen, yaitu: evaluasi input, evaluasi proses, evaluasi produk, dan evaluasi *outcome*. Dengan demikian, model ini disebut model

IPPO. Sesuai dengan komponen tersebut, prosedur evaluasi yang harus diikuti adalah sebagai berikut. Pertama dilakukan evaluasi terhadap input, yaitu fasilitas, kompetensi, dan kualifikasi guru. Kedua dilakukan evaluasi terhadap proses layanan, yang menyangkut perencanaan, pelaksanaan, maupun pelaksanaan penilaian oleh guru. Ketiga, di-

lakukan evaluasi terhadap produk, yaitu tingkat pencapaian perkembangan anak. Keempat adalah evaluasi *outcome*, yaitu *life skills* anak.

Jika digambarkan, Model dan prosedur evaluasi program layanan pendidikan anak usia dini (evaluasi model IPPO) adalah sebagai berikut.



Gambar 2. Prosedur Evaluasi Model IPPO

#### Panduan Evaluasi Program Model IPPO

Panduan evaluasi program layanan pendidikan anak usia dini (model IPPO) terdiri atas ketentuan umum, langkah-langkah evaluasi setiap komponen, pedoman penyekoran, pengolahan skor, waktu evaluasi, rekomendasi, dan format laporan hasil evaluasi.

#### Hasil Validasi Ahli dan Praktisi

Komponen model evaluasi yang divalidasi terdiri atas: (1) komponen dan prosedur evaluasi, (2) panduan penggunaan, dan (3) instrumen beserta kisi-kisinya. Tujuannya adalah untuk melihat kejelasan prosedur dan kelengkapan komponen, kejelasan panduan, dan kejelasan instrumen. Hasilnya sebagai berikut.

#### Kejelasan Prosedur Evaluasi

Prosedur dievaluasi dari kejelasan langkah-langkah evaluasi, kejelasan kalimat, ketepatan ejaan/tanda baca, kelengkapan penjelasan, kemudahan, kesederhanaan, dan efisiensi waktu, tenaga, dan biaya. Penilaian menggunakan skala 5 dengan skor terendah 1 dan skor tertinggi 5. Berdasarkan penilaian oleh para ahli, rerata skor total adalah

4,14. Sesuai dengan pedoman konversi, rerata tersebut termasuk kategori baik atau dapat digunakan tanpa perbaikan. Berdasarkan penilaian yang dilakukan guru dan kepala TK, diperoleh peningkatan hasil, yaitu diperoleh rerata total sebesar 4,31. Berdasarkan rerata skor tersebut, dapat disimpulkan prosedur evaluasi sangat baik dan sudah dapat digunakan dengan tanpa perbaikan. Jika dirata-ratakan rerata dari kedua kelompok penilai, diperoleh rerata total 4,22. Berdasarkan rerata skor tersebut dapat disimpulkan bahwa prosedur evaluasi sudah baik dan dapat digunakann tanpa perbaikan.

#### Kejelasan Instrumen Evaluasi Layanan PAUD

Penilaian terhadap kejelasan instrumen diarahkan pada tiga aspek utama, yaitu: (a) kejelasan petunjuk instrumen, (b) kelengkapan indikator instrumen fasilitas, kompetensi guru, kualifikasi guru, perencanaan, pelaksanaan, penilaian, tingkat pencapaian perkembangan anak, *life skills*; dan (c) kesesuaian indikator dengan butir (validitas isi), dan (d) penggunaan bahasa Indonesia. Penilaian menggunakan skala 5, skor terendah 1 dan tertinggi 5.

Berdasarkan seluruh rerata skor yang diberikan oleh para ahli, diperoleh rerata total = 4,1. Sesuai dengan dengan pedoman konversi, rerata itu ada pada interval > 3,4-4,2 dan masuk dalam klasifikasi baik. Dengan demikian, instrumen sudah baik dan dapat digunakan tanpa perbaikan. Berdasarkan penilaian yang dilakukan oleh para guru dan kepala TK, diperoleh rerata skor total 4,29. Rerata skor tersebut sesuai dengan kriteria penilaian di atas juga tergolong baik. Rerata skor total dari kedua kelompok penilai adalah 4,2. Hal tersebut berarti bahwa, instrumen sudah baik dan dapat dipergunakan tanpa perbaikan.

*Kejelasan Panduan Penggunaan*

Penilaian terhadap panduan juga dilakukan dua tahap, yaitu tahap pertama oleh para pakar dan tahap kedua oleh para praktisi (guru TK) dan pengguna model (kepala TK). Penilaian difokuskan pada aspek yang sama baik yang dilakukan oleh pakar maupun oleh para praktisi, yaitu kejelasan petunjuk umum, kejelasan langkah evaluasi, kejelasan rekomendasi, kejelasan waktu pelaksanaan, cara pengolahan skor,

kelengkapan isi panduan, kejelasan kalimat, penggunaan ejaan/tanda baca, ketepatan kata/istilah, dan bentuk/ukuran huruf. Penilaian menggunakan skala 5 dengan skor terendah 1 dan skor tertinggi 5.

Hasil penilaian terhadap panduan evaluasi adalah sebagai berikut. Rerata skor total dari para ahli diperoleh sebesar 4,22. Rerata skor tersebut sesuai pedoman konversi termasuk dalam kategori baik. Setelah panduan direvisi sesuai dengan masukan para ahli, dilakukan uji keterbacaan dan FGD dengan para guru dan kepala TK di lapangan. Hasil penilaian oleh para guru dan kepala TK ternyata menunjukkan adanya peningkatan, yaitu dengan rerata skor 4,32. Rerata skor tersebut termasuk dalam kategori sangat baik. Rerata dari kedua kelompok penilai adalah 4,27. Skor rerata tersebut menunjukkan bahwa panduan yang dikembangkan sudah baik dan dapat dipergunakan tanpa perbaikan.

Hasil penilaian terhadap prosedur, instrumen, dan panduan oleh para ahli dan praktisi, serta pengguna model, dapat dilihat pada Tabel 2.

Tabel 2. Rekapitulasi Hasil Validasi Pakar

No.	Komponen yang Divalidasi	Skor Rerata Total		Rerata	Kualifikasi
		Ahli	Praktisi		
1.	Prosedur Evaluasi	4,14	4,31	4,22	Sangat baik
2.	Instrumen	4,1	4,29	4,2	Baik
3.	Panduan Evaluasi	4,22	4,32	4,27	Sangat baik

Model Pengukuran Instrumen

*Instrumen Fasilitas dan Kompetensi Guru*

Instrumen penilaian sarana/fasilitas layanan dan kompetensi guru PAUD atau yang subjeknya guru pemeriksaan validitas dilakukan dengan analisis faktor eksploratori atau *exploratory factor analysis (EFA)* dan reliabilitasnya dihitung dengan teknik konsistensi internal (*internal consistency*), yaitu dengan melihat koefisien *Alpha Cronbach*, dengan bantuan program *SPSS 17.00 for Windows*. Hasil uji coba adalah sebagai beri-

kut. Sebelum dilakukan analisis faktor, dilakukan uji kelayakan apakah data layak dianalisis dengan analisis faktor atau tidak. Ternyata dari tiga tahap uji coba, semua data layak dianalisis faktor. Hal ini dapat dilihat bahwa semua harga *Kaiser-Meyer-Olkin (KMO)* lebih besar daripada 0,5. Selain itu semua hasil uji *Bartlett's test of Sphericity* signifikan ( $sig < 0,05$ ).

Muatan faktor untuk semua instrumen ketiga tahap uji coba semuanya lebih besar dari 0,5 sehingga tidak ada butir yang gugur. Semua instrumen memiliki koefisien

reliabilitas di atas 0,7 untuk ketiga tahap uji coba. Dengan demikian, dapat dikatakan bahwa instrumen penilaian input, yaitu instrumen fasilitas dan kompetensi guru telah

memiliki validitas dan reliabilitas yang baik. Dengan demikian, telah dapat digunakan untuk mengumpulkan data. Keseluruhan hasil analisis dapat dilihat pada Tabel 3.

Tabel 3. Rekapitulasi Validitas dan Reliabilitas Instrumen Input

No	Instrumen	Uji ke-	KMO	Factor Loading		Butir Gugur	Total Varians	Koefisien Alpha
				Terendah	Tertinggi			
1	Fasilitas layanan PAUD	1	0,623	0,636	0,858	-	66,875%	0,94
		2	0,747	0,638	0,841	-	61,67%	0,91
		3	0,726	0,627	0,829	-	62,39%	0,99
2	Kompetensi Kepribadian	1	0,648	0,70	0,845	-	68,72%	0,91
		2	0,830	0,683	0,858	-	72,06%	0,95
		3	0,812	0,623	0,859	-	64,73%	0,94
3	Kompetensi Profesional	1	0,732	0,728	0,911	-	78,96%	0,96
		2	0,779	0,630	0,902	-	70,94%	0,94
		3	0,846	0,671	0,913	-	78,09%	0,96
4	Kompetensi Pedagogik	1	0,668	0,521	0,917	-	68,21%	0,95
		2	0,815	0,567	0,817	-	63,75%	0,94
		3	0,804	0,534	0,894	-	64,38%	0,94
5	Kompetensi Sosial	1	0,692	0,694	0,890	-	66,71%	0,91
		2	0,778	0,653	0,863	-	61,71%	0,89
		3	0,807	0,695	0,892	-	71,57%	0,933

Catatan: Semua Uji Bartlett's test of Sphericity signifikan (sig < 0,05)

*Instrumen Proses Layanan*

Instrumen yang berupa lembar observasi/pengamatan, pemeriksaan validitas dilakukan dengan *expert judgement*, sedangkan pemeriksaan reliabilitasnya menggunakan teknik *inter-rater reliability*, yaitu dengan melihat koefisien *Kappa (Cohen's kappa)*.

Hasil ketiga tahap uji coba dapat dijelaskan sebagai berikut. Pada uji coba terbatas, dilakukan pengamatan terhadap tiga orang guru. Nilai *Kappa* diperoleh berturut-turut 0,72, 0,712, dan 0,743. Rerata dari ketiganya adalah 0,725 (baik). Pada uji coba kedua (uji coba utama), dilakukan pengamatan terhadap empat orang guru dan nilai *Kappa* berturut-turut adalah 0,797, 0,799,

0,795, dan 0,818. Reratanya adalah 0,8 (istimewa). Pada uji coba ketiga, dilakukan pengamatan terhadap lima orang guru. Nilai *Kappa* diperoleh berturut-turut 0,795, 0,778, 0,789, 767, dan 0,814. Reratanya adalah 0,786 (istimewa). Berdasarkan ketiga tahap uji coba diperoleh rerata nilai *Kappa* sebesar 0,771 (istimewa). Dengan demikian, dapat disimpulkan bahwa instrumen proses (lembar observasi) yang dikembangkan tergolong sangat baik. Dengan kata lain, lembar observasi yang dikembangkan dalam penelitian ini sangat reliabel sehingga dapat digunakan untuk mengumpulkan data. Keseluruhan hasil uji coba disajikan pada Tabel 4.

Tabel 4. Rangkuman Hasil Uji coba Instrumen Proses Layanan PAUD

No.	Uji coba	Jml TK	Nilai Kappa					Rerata Kappa	Kategori
			1	2	3	4	5		
1.	I	3	0,720	0,712	0,743	-	-	0,725	Baik
2.	II	4	0,797	0,799	0,795	0,818	-	0,80	Istimewa
3.	III	5	0,795	0,778	0,789	0,767	0,814	0,786	Istimewa
	Jumlah	12	-	-	-	-	-	2,313	-
	Rerata Total	-	-	-	-	-	-	0,771	Istimewa

*Instrumen Produk (Pencapaian Perkembangan Anak) dan Outcome (Life Skills)*

Instrumen penilaian tingkat pencapaian perkembangan anak TK dan *Life Skills* validitasnya dihitung dengan *confirmatory factor analysis (CFA)*, dengan bantuan program *Lisrel 8.8*, dan reliabilitas dihitung memperhatikan *construct reliability (CR)* berdasarkan nilai  $\lambda$  dan *error varians* indikator. Instrumen tingkat pencapaian perkembangan anak meliputi: pencapaian perkembangan moral-keagamaan, sosial-emosional, bahasa, kognitif, dan fisik-motorik.

Berdasarkan hasil analisis, semua butir pada semua instrumen dari tiga tahap uji-coba signifikan ( $t > 1,96$ ). Ini berarti bahwa semua butir dapat digunakan untuk mengukur konstruksinya dengan baik. Pada uji coba

ketiga ada beberapa butir instrumen tingkat pencapaian perkembangan bahasa dan *life skills* memiliki *factor loading* lebih kecil daripada 0,5, yaitu 0,49, 0,48. Karena mendekati 0,5, maka dibulatkan menjadi 0,5. Dengan demikian, semua instrumen memiliki validitas konstruk yang baik. Dilihat dari kecocokan model, pada uji coba ketiga semua persyaratan *model fit* telah terpenuhi, baik *p-value* ( $\geq 0,05$ ), *RMSEA* ( $\leq 0,08$ ), maupun *GFI* ( $\geq 0,9$ ). *Construct reliability (CR)* semua instrumen berada di atas 0,7 pada ketiga tahap uji coba. Dengan demikian, berdasarkan tiga tahap uji coba, instrumen tingkat pencapaian perkembangan anak dan *life skills* telah memiliki validitas konstruk, reliabilitas, dan kecocokan model yang baik. Hasil analisis ketiga tahap disajikan pada Tabel 5.

Tabel 5. Ringkasan Hasil Analisis Model Pengukuran Instrumen

Instrumen	Jml. Btr	Uji ke-	Chi-Square		RMSEA	GFI	$\lambda < 0,5$	CR
			Nilai	p-value				
Perk. Moral-Keagamaan	25	1	308,30	0,07	0,029	0,87	2	0,89
		2	311,77	0,058	0,023	0,91	-	0,91
		3	307,31	0,075	0,019	0,94	-	0,91
Perk. Sosial-Emosional	26	1	330,54	0,081	0,027	0,86	-	0,92
		2	333,69	0,07	0,022	0,91	-	0,91
		3	331,38	0,089	0,018	0,93	-	0,92
Perk. Bahasa	24	1	282,39	0,060	0,030	0,87	-	0,70
		2	276,32	0,089	0,022	0,90	-	0,75
		3	286,48	0,051	0,02	0,94	2	0,82
Perk. Kognitif	26	1	331,27	0,066	0,028	0,86	-	0,87
		2	326,05	0,089	0,021	0,91	2	0,80
		3	330,72	0,075	0,018	0,93	1	0,76
Perk. Fisik-Motorik	27	1	356,76	0,077	0,027	0,86	4	0,72
		2	351,72	0,094	0,02	0,91	3	0,85
		3	355,86	0,076	0,018	0,93	-	0,82
Life Skills	30	1	439,09	0,081	0,025	0,84	1	0,72
		2	437,05	0,092	0,019	0,90	-	0,76
		3	447,32	0,055	0,018	0,92	2	0,74

Efektivitas Model Evaluasi Program Layanan PAUD (IPPO)

Untuk mengetahui efektivitas model evaluasi yang dikembangkan, maka model disebarkan kepada 3 orang kepala TK dan 17 orang guru TK untuk dinilai efektivitasnya. Suatu model evaluasi dikatakan baik atau efektif apabila memenuhi syarat: (a) komprehensif, yaitu mencakup banyak komponen atau subkomponen program, baik input, proses, produk maupun *outcome*, (b) praktis, yaitu bersifat sederhana dan mudah dalam penggunaannya, (c) ekonomis, yaitu tidak membutuhkan biaya yang besar, tenaga yang banyak, maupun waktu yang lama, serta (d) didukung oleh instrumen

pengumpul data yang valid dan reliabel. Selain itu, model evaluasi harus memiliki kejelasan bahasa.

Untuk itu, penilaian difokuskan pada aspek-aspek: (a) kejelasan, (b) kekomprehensifan, (c) kepraktisan, dan (d) keekonomisan. Untuk mengetahui validitas dan reliabilitas dilakukan uji coba lapangan.

Penilaian menggunakan skala 5, yaitu: 5 = sangat baik, 4 = baik, 3 = cukup, 2 = kurang baik, dan 1 = tidak baik. Berdasarkan penilaian, dihitung rerata skor total. Rerata skor hasil penilaian terhadap model evaluasi layanan pendidikan anak usia dini disajikan dalam Tabel 6.

Tabel 6. Hasil Penilaian terhadap Efektivitas Model IPPO

No.	Aspek yang Dinilai	Rerata Skor	Rerata per Aspek
1.	Kejelasan prosedur/langkah-langkah evaluasi	4,2	4,2
2.	Kejelasan/kekomunikatifan bahasa	4,2	
3.	Kekomprehensifan komponen model	4,4	4,31
4.	Kekomprehensifan indikator instrumen	4,3	
5.	Kepraktisan panduan	4,5	4,4
6.	Kepraktisan instrumen	4,3	
7.	Keekonomisan dari segi waktu	4,5	4,3
8.	Keekonomisan dari segi biaya	4,2	
9.	Keekonomisan dari segi tenaga	4,2	
Jumlah rerata		38,8	17,21
Rerata total		4,3	4,3

Berdasarkan data pada tabel 6 di atas, dapat dikatakan secara umum, model yang dikembangkan sangat efektif. Jika dilihat skor pada masing-masing aspek, diperoleh rerata skor sebagai berikut: (1) kejelasan 4,2, (2) kekomprehensifan 0,35, (3) kepraktisan 4,4, dan (4) keekonomisan 4,3. Berdasarkan rerata skor tersebut, dapat dikatakan bahwa: (1) dari aspek kejelasan, model jelas langkah-langkah dan bahasanya; (2) dari segi kekomprehensifan, baik komponen model maupun indikator, model sangat komprehensif, (3) dari segi kepraktisan panduan

maupun instrumennya, model sangat praktis; dan (4) dari segi keekonomisan biaya, tenaga, dan waktu, model sangat ekonomis.

Untuk memperoleh informasi yang akurat, diperlukan instrumen pengumpul data yang valid dan reliabel. Berdasarkan uji coba lapangan, ternyata instrumen yang dikembangkan dalam penelitian ini telah memiliki validitas, reliabilitas, maupun kecocokan model yang baik sebagaimana diuraikan di atas. Oleh karena semua persyaratan sebagai model yang efektif sebagaimana dikemukakan di atas telah terpenuhi,

maka dapat dikatakan bahwa model evaluasi program layanan pendidikan anak usia dini (model IPPO) sangat efektif untuk diterapkan.

## Simpulan dan Saran

### Simpulan

Berdasarkan seluruh uraian yang disajikan pada bab IV, maka dapat ditarik simpulan sebagai berikut: 1) model evaluasi program layanan pendidikan anak usia dini yang dikembangkan dalam penelitian ini adalah model yang terdiri atas empat komponen, yaitu: evaluasi input, proses, produk, dan *outcome*. Model evaluasi ini disebut Model IPPO. Model evaluasi ini dilengkapi dengan prosedur evaluasi, panduan penggunaan, dan instrumen-instrumennya: (a) prosedur evaluasi program layanan pendidikan anak usia dini (Model IPPO) mengikuti langkah-langkah: (1) evaluasi input, (2) evaluasi proses, (3) evaluasi produk, dan (4) evaluasi *outcome*; (b) panduan penggunaan evaluasi program layanan PAUD (model IPPO) memuat: ketentuan umum, langkah-langkah evaluasi, pedoman penyekoran, pedoman pengolahan skor, waktu evaluasi, rekomendasi, dan format laporan hasil evaluasi; (c) menurut para pakar, praktisi, dan pengguna model, prosedur, instrumen, dan panduan yang dikembangkan sudah baik dan dapat digunakan tanpa perbaikan; d) Instrumen yang dikembangkan memiliki validitas, reliabilitas, dan memiliki kecocokan model yang baik. Semua indikator memiliki nilai  $t \geq 1,96$  dan nilai  $\lambda (\lambda) \geq 0,5$ ; koefisien reliabilitas di atas 0,7; *Chi-Square* dengan  $p\text{-value} \geq 0,05$ ,  $RMSEA \leq 0,08$ , dan  $GFI \geq 0,9$ ; 2) menurut penilaian praktisi (guru TK) dan pengguna model (kepala TK), model evaluasi program yang dikembangkan (Model IPPO) sangat efektif karena komponennya komprehensif, terdiri atas evaluasi input, proses, produk, dan *outcome*, praktis (sederhana dan mudah digunakan), ekonomis (tidak banyak memerlukan tenaga, biaya, dan waktu), dan didukung oleh instrumen yang valid dan reliabel.

### Saran

Berdasarkan simpulan tersebut, dapat dirumuskan rekomendasi sebagai berikut: (1) evaluasi model IPPO dapat dijadikan model oleh kepala TK maupun Dinas Pendidikan untuk mengevaluasi implementasi program layanan PAUD, selain akreditasi yang telah ada; (2) untuk menguji efektivitas model secara empiris, model perlu dimplementasikan sebelum digunakan secara luas; (3) dalam melakukan penilaian terhadap proses layanan di kelas atau di halaman bermain, kepala TK hendaknya melakukan pengamatan beberapa kali (tidak hanya sekali saja); (4) agar diperoleh hasil evaluasi yang objektif, guru, kepala TK, dan orang tua diharapkan mengisi instrumen dengan cermat dan sejujurnya, sesuai dengan keadaan yang sebenarnya; (5) Kepala PAUD/TK dalam melakukan penilaian terhadap kompetensi sosial dan kepribadian guru perlu melakukan *cross check* dengan teman sejawat atau guru yang lain untuk memperoleh data yang lebih objektif; 6) dalam menganalisis data/skor hasil evaluasi, evaluator sebaiknya menggunakan program *Excel for Windows* untuk lebih memudahkan pekerjaan; 7) dalam melakukan analisis data hasil evaluasi, sebaiknya setiap lembaga PAUD memiliki Team Analisis Data, yang terdiri atas kepala dan wakil kepala PAUD dan guru.

### Daftar Pustaka

- Anwar & Ahmad, Arsyad. (2007). *Pendidikan anak usia dini (pendidikan praktis bagi ibu dan calon ibu)*. Bandung: Alfabeta.
- Aqib, Zaenal. (2011). *Pedoman teknis penyelenggaraan PAUD (Pendidikan anak usia dini)*. Bandung: Nuansa Aulia.
- Ashiabi, G. S. (2007). Play in the peschool classroom: its socioemosional significance and the teacher's role in play [Versi elektronik]. *Early Childhood Education Journal*, 35 (2), 199-207.
- Azwar, Saifuddin. (2012). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar Offset.

- Beard, K. Y. & Sugai, G. (2004). First step to success: An early intervention for elementary children at risk for anti-social behavior [Versi elektronik]. *Behavioral Disorders*, 29 (4), 396-409.
- Berk, L. E. (2007). *Development throught the lifespan (4<sup>th</sup> ed)*. Boston: Pearson Education, Inc.
- Borg, W. R. & Gall, M. D. (1983). *Educational research: An introduction (4<sup>th</sup> ed)*. New York & London : Longman.
- Depdiknas. (2009). *Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 58, Tahun 2009, tentang Standar Pendidikan Anak Usia Dini*.
- Gozali, Imam. & Fuad. (2008). *Struktural eqition modeling: Teori, konsep, dan aplikasi dengan program Lisrel*. Semarang: Badan Penerbit Universitas Diponegoro.
- Hadis, Abdul & Nurhayati B. (2010). *Manajemen mutu pendidikan*. Bandung: Alfabeta.
- Hayat, Bahrul, & Yusuf, Suhendra. (2010). *Benchmark internasional mutu pendidikan*. Jakarta: Bumi Aksara.
- Hermawati, Istiana. (2007). *Evaluasi program pendidikan anak usia dini (PAUD) bagi anak dari keluarga miskin di tempat penitipan anak (TPA) Beringharjo Yogyakarta*. Departemen Sosial RI: Balai Besar Penelitian dan Pengembangan Pelayanan Keajahteraan Sosial (B2-P3KS) Yogyakarta.
- Hiryanto, dkk. (2004). *Pemetaan tingkat pencapaian mutu program pendidikan anak usia dini (PAUD) di Provinsi DIY*. (Laporan penelitian, tidak diterbitkan). Yogyakarta: Lembaga penelitian UNY.
- Janawi. (2011). *Kompetensi guru: Citra guru professional*. Bandung: Alfabeta.
- Kaufman, R. & Thomas, S. (1980). *Evaluation without fear*. New York: New Viewpoints.
- Kemendiknas. (2011). *Petunjuk Teknis Penyaluran Bantuan Alat Permainan Edukatif*. Jakarta: Direktorat Pembinaan Pendidikan Anak Usia Dini.
- Kunandar. (2007). *Guru professional: Implementasi kirikulum tingkat satuan pendidikan (KTSP) dan sukses dalam sertifikasi guru*. Jakarta: Rajagrafindo Persada.
- Latan, Hengky. (2012). *Struktural equation modeling: Konsep dan aplikasi menggunakan program Lisrel 8.80*. Bandung: Alfabeta.
- Mann, E. A. & Reinolds, A. (2006). Early intervention and delinquency prevention: Evidence from the Chicago longitudinal study [Versi elektronik]. *Social Work Research*, 30 (3), 153-167.
- Mardapi, Djemari. (2012). *Pengukuran, penilaian, dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- Nugraha, Ali. (2010). *Evaluasi pembelajaran untuk anak usia dini*. Jakarta: Universitas Pendidikan Indonesia.
- Samuelsson, I. P. (2011). Why we should begin early with ESD: The role of early childhood education [Versi elektronik]. *IJEC*, 43, 103-118.
- Sudijono, Anas. (2011). *Pengantar evaluasi pendidikan*. Jakarta: PT Raja Grafindo Persada.
- Sujiono, Yuliani Nurani. (2009). *Konsep dasar pendidikan anak usia dini*. Jakarta: PT Indeks.
- Widoyoko, S.E.P. (2008). *Pengembangan model evaluasi program pembelajaran ilmu pengetahuan sosial di SMP*. Disertasi doctor, tidak dipublikasikan. Universitas Negeri Yogyakarta.
- Woolfolk, A. (2007). *Educational psychology (10<sup>th</sup> ed)*. New York: Pearson Education, Inc.

## **PENGEMBANGAN TES DIAGNOSTIK KESULITAN BELAJAR MATEMATIKA DI SD**

<sup>1)</sup>M. Duskri, <sup>2)</sup>Kumaidi, <sup>3)</sup>Suryanto

<sup>1)</sup>UIN Ar-Raniry Banda Aceh, <sup>2)</sup>Universitas Muhammadiyah Surakarta,

<sup>3)</sup>Universitas Negeri Yogyakarta

<sup>1)</sup>duskri\_iainaceh@yahoo.co.id, <sup>2)</sup>kuma426@yahoo.com.

### **Abstrak**

Tujuan penelitian untuk: (1) menemukan prosedur pengembangan tes diagnostik kesulitan belajar matematika SD/MI, (2) mengetahui kualitas butir tes diagnostik yang dikembangkan, dan (3) mengetahui informasi yang dapat dimunculkan dari hasil analisis tes diagnostik. Penelitian ini merupakan penelitian pengembangan dari Borg & Gall. Subjek uji sebanyak 542 orang siswa kelas VI SD/MI di Kota Banda Aceh dan Aceh Besar. Analisis kuantitatif untuk mengetahui informasi butir tes menggunakan program ITEMAN dan Program R. Hasil penelitian menunjukkan: (1) pengembangan tes diagnostik kesulitan belajar matematika di SD ini meliputi: studi pendahuluan, studi literatur dan hasil-hasil penelitian, analisis masalah, merumuskan *learning continuum*, merumuskan peta konsep, menyusun tes *essay*, polarisasi jawaban siswa, menyusun tes bentuk pilihan ganda, validasi pakar melalui *focus group discussion*, uji coba terbatas, dan uji yang diperluas, (2) indeks daya beda butir tes antara 0,391 sampai dengan 2,317, indeks kesukaran butir tes antara -2,158 sampai dengan 2,528, kecocokan uji tes dengan kemampuan peserta ( $\theta$ ) antara -2,00 sampai dengan 2,60, dan fungsi informasi tes antara 0,111 sampai dengan 3,879, dan (3) informasi yang dapat dimunculkan dari tes meliputi: hasil tes secara klasikal dan individual, grafik ketuntasan belajar, profil individual, analisis salah konsepsi dan saran remedial.

**Kata kunci:** *tes diagnostik, kesulitan belajar matematika*

## **DEVELOPING DIAGNOSTIC TEST OF MATHEMATICS LEARNING DIFFICULTIES IN ELEMENTARY SCHOOLS**

<sup>1)</sup>M. Duskri, <sup>2)</sup>Kumaidi, <sup>3)</sup>Suryanto

<sup>1)</sup>UIN Ar-Raniry Banda Aceh, <sup>2)</sup>Universitas Muhammadiyah Surakarta,

<sup>3)</sup>Universitas Negeri Yogyakarta

<sup>1)</sup>duskri\_iainaceh@yahoo.co.id, <sup>2)</sup>kuma426@yahoo.com.

### **Abstract**

The purposes of this study were: (1) to find a method to develop diagnostic test of mathematics learning difficulties in primary schools, (2) to determine the diagnostic test characteristics developed, and (3) to determine the information that could be generated by using diagnostic test. This study was a development research by Borg & Gall. The participants of main field testing were 542 grade VI students of primary schools in Banda Aceh City and Aceh Besar Regency. The quantitative analysis to determine the information of the test items developed used ITEMAN program and R Program. The results of this study are: (1) in developing the diagnostic test of the mathematics learning difficulties, the research phase included preliminary study, literature and research results study, problem analysis, formulating learning continuum, formulating concept map, preparing essay-form tests, students' answers polarization, preparing multiple-choice tests, experts' validation with focus group discussion, preliminary and readability tests, operational field testing, and main field testing to obtain model that fitted theoretically and empirically, (2) discriminating power of test items index ranges from 0.391 to 2.317, the difficulty test items index ranges from -2.158 to 2.528, the test compatibility testing with the participants ability ( $\theta$ ) ranges from -2.00 to 2.60 and the test information function ranges from 0.111 to 3.879, and (3) the information that can be generated from the diagnostic test includes classically and individually test results, graphic of learning outcomes attainment, individual profile, every answers analysis, false conception and remedial suggestion analysis.

**Keywords:** *diagnostic test, mathematics learning difficulties*

## Pendahuluan

Salah satu ciri dari mata pelajaran matematika adalah bertumpu pada pengetahuan atau pengalaman sebelumnya, deduktif dan abstrak. Belajar matematika pada hakikatnya belajar tentang konsep, struktur konsep dan mencari hubungan antarkonsep dan strukturnya (Subarinah, 2006, p.1). Oleh sebab itu, jika terjadi miskonsepsi pada konsep sebelumnya, maka dapat dipastikan akan berakibat miskonsepsi pula pada konsep berikutnya. Apabila miskonsepsi ini tidak dibenahi, maka peserta didik akan mengalami kesulitan dalam belajar matematika. Bagi peserta didik, kesulitan dalam mempelajari matematika akan berakibat keputusasaan, tidak berminat dan pada akhirnya akan membenci pelajaran matematika.

Berdasarkan hasil survei dari *Programme for International Student Assessment (PISA)* bahwa kemampuan matematika anak-anak Indonesia dalam usia kisaran 15 tahun di dunia internasional berada pada peringkat yang belum memuaskan. Sebagaimana dirilis dalam laman Litbang Kemendikbud, bahwa Indonesia selalu berada di bawah skor rata-rata internasional yaitu 500. Pada tahun 2000, Indonesia menduduki peringkat 39 dari 41 negara dengan skor rata-rata 367, tahun 2003 menduduki peringkat 38 dari 40 negara dengan skor rata-rata 361, tahun 2006 menduduki peringkat 50 dari 57 negara dengan skor rata-rata 391, dan tahun 2009 menduduki peringkat 61 dari 65 negara dengan skor rata-rata 371 (Litbang Kemendikbud, 2011). Hasil PISA matematika tahun 2009, ternyata 43,5% siswa Indonesia tidak mampu menyelesaikan soal PISA paling sederhana, 33,1% hanya mampu mengerjakan soal jika pertanyaan dari soal kontekstual diberikan secara eksplisit serta semua data yang dibutuhkan untuk mengerjakan soal diberikan secara tepat, dan hanya 0,1% siswa Indonesia yang mampu mengembangkan dan mengerjakan pemodelan matematika yang menuntut keterampilan berpikir dan penalaran (Wijaya, 2012, p.2). Hasil-hasil tersebut menunjukkan bahwa kemampuan pemecahan masalah bagi sebagian besar siswa Indonesia masih rendah, hal

ini perlu menjadi perhatian guru-guru dalam melakukan pembelajaran matematika tidak hanya berlatih mengerjakan soal yang secara eksplisit telah tersedia semua argumen yang dibutuhkan, akan tetapi lebih dari itu, yaitu membelajarkan siswa agar mampu berfikir kritis. Dengan demikian, siswa akan mampu menyelesaikan persoalan-persoalan matematika yang menuntut pemikiran dan penalaran.

*Trends in International Mathematics and Science Study (TIMSS)* pada tahun 1999, 2003 dan 2007 melakukan studi internasional tentang prestasi matematika dan sains siswa sekolah lanjutan tingkat pertama. Bidang matematika yang diuji dalam studi tersebut adalah kemampuan dalam materi bilangan, aljabar, geometri, data dan peluang. Berdasarkan hasil survei tersebut, ternyata kemampuan rata-rata siswa-siswi Indonesia masih di bawah rata-rata internasional. Khusus untuk bidang matematika, pada tahun 1999 skor rata-rata internasional 487, sedangkan Indonesia mendapat skor rata-rata 403 dan menduduki peringkat 34 dari 38 negara yang disurvei. Pada tahun 2003 skor rata-rata internasional 467, sedangkan Indonesia mendapat skor rata-rata 411 dan menduduki peringkat 35 dari 46 negara yang disurvei. Pada tahun 2007 skor rata-rata internasional 500, sedangkan Indonesia mendapat skor rata-rata 397 dan menduduki peringkat 36 dari 49 negara yang disurvei (Litbang Kemdikbud, 2011).

Sejalan dengan pernyataan di atas, berdasarkan hasil laporan dari penelitian yang dilakukan oleh Tim Pusat Pengembangan Penataran Guru Matematika yang meneliti pada beberapa Sekolah Dasar di Indonesia mengungkapkan bahwa 51% siswa kesulitan dalam aspek berhitung, 50% siswa kesulitan dalam penguasaan konsep, dan 49% siswa mengalami kesulitan dalam menyelesaikan soal cerita (Tim PPPG Matematika, 2001, p.18). Selanjutnya, pada tahun 2002 berdasarkan hasil penelitian dari Tim PPPG Matematika mengungkapkan bahwa di beberapa wilayah Indonesia, sebagian besar siswa SD kesulitan dalam menyelesaikan soal-soal cerita dan menerjemahkan soal cerita ke da-

lam model matematika (Tim PPPG Matematika, 2002, p.71).

Rendahnya mutu pendidikan, termasuk hasil pembelajaran yang dicapai sebagaimana yang telah disebutkan di atas (termasuk pembelajaran matematika) dapat disebabkan oleh berbagai faktor, antara lain karakteristik mata pelajaran, siswa dan guru. Matematika dibangun atas konsep-konsep yang abstrak dan deduktif, akibatnya sukar dipahami oleh sebagian besar siswa.

Soedjadi (1996, p.2) menyatakan bahwa siswa SLTP lemah dalam penguasaan geometri, pecahan, dan soal cerita. Lemahnya penguasaan siswa pada jenjang SMP ini boleh jadi disebabkan oleh lemahnya penguasaan siswa terhadap matematika ketika berada pada jenjang SD/MI. Penyebab lain karena matematika menggunakan simbol-simbol, berfikir deduktif, dan abstrak. Bagi siswa yang masih mengalami kendala pada materi tertentu, perlu diberikan bimbingan tersendiri, misalnya dalam bentuk *remedial teaching*. Menurut Ischak & Warji (1992, p.35) rendahnya hasil belajar matematika dimungkinkan penyebabnya karena tidak efektif (maksimal) pemberian bantuan belajar berupa *remedial teaching* kepada siswa yang mengalami kesulitan.

Proses pembelajaran dapat berjalan efektif dan sukses apabila perbedaan-perbedaan individu (*individual differences*) yang ada di antara peserta didik mendapat perhatian (Suwarsono, 1987, p.69) dan dapat dideteksi. Perbedaan-perbedaan tersebut akan berpengaruh terhadap tingkat pemahaman siswa dalam mempelajari materi tertentu. Diketahuinya perbedaan-perbedaan individual (terutama kesulitan dalam memahami materi pelajaran), faktor yang ikut menjadi penyebab kesulitan, dan faktor utama penyebab kesulitan tersebut, maka kegiatan pembelajaran yang akan dilakukan oleh guru dapat disesuaikan. Untuk mengetahui perbedaan tingkat pemahaman peserta didik (terutama kesulitan dalam memahami materi pelajaran), faktor yang ikut menjadi penyebab kesulitan, dan faktor utama penyebab kesulitan tersebut dapat dilakukan dengan tes diagnostik kesulitan belajar.

Diagnosis dalam pendidikan merupakan konsep yang luas, meliputi identifikasi kekuatan dan kelemahan peserta didik (Sriati, 1993; Thorndike dan Hagen, 1977, p.72). Sheehan (1997, p.333) menyatakan bahwa hasil tes bukan hanya menunjukkan berapa skor yang benar dan salah dari setiap siswa, namun memberikan informasi pola pencapaian ketuntasan kemampuan siswa. Informasi tersebut dapat membantu siswa dan guru untuk memahami lebih baik makna dari skor itu sendiri dalam kaitannya dengan peningkatan hasil belajar. Prosedur untuk menterjemahkan hasil tes dalam kaitannya dengan proses pembelajaran dapat dirancang melalui *student-level diagnostic* atau *group-level diagnostic*. Diagnosis level siswa secara individual menggambarkan kekuatan dan kelemahan tiap individual siswa masing-masing. Sedangkan diagnosis secara kelompok menggambarkan kekuatan dan kelemahan siswa yang didasarkan pada titik-titik tertentu berdasarkan laporan skala skor dari tes itu.

Waskito & Kumaidi (2010) mengembangkan model analisis prestasi belajar untuk mengidentifikasi kesulitan belajar peserta didik dengan menyusun butir tes objektif pilihan ganda dan esai serta mengembangkan perangkat lunak (*software*) komputer yang dapat mengolah data dan laporan hasil tes diperoleh lebih cepat. Hasil analisisnya berupa: (1) profil diri individual peserta didik yang menggambarkan tentang penguasaan peserta didik terhadap kompetensi dasar yang telah dipelajari serta saran pembelajaran yang harus dilakukan untuk mengatasi kesulitan tersebut; (2) profil kelas yang memberikan informasi tentang siapa saja di kelas tersebut yang belum menguasai indikator soal; dan (3) grafik yang menggambarkan penguasaan setiap siswa terhadap materi pelajaran yang dites.

Penelitian yang dilakukan oleh Fauzan (2010) mendeskripsikan kesulitan-kesulitan yang dialami peserta didik terhadap materi ajar sesuai dengan indikator yang telah dirumuskan. Penelitian yang dilakukan oleh Waskito & Kumaidi (2010) memberikan informasi tentang skor dan nilai tiap peserta

didik dan memberikan saran apa saja yang harus dipelajari untuk mengatasi kesulitan tiap butir tes, serta memberikan informasi prestasi setiap siswa dibandingkan dengan teman sekelasnya.

Berdasarkan hal tersebut, peneliti menggabungkan kedua hasil penelitian dari Fauzan (2010) dan Waskito & Kumaidi (2010), dan melengkapinya dengan beberapa informasi penting lainnya. Hasil penelitian ini diharapkan mampu memberikan informasi tentang grafik kedudukan siswa di kelasnya dibandingkan dengan nilai Kriteria Ketuntasan Minimal (KKM) yang ditetapkan serta skor dan nilai yang diperolehnya, profil individual, deskripsi tentang konsep-konsep apa saja yang sudah dipahami, konsep-konsep apa saja yang belum dipahami, disertai dengan informasi dugaan kuat penyebab peserta didik mengalami kesulitan khususnya penguasaan materi prasyarat pada setiap konsep tersebut, dan memberikan saran materi pembelajaran apa saja yang harus diremedialkan terlebih dahulu untuk mengatasi kesulitan tersebut. Secara teoretis ini mungkin dilakukan, karena berdasarkan *learning continuum* atau melalui tela'ah materi prasyarat akan dapat diprediksikan secara tepat penyebab kesulitan yang dialami siswa dan alternatif pemecahannya. Hal ini sangat membantu guru dalam merencanakan dan melaksanakan *remedial teaching* terhadap materi yang belum dikuasai oleh siswa.

Permasalahan dibatasi pada pengembangan tes diagnostik kesulitan belajar matematika di SD. Tes bentuk pilihan ganda, tes uraian, dan wawancara telah banyak digunakan untuk mengungkap kesulitan belajar yang dialami siswa terhadap materi pelajaran matematika. Khusus untuk tes pilihan ganda, *option* tiap butirnya masih jarang didasarkan pada polarisasi jawaban siswa terhadap butir tes tersebut dalam bentuk *essay*. Apabila hal itu dilakukan, akan dapat difungsikan sebagai alat untuk mendeteksi penyebab siswa tidak mampu menjawab butir tes tersebut dengan benar.

Tes diagnostik yang dikembangkan ini khusus untuk tes diagnostik kesulitan belajar matematika jenjang SD/MI pada kelas

VI. Dipilih kelas VI dengan asumsi bahwa mereka telah mempelajari pelajaran matematika selama beberapa tahun di bangku SD/MI, dengan harapan cakupan materi tes yang dikembangkan dapat menjangkau mulai dari kelas-kelas sebelumnya sampai dengan kelas VI SD/MI terutama untuk memprediksi kemungkinan kesalahan-kesalahan konsep yang mungkin telah dialami oleh peserta didik selama beberapa tahun.

Tujuan dari penelitian ini adalah untuk: (1) menemukan prosedur pengembangan tes diagnostik kesulitan belajar matematika SD/MI, (2) mengetahui kualitas butir tes diagnostik kesulitan belajar matematika SD/MI yang dikembangkan, dan (3) mengetahui informasi yang dapat dimunculkan dari hasil analisis tes diagnostik kesulitan belajar matematika SD/MI yang dikembangkan.

### Metode Penelitian

Penelitian ini merupakan penelitian pengembangan (*research and development*), yakni pengembangan tes diagnostik kesulitan belajar matematika di SD. Produk yang akan dihasilkan dari penelitian ini adalah instrumen tes diagnostik kesulitan belajar matematika SD dan desain perangkat lunak program komputer untuk analisis hasil diagnostiknya.

Produk dari pengembangan tes diagnostik kesulitan belajar matematika di SD ini berupa instrumen tes diagnostik kesulitan belajar matematika SD/MI berbentuk pilihan ganda khusus materi bilangan. Setiap butir tes yang dikembangkan, *option-option* yang dihasilkan didasarkan pada polarisasi jawaban peserta yang diperoleh dari uji tes awal berbentuk *essay*. Kemungkinan *option* ini didapat dari hasil uji desain tes *essay* (dalam studi pendahuluan untuk kebutuhan *need assessment*) yang dilakukan di lapangan. Berdasarkan studi pendahuluan ini, didesain butir tes-butir tes diagnostik kesulitan belajar matematika SD/MI berbentuk pilihan ganda, dilanjutkan dengan *focus group discussion* (FGD), diujicobakan secara terbatas, dilakukan analisis kuantitatif dengan program iteman dan analisis kualitatif melalui FGD. Selanjutnya, dikembangkan lagi menjadi bu-

tir tes-butir tes yang akan digunakan untuk uji tes yang diperluas, dilakukan analisis kuantitatif dengan program *Iteman*, program R dengan 1 parameter dan 2 parameter, analisis kualitatif melalui FGD sampai kepada tahap revisi produk akhir.

Kriteria-kriteria peserta yang dilibatkan dalam FGD di antaranya paling tidak: (1) menguasai materi matematika SD/MI, (2) minimal telah mengajar mata pelajaran/bidang studi matematika selama 5 tahun, (3) guru bidang studi matematika yang terlibat sebagai instruktur atau narasumber pada kegiatan KKG/PKG, dan (4) dosen pada Prodi Pendidikan Matematika/PGSD/PGMI minimal berpendidikan S-2 Pendidikan Dasar/Pendidikan Matematika dan telah berpengalaman mengajar minimal 3 tahun di PGSD/PGMI.

Kualitas dari produk instrumen tes diagnostik kesulitan belajar matematika SD/MI diteliti berdasarkan: (1) penilaian para ahli melalui *expert judgment* dengan menggunakan teknik *focus group discussion* (FGD); (2) analisis menggunakan teori tes klasik (program ITEMAN), *item response theory* (IRT) dengan satu dan dua parameter (program R); dan (3) uji lapangan. Sedangkan untuk menguji kualitas produk berupa desain program komputer untuk analisis hasil diagnostiknya diteliti berdasarkan: (1) uji coba produk dengan melibatkan sejumlah guru dan fasilitator dalam pembelajaran; dan (2) verifikasi hasil uji coba.

Subjek coba penelitian ini sebanyak 542 siswa kelas VI SD/MI di Kota Banda Aceh dan Aceh Besar Provinsi Aceh. Diambil siswa kelas VI dengan pertimbangan bahwa mereka telah mempelajari sejumlah materi mulai dari kelas I sampai dengan kelas VI.

Data empirik merupakan data primer yaitu data yang diperoleh dari sumber primer yaitu respons siswa terhadap butir tes, para guru, dan *expert* dalam bidang pendidikan matematika, dan pakar evaluasi pendidikan yang tergabung dalam *Focus Group Discussion*. Data yang diperoleh dianalisis, dan digunakan untuk pengambilan keputusan dalam penelitian pengembangan ini.

Data kualitatif berupa masukan dari para ahli yang menjadi peserta *Focus Group Discussion*, yaitu ahli pendidikan matematika (dosen pendidikan matematika, guru yang mengajar matematika), dan ahli pengukuran dan evaluasi (dosen pengukuran atau evaluasi pendidikan), serta *reviewer* dari dosen Prodi PGMI/Pendidikan Matematika, mahasiswa Program Pascasarjana Prodi Pendidikan Matematika, dan guru yang mengajar matematika.

Masukannya berupa kebenaran konsep matematika yang diujikan dalam instrumen (butir tes), kesesuaian indikator kemampuan dengan Kompetensi Dasar, kesesuaian butir soal-butir soal dengan materi yang terkandung dalam kurikulum yang didasarkan pada Kompetensi Dasar, kesesuaian peta konsep yang dibuat untuk mendapatkan informasi materi pra-syarat yang sesuai dengan materi pokok, konstruksi tes, bahasa yang digunakan (menggunakan bahasa Indonesia yang benar dan dapat dimengerti oleh usia anak sekolah dasar), ketepatan instrumen untuk mendiagnosis kesulitan belajar matematika siswa sekolah dasar khususnya berkenaan dengan materi bilangan, kesesuaian rumusan pilihan alternatif jawaban tiap butir dengan dugaan terjadinya penyebab salah konsepsi (penyebab siswa mengalami kesulitan dalam menjawab butir soal), dan kesesuaian saran remedial pembelajaran yang diajukan untuk mengatasi salah konsepsi. Kejelasan petunjuk soal, kecukupan alokasi waktu yang tersedia untuk menyelesaikan soal, dan kelengkapan perangkat yang harus disediakan untuk mendiagnosis kesulitan belajar matematika juga perlu mendapatkan masukan.

Data kualitatif lainnya berupa penilaian tim validator terhadap perangkat tes diagnostik secara keseluruhan baik sebelum uji coba maupun setelah direvisi yang didasarkan pada uji coba lapangan. Data ini diperlukan untuk mendapatkan informasi tentang reliabilitas antar validator terhadap validasi instrumen tes yang mereka lakukan.

Data kuantitatif berupa skor peserta didik setelah merespons tes diagnostik *multiple choice* dalam uji coba terbatas dan uji

yang diperluas. Data tersebut digunakan untuk validasi empiris terhadap kualitas butir tes berupa tingkat kesulitan soal, daya beda soal, keberfungsian pengecoh tiap butir tes, dan reliabilitas instrumen tes secara keseluruhan. Khusus untuk data empiris setelah uji yang diperluas, di samping informasi tersebut juga diperlukan informasi berupa estimasi kemampuan peserta, estimasi daya beda soal tiap butir, estimasi tingkat kesulitan tiap butir tes, dan fungsi informasi tes.

Instrumen pengumpulan data terdiri atas: (1) tes diagnostik kesulitan belajar matematika SD berbentuk *essay* bertujuan *need assessment* untuk mendapatkan kondisi riil yang dialami oleh siswa, dari sini akan didapatkan polarisasi jawaban siswa, digunakan untuk merancang *option* tiap butir tes, (2) lembar validasi soal, (3) tes diagnostik kesulitan belajar matematika SD berbentuk pilihan ganda yang bertujuan untuk mengetahui kecocokan model, dan (4) rubrik penilaian secara kualitatif yang terdiri atas soal, kunci jawaban, analisis pengecoh (*distractor*), penyebab salah konsepsi, dan saran remedi pembelajaran. Data masukan dari para ahli melalui *expert judgment* terhadap produk awal (produk ke I) tes diagnostik kesulitan belajar matematika di SD (TDKB-MatSD) yang dikembangkan dianalisis secara deskriptif dengan memverifikasi masukan tersebut, dijadikan sebagai bahan untuk merevisi produk awal TDKB-MatSD.

Produk awal (produk ke I berupa tes *essay*) dilakukan uji pendahuluan untuk mendapatkan polarisasi jawaban peserta yang berguna untuk merancang *option* tiap butir tes. Berdasarkan hasil analisis uji pendahuluan dikembangkan lagi menjadi produk ke II berupa butir tes-butir tes pilihan ganda. Produk ke II ini dinilai oleh beberapa orang dosen Prodi PGMI/Pendidikan Matematika, dan didiskusikan lagi dengan tim reviewer untuk mendapatkan masukan perbaikan. Diskusi ini difokuskan pada kesesuaian rumusan indikator kemampuan dengan Kompetensi Dasar, butir tes yang dirumuskan, konstruk tes yang telah dirumuskan, *option* tiap butir tes, kebahasaan,

analisis pengecoh (*distractor*). Masukan dari reviewer dan hasil penilaian oleh beberapa orang dosen tersebut digunakan untuk memperbaiki produk ke II (TDKB-MatSD).

Setelah diperoleh butir tes-butir tes yang telah sesuai dengan kriteria (valid dan reliabel secara empiris) yang dianalisis melalui teori tes klasik (ITEMAN) yang menghasilkan produk III TDKB-MatSD, selanjutnya dilakukan uji yang diperluas untuk mendapatkan informasi keterlaksanaan di lapangan. Subjek penelitian untuk uji yang diperluas sebanyak 542 orang siswa kelas VI yang tersebar di 10 SD/MI Kota Banda Aceh dan Aceh Besar.

Data kuantitatif hasil uji yang diperluas dianalisis dengan teori respons butir model logistik 1 parameter dan 2 parameter melalui program R. Sebelum menggunakan analisis teori respons butir, terlebih dahulu dilakukan pemeriksaan asumsi unidimensional dan independen lokal (Hambleton, et al., 1991). Asumsi pertama, *unidimensional* menyatakan bahwa dalam setiap tes hanya ada satu kemampuan yang diukur oleh perangkat butir-butirnya. Asumsi ini sangat sulit dipenuhi karena begitu banyaknya faktor yang mempengaruhi tes, seperti kognitif, kepribadian, motivasi, kecemasan, dan lain sebagainya. Namun demikian, yang paling penting adalah adanya satu komponen yang paling dominan dalam menentukan kemampuan subjek. Salah satu cara pemeriksaan unidimensi adalah dengan melakukan faktor analisis yang menghasilkan *eigenvalue*. Apabila *eigenvalue* salah satu faktor mempunyai nilai lebih dominan dan beberapa kali dibandingkan dengan *eigenvalue* faktor yang lainnya, maka dapat dikatakan syarat unidimensi sudah terpenuhi (Naga, 1992, p.297). Penelitian ini menggunakan analisis program R untuk menentukan *matrix of tetrachoric correlations* antar-item dan dianggap memenuhi persyaratan apabila *eigenvalue* > 1.

Asumsi kedua, *local independence* (independensi lokal) yaitu respons peserta tes terhadap suatu butir tidak berhubungan dengan butir lainnya dalam tes tersebut. Independensi lokal dimaksudkan sebagai

letak pada suatu titik di kontinum parameter ciri peserta tes yakni  $\theta$  (Naga, 1992, p.170). Hal ini dapat diamati dari fungsi informasi butir tes dan fungsi respons butir tes. Selain dua asumsi di atas, asumsi yang juga harus dipenuhi adalah invariansi parameter. Asumsi ini menyatakan bahwa parameter butir tidak tergantung pada parameter siswa dan sebaliknya parameter siswa tidak tergantung parameter butir. Invariansi pada butir tes dibedakan dalam dua bagian, yaitu: (1) invariansi terhadap estimasi parameter-parameter butir tes, dan (2) invariansi terhadap parameter kemampuan (Hambleton, Swaminathan, & Rogers, 1991, p.19).

Memeriksa invariansi terhadap estimasi kemampuan, butir tes dipilah menjadi kelompok butir tes ganjil dan butir tes genap, dilakukan estimasi dengan program R sehingga diperoleh parameter kemampuan setiap siswa dari kedua kelompok. Kedua kelompok kemampuan dikorelasikan, apabila korelasinya tinggi, maka asumsi invariansi kemampuan terpenuhi (Syarifah, 2007, p.16). Korelasi estimasi parameter kemampuan peserta kelompok *item* ganjil dan *item* genap dilakukan dengan program SPSS versi 17.

Analisis melalui teori respons butir bertujuan untuk mendapatkan estimasi daya beda butir tes, estimasi tingkat kesukaran butir tes, estimasi kemampuan peserta, grafik fungsi informasi butir tes, dan grafik fungsi respons butir tes. Dali S. Naga (1992, pp.223-224) menyatakan bahwa batas nilai parameter kemampuan peserta yang berguna secara praktis terletak antara  $-4 \leq \theta \leq +4$ , parameter daya pembeda butir tes harus bernilai positif ( $a_i > 0$ ), parameter tingkat kesukaran butir tes membentang sepanjang nilai  $\theta$  (idealnya  $-2 \leq b_j \leq +2$ ), dan nilai *guessing* (tebakan)  $0 \leq c_j \leq 1$ , idealnya nilai  $c_j$  terletak disekitar 0,25. Berhubung subjek penelitian ini berjumlah 542 orang maka peneliti hanya menganalisis dengan model logistik 1 parameter dan 2 parameter, sehingga nilai tebakan dianggap 0. Estimasi karakteristik butir, estimasi kemampuan peserta, dan uji *fit* tidaknya butir tes terhadap model, yaitu baik model logistik 1 para-

meter maupun model logistik 2 parameter dilakukan dengan program R. Setelah dilakukan uji *fit* butir tes terhadap model, maka kedua model dibandingkan, yang paling banyak butir tesnya *fit*, model tersebutlah yang paling cocok. Uji kecocokan tiap butir tes dengan model dilakukan dengan membandingkan nilai  $\chi^2$  hitung dan nilai  $\chi^2$  tabel. Apabila  $\chi^2$  hitung < nilai  $\chi^2$  tabel dengan taraf signifikansi 0,05 atau nilai peluang yang dihasilkan (uji *fit* model melalui program R) lebih besar dari 0,05 maka butir tes tersebut cocok dengan model, namun apabila sebaliknya maka butir tersebut tidak cocok dengan model.

Uji kelayakan penggunaan *software* komputer TDKB-MatSD diuji coba langsung oleh 15 orang guru SD/MI di laboratorium komputer Prodi Pendidikan Matematika Fakultas Tarbiyah IAIN Ar-Raniry pada tanggal 2 Mei 2013. Guru diminta membaca hasil *print out* analisis komputer yang terdiri atas: (1) profil siswa secara individual berupa: skor, nilai yang diperoleh, ketuntasan individual, deskripsi indikator (materi) yang telah dikuasai, deskripsi indikator (materi) yang belum dikuasai, deskripsi dugaan penyebab siswa mengalami kesulitan terhadap materi tersebut, dan deskripsi saran remedial yang ditawarkan untuk mengatasi kesulitan tersebut, dan (2) profil secara klasikal berupa: ketuntasan secara klasikal dan individual, grafik ketuntasan secara individual dan klasikal, dan ketuntasan indikator (materi) yang diujikan. Hasil uji coba *software* dan *print out* analisisnya, guru diminta memberikan masukan dan penilaian terhadap produk-produk tersebut melalui instrumen uji validasi *software* komputer yang diberikan kepada tiap guru. Data tersebut berupa tanggapan terhadap tiap butir pernyataan: sangat setuju (4), setuju (3), kurang setuju (2), dan tidak setuju (1). Hasil tanggapan guru tersebut dihitung nilai rata-rata tiap butir pernyataan, dengan kriteria: rata-rata 1-1,50 berarti jelek; rata-rata 1,51-2,50 berarti cukup, rata-rata 2,51-3,50 berarti baik, dan rata-rata 3,51-4 berarti baik sekali.

Uji kelayakan dan efektivitas TDKB-MatSD dalam mendiagnosis kesulitan siswa

terhadap indikator yang diujikan, dilakukan dengan teknik uji beda antara dua rerata (Borg & Gall, 1983, pp.543-545). Rerata skor siswa yang berasal dari sekolah favorit dan rerata skor siswa yang berasal dari sekolah nonfavorit dibandingkan. TDK.B-MatSD dianggap efektif dalam mendiagnosis kesulitan belajar siswa, apabila diperoleh perbedaan yang signifikan antara hasil tes siswa yang berasal dari sekolah yang favorit dan sekolah nonfavorit. Pengujian ini dilakukan dengan program SPSS versi 17.0, apabila nilai signifikansinya  $<0,05$  berarti TDKB-MatSD yang telah dikembangkan dianggap efektif untuk mendiagnosis kesulitan belajar siswa.

### Hasil Penelitian dan Pembahasan

Secara keseluruhan butir tes yang dikembangkan memenuhi asumsi-asumsi yang mendasari analisis dengan teori respon butir. Asumsi pertama, unidimensi terpenuhi, karena uji unidimensi dengan menggunakan program R diperoleh *eigenvalue* data sebesar 2,4025 dan nilai signifikansi hasil perhitungan sebesar 0,0099 lebih kecil dari signifikansi yang dipilih 0,05. Hasil uji unidimensi ini juga menunjukkan bahwa butir yang dikembangkan hanya satu faktor, yang berarti validitas konstruksinya juga terpenuhi.

Asumsi kedua, *local independence* (independensi lokal) yaitu respons peserta tes terhadap suatu butir tidak berhubungan dengan butir lainnya dalam tes tersebut. Independensi lokal dimaksudkan sebagai letak pada suatu titik di kontinum parameter ciri peserta tes yakni  $\theta$  (Hambleton, Swaminathan, & Rogers, 1991, p.10). Hal ini dapat diamati dari fungsi informasi butir tes dan fungsi respons butir tes. Hampir semua butir tes memenuhi fungsi informasi yang ideal dan fungsi respons butir yang mendekati fungsi ojaif normal. Hasil uji Khi-Kuadrat diperoleh nilai Khi-Kuadrat hitung jauh lebih kecil dari Khi-Kuadrat tabel, berarti independensi lokal terpenuhi.

Asumsi invariansi parameter menyatakan bahwa parameter butir tidak tergantung pada parameter siswa dan sebaliknya para-

meter siswa tidak tergantung parameter butir. Invariansi pada butir tes dibedakan dalam dua bagian, yaitu: 1) invariansi terhadap estimasi parameter-parameter butir tes, dan 2) invariansi terhadap parameter kemampuan (Hambleton, Swaminathan, & Rogers, 1999, p.19).

Uji invariansi parameter kemampuan, dilakukan pemilahan butir tes ganjil dan butir tes genap. Kedua kelompok kemampuan dikorelasikan, apabila korelasinya tinggi, maka asumsi invariansi kemampuan terpenuhi (Syarifah, 2007, p.16). Hasil uji invarian parameter menunjukkan bahwa korelasi antara estimasi kemampuan peserta butir ganjil dan kemampuan peserta butir genap sebesar 0,822, sehingga dapat dikatakan bahwa asumsi invarian terpenuhi.

Secara umum, karakteristik butir soal yang telah dikembangkan memiliki tingkat kesukaran (b) dari -2,158 sampai dengan 2,5283; daya beda (a) dari 0,391 sampai dengan 2,317, dan tebakan (c) memiliki nilai minimum sebesar 0 (karena model yang dipilih adalah model logistik 2 parameter) dan nilai informasi butir berkisar antara 0,111 sampai dengan 3,879.

Pemilihan model logistik yang digunakan untuk analisis butir tes terlebih dahulu dilakukan uji *fit* butir tes terhadap model. Penelitian ini dipilih model logistik 2 parameter dengan pertimbangan bahwa setelah diuji *fit* model ternyata model logistik 2 parameter lebih banyak butir tes yang *fit* dibandingkan dengan model logistik lainnya.

Fungsi informasi uji tes ditentukan oleh fungsi informasi tiap-tiap butir tes yang membentuk uji tes tersebut. Berdasarkan respons peserta terhadap uji tes tertentu dapat dihitung informasi maksimal yang dapat diberikan oleh uji tes tersebut dengan memasukkan nilai rentang  $\theta$  tertentu (misalnya dari -2 sampai +2), daya beda tiap butir tes, dan tingkat kesukaran tiap butir tes. Uji tes yang dikembangkan memberikan informasi maksimal (tertinggi) pada kemampuan peserta  $\theta = 0$  dengan nilai informasi sebesar 31,636.

Secara keseluruhan, uji tes memberikan informasi yang tinggi pada rentangan kemampuan peserta antara -1 dan +1, sedangkan di luar kemampuan tersebut akan memberikan informasi yang rendah. Hal ini menunjukkan bahwa uji tes yang dikembangkan secara umum sangat cocok bagi peserta dengan kemampuan sedang. Dengan kata lain uji tes yang dikembangkan dapat digunakan untuk melakukan diagnosis kesulitan belajar.

Uji kelayakan dan efektivitas TDKB-MatSD untuk mendiagnosis kesulitan siswa terhadap indikator yang diujikan, dilakukan dengan teknik uji beda antara dua rerata (Borg & Gall, 1983, pp.543-545). Rerata skor siswa yang berasal dari sekolah berkategori tinggi (dalam hal ini sebagai kriterianya adalah sebagai SD favorit) dan rerata skor siswa yang berasal dari sekolah berkategori sedang (dalam hal ini sebagai kriterianya SD yang kurang favorit) dibandingkan. Pengujian ini dilakukan dengan program SPSS versi 17.0, apabila nilai signifikansinya  $<0,05$  berarti TDKB-MatSD yang telah dikembangkan dianggap efektif dalam mendiagnosis kesulitan belajar siswa.

Sebagai sampel untuk uji kelayakan dan efektivitas dalam penelitian ini, berdasarkan hasil tes uji yang diperluas, diambil 2 (dua) SD di Kota Banda Aceh, 2 (dua) SD di Aceh Besar, dan 2 (dua) MIN, masing-masing pasangan tersebut diperbandingkan dengan syarat sekolah seperti tersebut di atas.

Nilai rata-rata siswa SD Negeri Garot Kabupaten Aceh Besar sama dengan 71,32 dengan standar deviasi 15,67 dengan jumlah sampelnya 76 orang, dan nilai rata-rata siswa SD Negeri Lambheu Aceh Besar sama dengan 55,92, standar deviasi 13,69, dengan jumlah sampelnya 54 orang. Hasil analisis menunjukkan bahwa nilai  $t$  hitung sebesar 5,818 dengan signifikansi  $< 0,0001$  jauh lebih kecil dari nilai  $\alpha=0,05$ , berarti bahwa ada perbedaan yang signifikan hasil tes yang diperoleh siswa SD favorit bila dibanding dengan siswa SD yang kurang favorit.

Nilai rata-rata siswa SD Negeri 26 Kota Banda Aceh sama dengan 49,57, standar deviasi 17,60 dengan jumlah sampelnya 30 orang, dan nilai rata-rata siswa SD Negeri 9 Kota Banda Aceh sama dengan 40,05, standar deviasi 12,35, dengan jumlah sampelnya 31 orang. Hasil analisis menunjukkan bahwa nilai  $t$  hitung sebesar 2,452 dengan signifikansi 0,017 jauh lebih kecil dari nilai  $\alpha= 0,05$ , berarti ada perbedaan yang signifikan hasil tes yang diperoleh siswa SD favorit bila dibanding dengan siswa SD yang kurang favorit.

Nilai rata-rata siswa MIN Tungkob Kab. Aceh Besar sama dengan 34,78, standar deviasi 16,42 dengan jumlah sampelnya 133 orang, dan nilai rata-rata siswa MIN Rukoh Kota Banda Aceh sama dengan 32,06, standar deviasi 18,12, dengan jumlah sampelnya 76 orang. Hasil analisis menunjukkan bahwa nilai  $t$  hitung sebesar 1,111 dengan signifikansi 0,268 jauh lebih besar dari nilai  $\alpha=0,05$ . Hal ini menunjukkan bahwa tidak ada perbedaan yang signifikan hasil tes yang diperoleh siswa MIN Tungkob Kab. Aceh Besar bila dibanding dengan siswa MIN Rukoh Kota Banda Aceh, di mana kedua MIN tersebut sama-sama salah satu MIN favorit.

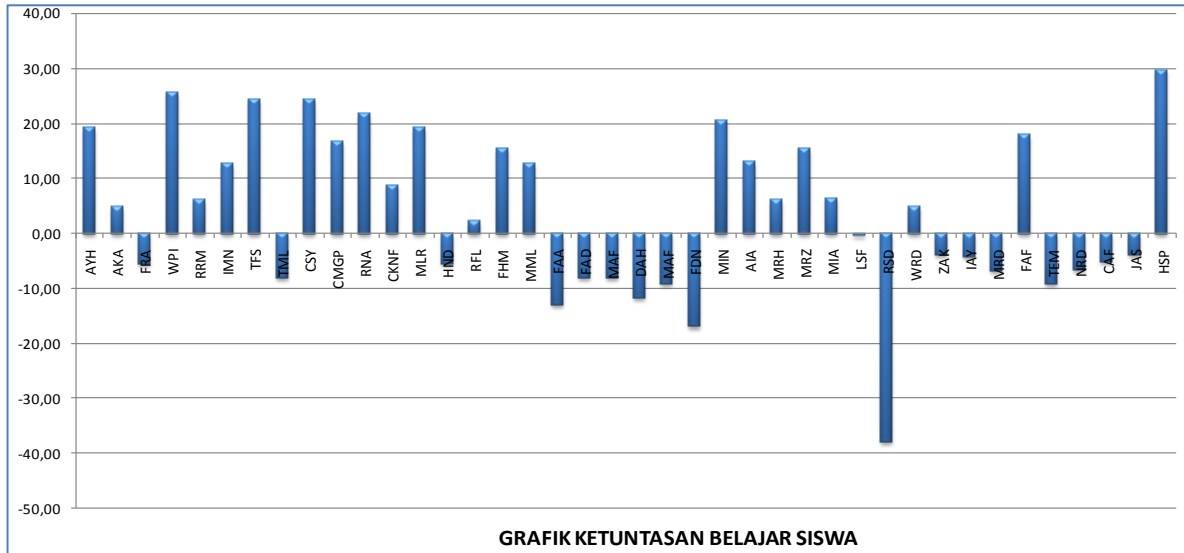
Berdasarkan tiga kelompok sampel yang telah dijelaskan di atas, menunjukkan bahwa TDKB-MatSD yang dikembangkan mampu membedakan secara signifikan kelompok yang favorit dan kelompok non-favorit. Ini berarti TDKB-MatSD yang dikembangkan telah berfungsi sebagai instrumen untuk mendiagnosis kesulitan belajar siswa SD khususnya materi bilangan.

*Software* komputer yang dirancang untuk melakukan analisa hasil diagnosis kesulitan belajar matematika adalah *software Microsoft Office Excel* 2007. Menu utama pada layar pembukanya adalah Data Siswa dan Hasil Tes, Nilai, Profile Siswa, Analisis Semua Jawaban, Analisis Jawaban Benar, Analisis Jawaban Salah, Grafik Ketuntasan Belajar, dan Analisis Indikator.

Menu Grafik (diadaptasi dari Waskito dan Kumaidi, 2010) berisi informasi secara klasikal tentang grafik ketuntasan belajar

siswa di kelasnya bila dibandingkan dengan nilai KKM yang ditetapkan. Apabila nilai yang diperoleh siswa lebih besar dari nilai KKM yang ditetapkan, maka grafiknya akan

muncul di atas garis KKM, apabila nilai yang diperoleh siswa di bawah nilai KKM, maka grafiknya akan berada di bawah garis KKM.



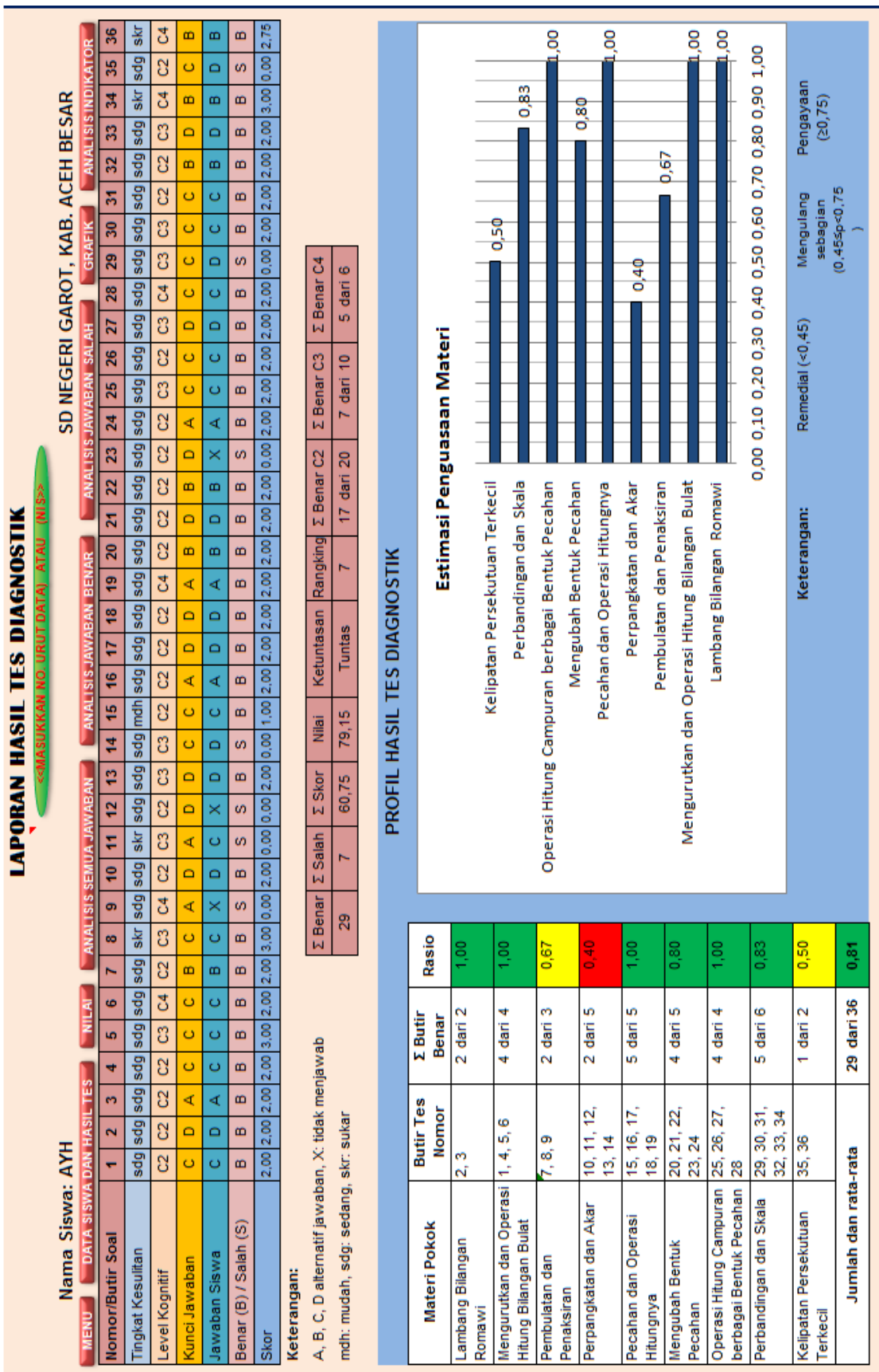
Gambar 1. Grafik Ketuntasan Belajar

Semakin tinggi nilai perolehan di atas KKM maka grafiknya akan semakin tinggi pula. Sebaliknya apabila nilai yang diperoleh siswa jauh di bawah KKM maka grafiknya akan menghunjam ke bawah. Informasi ini sangat berguna bagi guru untuk memperoleh informasi siswa mana saja yang mengalami kesulitan belajar dan siswa mana pula yang tidak mengalami kesulitan belajar. Siswa yang grafiknya di atas garis KKM, terutama siswa-siswa yang grafiknya lebih tinggi dari siswa yang lainnya, dapat dimanfaatkan oleh guru sebagai teman sejawat dalam melakukan remedial terhadap peserta didik yang grafiknya di bawah garis KKM.

Analisis Indikator berisi informasi secara klasikal tentang daftar indikator kemampuan yang diujikan dan nama-nama siswa yang merespons indikator tersebut. Hasil respons tersebut berupa informasi benar (B) atau salah (S) tiap indikator yang diujikan bagi setiap siswa, jumlah siswa yang menjawab dengan benar, jumlah siswa yang belum berhasil menjawab dengan benar,

jumlah siswa yang tidak menjawab, persentase jawaban benar, dan ketuntasan secara klasikal terhadap tiap indikator yang diujikan. Hasil *print out* dari Analisis Indikator yang diujikan sangat berguna bagi guru untuk mendapatkan informasi tentang siswa-siswi mana saja yang menjawab benar dan siswa-siswa mana pula yang belum berhasil menjawab dengan benar indikator yang diujikan. Dengan demikian, guru dapat dengan tepat melakukan remedial bagi siswa yang belum berhasil dan memberikan pengayaan bagi siswa yang sudah berhasil berkenaan dengan indikator yang diujikan.

Hasil analisis komputernya, bukan hanya informasi secara klasikal, tetapi juga individual berupa Profile Hasil Diagnostik secara umum (Gambar 2), hasil analisis semua butir yang direspons oleh siswa ditampilkan dalam menu Analisis Semua Jawaban, analisis jawaban untuk butir yang dijawab benar dalam menu Analisis Jawaban Benar, analisis jawaban untuk butir yang dijawab salah dalam menu Analisis Jawaban Salah.



Gambar 2. Laporan Hasil Diagnostik dan Profilnya

Laporan hasil diagnostik secara umum dalam bentuk Profile Hasil Diagnostik diadaptasi dari model laporan hasil diagnostik yang dikembangkan oleh *Department of Mathematics, Keele University*, yang dikembangkan oleh Rupp, Templin, & Henson (2010, p.31). Perpaduan dari kedua bentuk laporan tersebut dan beberapa informasi penting lainnya, termasuk laporan hasil diagnostik yang dikembangkan berisi informasi secara individual tentang semua jawaban yang direspons terhadap butir soal. Semua laporan tersebut dapat digolongkan ke dalam penilaian autentik, dimana menampilkan sesuai dengan kompetensi yang sebenarnya dimiliki oleh masing-masing siswa menurut kompetensi dasar yang telah dipelajarinya.

Tanggapan guru terhadap produk software komputer untuk menganalisis hasil TDKB-MatSD, diperoleh informasi bahwa software yang dikembangkan layak dan dapat digunakan untuk keperluan mendiagnosis kesulitan belajar matematika jenjang SD/MI dengan rata-rata 3,689 (dalam skala 1 - 4). Hasil prin-outnya dapat dipahami dan berguna sebagai informasi untuk perbaikan pembelajaran.

## Simpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dikemukakan, maka kesimpulan temuan penelitian ini sebagai berikut.

Pengembangan tes diagnostik kesulitan belajar matematika jenjang SD/MI (TDKB-MatSD) ini menggunakan model Research & Development. Tahap penelitian meliputi studi pendahuluan, studi literatur dan hasil-hasil penelitian, analisis masalah, merumuskan *learning continuum*, merumuskan peta konsep, menyusun tes berbentuk *essay*, polarisasi jawaban siswa, dan menyusun tes berbentuk pilihan ganda. Tahap pengembangan meliputi validasi pakar atau praktisi dalam bentuk *focus group discussion*, uji pendahuluan dan keterbacaan, uji coba terbatas, dan uji yang diperluas sehingga diperoleh model yang *fit* secara teoretis dan empiris.

Butir tes yang dikembangkan cocok digunakan untuk diagnosis kesulitan belajar matematika materi bilangan kelas VI SD/MI. Hasil analisis dengan program ITEM-AN diperoleh informasi bahwa kisaran tingkat kesukaran butir antara 0,192 sampai dengan 0,731, kisaran daya beda butir tes 0,221 sampai dengan 0,644, indeks reliabilitas tes 0,889, dan rata-rata daya beda soal 0,579, sedangkan hasil analisis dengan program R diperoleh indeks daya beda butir tes berkisar antara 0,391 sampai dengan 2,317, indeks kesukaran butir tes berkisar antara -2,158 sampai dengan 2,528, kecocokan uji tes dengan kemampuan peserta  $\theta$  berkisar antara -2,00 sampai dengan 2,60, dan fungsi informasi tes berkisar antara 0,111 sampai dengan 3,879.

Fungsi informasi uji tes secara keseluruhan memberikan informasi tertinggi pada kemampuan peserta  $\theta$  sama dengan 0,00 sebesar 31,636. Berarti butir tes yang dikembangkan cocok untuk kemampuan rata-rata peserta, dan sangat cocok digunakan sebagai tes diagnostik.

TDKB-MatSD yang dikembangkan mampu membedakan secara signifikan kelompok yang favorit dan kelompok non-favorit, hal ini menunjukkan bahwa butir tes yang dikembangkan berfungsi sebagai instrumen untuk mendiagnosis kesulitan belajar siswa SD khususnya materi bilangan.

Laporan hasil analisis tes diagnostik yang dimunculkan bermanfaat bagi guru untuk merencanakan dan melaksanakan perbaikan pembelajaran matematika SD/MI khususnya materi bilangan baik secara individual maupun klasikal dalam bentuk remedial teaching.

## Daftar Pustaka

- Borg, W. R., & Gall, M. D. (1983). *Educational research: An introduction*. London: Longman, Inc.
- Fauzan. (2010). *Pengembangan tes diagnostik kesulitan belajar matematika di SMA*. Disertasi. Yogyakarta: PPs UNY.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamental of*

- item response theory*. Newbury Park: SAGE Publications, Inc.
- Kumaidi. (2005). *Profil siswa dan kelas untuk bantuan peningkatan pembelajaran*. Makalah disampaikan dalam Seminar Nasional Hasil Penelitian tentang Evaluasi Hasil Belajar serta Pengelolaannya di PPs UNY.
- Naga, D. S. (1992). *Pengantar teori sekor pada pengukuran pendidikan*. Jakarta: Gunadarma.
- Puspendik, (2010). *Survei internasional TIMSS*. Diakses dari <http://litbangkemdiknas.net/detail.php?id=214>, tanggal: 13 Oktober 2011.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: theory, methods, and applications*. New York: The Guilford Press.
- Sheehan, K. M. (1997). A tree-based approach to proficiency scaling and diagnostic assessment. *Journal of educational measurement*. Volume 34, Issues 4, pages 333-352, December 1997.
- Sriati, Arti. (1994). Kesulitan belajar matematika siswa SMA: pengkajian diagnostik. *Jurnal kependidikan*, Nomor 2, Tahun XXIV. Yogyakarta: UNY
- Subarinah, Sri. (2006). *Pembelajaran matematika sekolah dasar*. Jakarta: Depdiknas.
- Suwarsono, St. (1987). *Trait-treatment interaction dalam pendidikan matematika*. Yogyakarta: IKIP Sanata Dharma Yogyakarta.
- Syarifah. (2007). Persyaratan analisis instrumen sebagai pra-syarat ketepatan hasil analisis dalam penelitian pendidikan. *Jurnal penelitian MIPA*. Vol. 1 No. 1, Desember 2007.
- Thorndike, R. L., & Hagen, E. P. (1972). *Measurement and evaluation in psychology and education*. New York: John Willey & Sonc.
- Waskito & Kumaidi. (2010). *Asesmen untuk memantau kualitas pendidikan*. Pengembangan model analisis prestasi belajar untuk mengidentifikasi kesulitan belajar peserta didik. Jakarta: UIN Syarif Hidayatullah Jakarta.
- Wijaya, Ariyadi. (2012). *Pendidikan matematika realistik suatu alternatif pendekatan pembelajaran matematik*. Yogyakarta: Graha Ilmu.

## **BIAYA OPERASI SEKOLAH MENENGAH KEJURUAN NEGERI KATEGORI SBI/RSBI DI DIY**

*Moch Alip*

Fakultas Teknik Universitas Negeri Yogyakarta

moch\_alip@uny.ac.id

### **Abstrak**

Penelitian bertujuan mengungkap besar biaya operasi (BO) dan dana operasi nonpersonalia dalam penyusunan anggaran Sekolah Menengah Kejuruan Bertaraf Internasional (SMK-BI) di Daerah Istimewa Yogyakarta (DIY). Penelitian evaluatif populatif ini menggunakan model *discrepancy* dengan pendekatan penghitungan biaya berbasis kegiatan (*activity based costing*, disingkat ABC). Data dikumpulkan melalui wawancara dan kaji dokumen kemudian dianalisis secara kualitatif dan kuantitatif. Hasil menunjukkan bahwa pengelola belum pernah menghitung besar BO nonpersonalia per Program Keahlian (PK) sehingga anggaran disusun per sekolah dan tidak mengakomodasi perbedaan kebutuhan bahan dan alat habis pakai (BAHP) antar PK. Besar BO nonpersonalia SMK-BI di DIY hampir sama dengan standar biaya SMK SSN, yaitu Rp2.166.237,00 untuk kelompok bidang non-teknik dan sebesar Rp2.287.066,00 untuk kelompok bidang teknik. Dana operasi nonpersonalia dari Pemerintah sekitar 10% s.d. 16% dari kebutuhan sehingga tidak cukup untuk pengadaan BAHP sekitar 10,5% s.d. 19,4% (idealnya 10% s.d. 30%) sehingga yang ditanggung orang tua siswa sebesar 84% s.d. 90%. Pemerintah perlu mengalokasikan dana operasi program peningkatan mutu sekolah seperti SMK-BI sesuai kebutuhan dan mendorong penerapan anggaran berbasis kinerja.

**Kata kunci:** *biaya operasi SMK*

## **OPERATIONAL COST OF STATE VOCATIONAL HIGH SCHOOLS IN SBI/RSBI CATEGORY IN DIY**

*Moch Alip*

Fakultas Teknik Universitas Negeri Yogyakarta

moch\_alip@uny.ac.id

### **Abstract**

The objectives of this research are to reveal nonpersonal operational cost and funding in composing the budget of world class of vocational school in Yogyakarta Special Province. The evaluatif populatif research used a discrepancy model and an activity based costing approach. Data were collected through interview and document analysis and analyzed by qualitative and quantitative technique. The result showed that the management had not calculated nonpersonal operational cost of each study program, so the school budget did not accommodate the difference cost of consumable material between study programmes. Based on the budget documents, the nonpersonal operational cost per student per year is Rp2.166.237,00 for non-technical programmes and Rp2.287.066,00 for technical programmes. The cost was not much higher than that for standard schools. The fund from government is varies from 10% up to 16%. It is not enough for consumable materials of workshop activities that up to 10,5% for non-technical programmes and 19,4% for technical programmes. The small fund from government made the parent have to pay 84% up to 90% of nonpersonal operational cost, which is too expensive for poor families. The government should allocate appropriate fund when launching special programmes for improving quality of education and implement performance based budgeting.

**Keywords:** *operational cost of vocational high school*

## Pendahuluan

Peningkatan mutu pendidikan termasuk sekolah menengah kejuruan (SMK) merupakan amanat Undang-Undang Republik Indonesia No. 20 Tahun 2003 (UU No 20 Th 2003) tentang Sistem Pendidikan Nasional (Sisdiknas). Salah satu tindak lanjutnya adalah penyelenggaraan rintisan sekolah ber-taraf internasional (RSBI) atau sekolah ber-taraf internasional (SBI) berlandaskan Peraturan Pemerintah Republik Indonesia Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan (PP 19 Th. 2005). Program tersebut menimbulkan pro dan kontra. Pembantu Rektor I UNS (Suara Karya, 27 Mei 2010) menyatakan bahwa pemberian label SBI atau RSBI memang diperlukan sebagai upaya peningkatan mutu supaya lulusan dapat bersaing dengan lulusan dari negara lain tetapi Pemerintah harus ketat dalam menerapkan standar karena biaya mahal tidak menjamin mutu. Namun, Walikota Solo kurang setuju terhadap program tersebut karena biayanya dinilai mahal (Solo Pos 31 Januari 2010). Di sisi lain, pihak sekolah menyatakan bahwa penyelenggaraan SMK ber-taraf internasional (SMK-BI) belum diikuti dengan pendanaan yang memadai. Brock, Marshall, dan Tucker (Washington Post, 2009) mengemukakan 10 langkah menuju *World-Class Schools*, antara lain: pengelolaan yang akuntabel dan penyediaan dana dari pemerintah bagi anak keluarga miskin.

Amat Jaedun (2010) melaporkan bahwa penyelenggaraan SMK RSBI di DIY menghadapi kendala terbatasnya dana, namun belum ada informasi berapa yang dibutuhkan, berapa dana tersedia dari pemerintah khususnya bagi anak miskin, dan berapa yang harus ditanggung masyarakat/orang tua peserta didik. Pemerintah sudah menetapkan besar biaya operasi (BO) untuk sekolah standar nasional (SSN), namun BO untuk SBI belum. Oleh karena itu perlu penelitian untuk memperoleh informasi tentang berapa besar biaya dan dana operasi SMK-BI di DIY sebagai bahan penyusunan kebijakan pemerintah tentang penyediaan dana penyelenggaraan program peningkatan mutu sekolah.

Pendidikan kejuruan merupakan bagian penting dari sistem pendidikan di Austria, German, Liechtenstein, dan Swiss, seperti dikemukakan Shapiro and Purpel (2003) sebagai berikut: "*Vocational education is an important part of the education systems in Austria, Germany, Liechtenstein and Switzerland ...*". Rauner dan Maclean (2008) menyatakan bahwa "*Vocational education is considered a key factor for improving or maintaining the competitiveness of enterprises and national economies*". Lerman (dalam Rauner dan Maclean) menyatakan bahwa:

*"Broader benefit of Vocational Education and Training (VET) for individual and communities. ... VET raises self-esteem, self confidence, social skills, social contacts, and motivation to learn, and clarifies future career prospects. These effects, in turn, improve physical health and reduce depression, teenage pregnancy, and criminal activity."*

Pendidikan kejuruan selain menyumbang peningkatan daya saing usaha dan ekonomi nasional suatu negara dan masyarakat juga meningkatkan kondisi sosial dan psikologis individu peserta didik seperti *self-esteem*, ketrampilan sosial, dan motivasi belajar demi masa depannya, dan selanjutnya menurunkan tingkat depresi, angka kehamilan remaja, dan kriminalitas.

Swiss Agency for Development and Cooperation (SDC) (2005), State Departments of Education of USA (2002), dan Gasskov (2000) mengemukakan konsep senada bahwa pendidikan dan pelatihan (diklat) kejuruan adalah diklat untuk bekerja di industri dan meningkatkan produktivitas personal dan sosial, namun pendanaannya memerlukan campur tangan pemerintah untuk memastikan pemerataan layanan sesuai dengan kebutuhan per jenis program dan kondisi sosial ekonomi peserta didik. Dengan demikian diklat kejuruan tetap diperlukan walaupun memerlukan biaya lebih besar (Onstenk dalam Rauner and Maclean, 2008) walaupun ada yang kurang setuju karena nilai balik diklat kejuruan (*rate of return*) relatif sama dengan sekolah umum (Psacharopoulos, 1985; Tilak dalam Keeves dan Watanabe, 2002; Hoeckel, 2008).

Chalker dan Haynes (1995) menyatakan bahwa *World-Class Schools* merupakan upaya peningkatan mutu pendidikan melalui penerapan standar baru yang mencakup: pendanaan, waktu belajar, jumlah siswa per rombongan belajar (*class size*), persiapan, kurikulum, system penilaian, tatakelola, dan budaya mutu. Program peningkatan mutu SMK melalui program SMK-BI atau sejenisnya juga dilakukan Singapura karena tuntutan global dan sebagai modal pengembangan ekonomi, seperti dilaporkan Law Song Seng (2007) sebagai berikut:

*“Vocational technical education (VTE) system play a crucial role in the social and economic development of a nation. Owing to their nature, they are continuously subject to the forces driving change in the school, industry, society. ... the modern history of VTE, highlights the transformation of the Institute of Technical Education (ITE) as a world class post secondary institution in Singapore”*

Stewart (2010) menyampaikan pendapat sejenis sebagai berikut: *“Singapore has developed a close linkage between education and economic development, studying where they want the economy to grow and then educating a rising workforce to be prepared accordingly”*.

Di Germany, dana penyelenggaraan sekolah kejuruan menjadi tanggungjawab berbagai pihak terkait, seperti dikatakan Shapiro and Purpel (2003) sebagai berikut *“... in Germany a law (the Berufsausbildungsgesetz) was passed in 1969 which regulated and unified the vocational training system and codified the shared responsibility of the state, the unions, associations and chambers of trade ...”*.

Di USA, penyediaan dana penyelenggaraan pendidikan merupakan masalah dinamis (Lunenbergs dan Ornstein, 2000). Pada awal tahun 1930-an pemerintah lokal menanggung sekitar 82,7% biaya pendidikan. Pada tahun 1999 proporsi bergeser menjadi: pemerintah lokal 39,9%; negara bagian 43,9%; Pemerintah Federal 6,1%; dan sumber lain 10,2%. Besar dana yang disediakan bervariasi antarnegara bagian. Misal, di New Jersey US\$ 9318 dan di Utah US\$ 3670 per siswa per tahun. Lerman (dalam Rauner and Maclean, 2008) menyatakan

bahwa *“...State’s responsibility for financing Vocational Education and Training (VET) is often ambiguous”*. Pendanaan pendidikan kejuruan di negara sedang berkembang belum seperti yang diharapkan, karena beberapa hal, seperti daya tampung industri lebih kecil dari jumlah siswa yang akan praktik di industri.

Hoeckel (2008) menyatakan bahwa diklat kejuruan yang baik memerlukan biaya dua sampai tiga kali lipat dari pendidikan umum dan makin mahal bila terkait dengan okupasi yang menuntut penguasaan teknologi baru yang spesifik dan cepat berkembang karena peralatan cepat usang sehingga pemutakhiran peralatan lebih sering dilakukan. Mun C. Tsang (1997) mengemukakan bahwa di Malaysia dan Philipina, biaya pendidikan kejuruan hampir tiga kali lipat dari pendidikan umum karena memerlukan material, peralatan, dan fasilitas fisik yang lebih banyak/luas dan kompleks, serta rasio guru-siswa yang lebih kecil dibanding pendidikan umum.

USA juga meluncurkan program peningkatan mutu seperti dilaporkan Achieve, Inc (2007) sebagai berikut: *“...Ohio’s competitiveness depends on the ability of its school system to produce an educated and skilled populace”*. Dilaporkan juga bahwa program peningkatan mutu sekolah berkaitan dengan system pendanaan:

*“Ensure that funding is fairly allocated and linked to Accountability ... An excellent, well-managed funding system is critical to provide support at all level of system. Evidence from around the world shows that high-performing countries all follow two common-sense principles. First, they assume a certain minimum amount of money required to educate a student. Second, they acknowledge that some students, because of disadvantages and specific need, will cost more than this minimum amount”*.

Jadi, perlu dipastikan bahwa dana pendidikan dialokasikan secara fair dan akuntabel, terutama untuk sekelompok peserta didik dari keluarga yang kurang beruntung secara sosial ekonomi atau karena berkebutuhan khusus.

Dengan demikian, peningkatan mutu melalui SMK-BI merupakan tuntutan yang perlu dilakukan oleh suatu negara dalam menghadapi persaingan global. Konsekuensinya adalah komitmen pemerintah untuk menyediakan dana yang memadai, terutama untuk SMK yang memerlukan dana lebih besar dan sebagian besar peserta didiknya berasal dari golongan ekonomi lemah. Konsekuensi lain adalah manajemen anggaran sekolah yang akuntabel.

Poston (2011) menyatakan bahwa ada empat macam pola penganggaran, yaitu *“line-item or function-object budgeting, activity or program budgeting, incremental budgeting, dan performance budgeting”*. *Function-object budgeting* dapat mengakomodasi perbedaan kebutuhan biaya antarprogram. *Performance-Based Budgeting* (anggaran berbasis kinerja) cocok untuk program dengan tujuan unik. Pola ini lebih akuntabel karena menggambarkan lebih jelas hubungan antara aktivitas yang perlu dilakukan, besar biaya yang dibutuhkan, dan metode evaluasi untuk memastikan apakah tujuan telah tercapai. Penyelenggaraan SBI dapat dikategorikan memiliki tujuan unik (mutu lebih tinggi) sehingga lebih tepat menerapkan anggaran berbasis kinerja dan satuan biaya sesuai kebutuhan.

Menurut Daljono (2011), biaya (*cost*) merupakan pengorbanan sumber ekonomi untuk terlaksananya suatu aktifitas yang memerlukan bahan, tenaga kerja, dan sumber daya *overhead* dapat dihitung menggunakan metode ABC (*activity based costing*). Bahan dan tenaga kerja dibedakan menjadi langsung dan tidak langsung. Sumber daya *overhead* mencakup pemakaian *supplies*, penyusutan nilai alat dan fasilitas, pemeliharaan alat dan fasilitas, penggunaan daya dan jasa, kegiatan supervisi dan pembinaan, dan bahan yang sulit dirinci.

Abbas Ghozali (2010) melaporkan bahwa biaya operasi pendidikan (BOP) mencakup tidak langsung dan langsung. Biaya tidak langsung antara lain biaya untuk rapat, pemantauan dan kegiatan manajemen lain. Biaya langsung mencakup biaya personal dan nonpersonal. Biaya personal adalah gaji dan tunjangan untuk guru dan tenaga

kependidikan lainnya. Biaya nonpersonal adalah untuk pengadaan bahan dan alat habis, perawatan fasilitas, dan daya dan jasa seperti listrik, telepon, *internet*, air dan sejenisnya. Besar dana yang disediakan Pemerintah untuk SMKN di beberapa propinsi Indonesia pada tahun 2004 sudah lebih besar dari SMAN, namun baru sekitar 50% dari total anggaran sekolah dan besar dana per siswa per tahun disajikan pada Tabel 1.

Tabel 1. Perbandingan Dana per Siswa untuk SMA dan SMK

Sekolah	Besar	Sumber		
		APBN/D	Lain	Total
SMA Negeri	Rp	431.000	503.000	934.000
	%	46,15	53,65	100
SMK Negeri	Rp	711.000	519.000	1.230.000
	%	57,78	42,22	100

Angka tersebut memperkuat laporan Lewin (2001) bahwa negara sedang berkembang telah berhasil menaikkan angka partisipasi pendidikan, namun belum mampu menaikkan alokasi subsidi dana pendidikan. Akibatnya, mutu pendidikan kejuruan belum memenuhi tuntutan dunia kerja atau biaya SMK yang baik makin tidak terjangkau oleh peserta didik yang berasal dari golongan ekonomi lemah.

Klein et.al (2000) dan Bray and Thomas (1998) menyebutkan bahwa biaya pendidikan mencakup komponen gaji guru dan staf pendukung, pengadaan dan perawatan peralatan, dan bahan habis pakai. Mun C. Tsang membagi biaya pendidikan menjadi tiga, yaitu: institusi, personal peserta didik, dan *society*. Biaya institusi dibagi menjadi operasi dan kapital. Biaya operasi mencakup biaya personal dan non-personal. Personal terdiri atas guru, administrator, dan staf pendukung. Non-personal mencakup bahan habis, bahan pembelajaran, *utilities*, perawatan dan perbaikan ringan peralatan, dan layanan lain. Biaya kapital mencakup gedung, peralatan, mebel, dan lahan. Biaya personal peserta didik mencakup biaya buku pribadi, transportasi, dls. Biaya *society* tidak dirinci.

Dengan demikian bisa dimengerti bila biaya SMK lebih mahal dari SMA. Di Filipina, perbandingan *recurrent cost* (biaya rutin) pendidikan kejuruan hampir dua kali lebih besar dari pendidikan umum seperti disajikan pada Tabel 2.

Tabel 2. Perbandingan Biaya Sekolah Umum dan Sekolah Kejuruan Teknik

No	Bidang	Persentase (%) Biaya			Jumlah Per siswa
		Staf	Layanan siswa	Lainnya	
1	Umum	85,0	13	1,4	6.080
2	TVET	49,0	49	1,9	11.401

(Peano, Vergel de Dios, Atchoarena, and Mendoza, 2008)

Besar biaya untuk SMK juga terkait dengan jumlah dan jenis alat yang digunakan (Nolker, 1983). Jumlah alat menggunakan satuan “*Student Place*”, yaitu tempat praktek yang dilengkapi dengan alat utama dan alat pendukung untuk pembentukan kompetensi siswa, baik yang digunakan secara individu atau kelompok. Jumlah dan jenis peralatan praktek berdampak pada besar biaya perawatan dan perbaikan, luas gedung laboratorium dan/atau bengkel, serta biaya kapital yang diperlukan. Rumus dasar tentang luas ruang praktek (*work place*) untuk setiap “*Student Place*” adalah: 5 - 7 m<sup>2</sup> untuk kegiatan pembentukan kompetensi kognitif dan 8 - 10 m<sup>2</sup> untuk praktek pembentukan kompetensi psikomotor. Jumlah guru yang dibutuhkan juga banyak sesuai dengan jenis bidang kompetensi yang dibutuhkan dalam proses pembentukan keterampilan kognitif, maupun keterampilan psikomotor.

Dengan demikian, dapat disimpulkan bahwa sekolah menengah kejuruan (SMK) memerlukan biaya lebih besar dari sekolah menengah umum (SMU) karena memerlukan fasilitas, bahan habis pakai, dan guru lebih banyak. Besar biaya bervariasi antar-jenis kompetensi. Namun, peningkatan mutu SMK harus diselenggarakan karena diperlukan peserta didik untuk memperoleh kompetensi sebagai modal mencari pekerjaan

dan juga diperlukan dunia usaha dan industri.

Terkait dengan hal tersebut, penelitian ini bertujuan untuk mengungkap besar biaya operasi dan dana operasi nonpersonalia dalam penyusunan anggaran Sekolah Menengah Kejuruan Bertaraf Internasional di Daerah Istimewa Yogyakarta.

### Metode

Jenis penelitian ini adalah evaluatif model *discrepancy* menggunakan pendekatan penghitungan biaya berbasis kegiatan (*activity based costing*, disingkat ABC), namun tidak secara penuh karena kegiatan tertentu tidak ada datanya dan pola penyusunan anggaran berbeda antarsekolah terteliti. ABC dipilih berdasarkan asumsi bahwa program keahlian tertentu memerlukan lebih banyak bahan praktek sehingga biaya dan anggaran lebih besar (Daljono, 2011; dan Hoeckel, 2008). Data dari dua pola anggaran yang berbeda ditransfer ke dalam pola yang sama, yaitu berbasis pengeluaran seperti yang dipakai pada Permendiknas No 69 Tahun 2009.

Penelitian ini dilakukan di SMK-BI di Propinsi Daerah Istimewa Yogyakarta pada tahun 2011, namun data yang diteliti adalah dokumen anggaran sekolah tahun 2010. Populasi penelitian ini adalah sekolah menengah kejuruan negeri (SMKN) di DIY yang sudah dinyatakan sebagai SBI dan RSBI sebanyak 10 sekolah dan tidak dilakukan sampling. Berdasarkan bidangnya, sekolah dikelompokkan menjadi dua, yaitu SMK Teknik dan Non-Teknik. Variabel pada penelitian ini adalah besar biaya operasi non-personalia, besar dana tersedia, dan komponen kegiatan pada program pembentukan kompetensi lulusan SMK.

Data dikumpulkan melalui wawancara terstruktur dan kaji dokumen. Wawancara relatif singkat karena pada awal wawancara sudah diketahui bahwa istilah “biaya operasi sekolah” dan Permendiknas No 69 Th 2009 belum dikenal oleh pengelola sekolah, sehingga pengumpulan data fokus melalui kaji dokumen Rencana Kegiatan dan Anggaran Sekolah (RKAS). Sumber data (informan)

adalah kepala sekolah, petugas yang ditunjuk kepala sekolah untuk memegang dokumen RKAS dan pengelola anggaran sekolah. Data dianalisis secara deskriptif kualitatif dan kuantitatif. Analisis kualitatif digunakan untuk mengelompokkan jenis pengeluaran. Analisis kuantitatif digunakan untuk menganalisis besar anggaran. Validasi data dilakukan melalui triangulasi sumber data primer, yaitu dokumen diperoleh langsung dari (a) pimpinan sekolah (kepala dan wakil kepala sekolah), (b) pengelola anggaran sekolah, dan (c) kepala Tata Usaha.

### Definisi Operasional

Dalam penelitian ini, ada beberapa hal yang perlu didefinisikan terlebih dahulu, yaitu: (a) dana adalah rupiah yang tersedia untuk suatu kegiatan; (b) anggaran adalah alokasi dana untuk suatu kegiatan yang direncanakan; (c) biaya adalah rupiah yang perlu dibayarkan untuk suatu komponen kegiatan; (d) Biaya personalia adalah gaji dan tunjangan melekat yang harus dibayarkan kepada personil pelaksana kegiatan; (e) biaya operasi nonpersonalia adalah rupiah yang harus dibayar untuk bahan dan alat habis pakai (BAHP) dan jasa yang diperlukan guna terselenggaranya kegiatan yang meliputi: (1) alat tulis sekolah (ATS) untuk pengelolaan sekolah dan proses belajar; (2)

sumber daya listrik, air, jasa telekomunikasi; (3) pemeliharaan dan perbaikan ringan sarana dan prasarana; (4) transportasi perjalanan dinas pendidik dan tenaga kependidikan; (5) konsumsi untuk rapat-rapat, perlombaan dsj; (6) asuransi keselamatan sekolah, pendidik, tenaga kependidikan, dan siswa (peserta didik), kecelakaan siswa saat praktek kerja di industri, dll; (7) jasa pembina kegiatan ekstra kurikuler, (8) BAHP untuk kegiatan praktikum, kebersihan, olah raga, dsj; (9) biaya praktek kerja industri dan uji kompetensi; (10) evaluasi hasil belajar dan pelaporan sekolah kepada pihak yang berwenang; (11) penyusunan kurikulum; (12) kemitraan; (13) pengelolaan, dan (14) kegiatan yang belum teridentifikasi tetapi perlu dilakukan untuk keberhasilan suatu program.

### Hasil

Hasil penelitian mencakup acuan penyusunan RKAS, hasil penghitungan besar BO nonpersonalia, dan distribusi penggunaan dana berdasarkan hasil kaji dokumen anggaran SMKN-BI/RSBI atau unggulan di DIY.

Acuan Penyusunan Rencana Kegiatan Anggaran Sekolah (RKAS) diketahui melalui wawancara. Hasil wawancara tentang acuan penyusunan RKAS disajikan pada Tabel 3.

Tabel 3. Acuan Penyusunan RKAS

No	Acuan Penyusunan Anggaran Operasi	Jawaban		Ringkasan Hasil Wawancara
		Ya	Tidak	
1	Apakah sudah mengacu pada Permendiknas No 69 Th 2009	0	9	Pengelola sekolah belum tahu bahwa ada Permendiknas tersebut sehingga belum ada yang menggunakannya sebagai acuan.
2	Bila tidak mengacu pada Permendiknas No 69 Th 209, Apa acuan penyusunan anggaran sekolah?"	9	0	Pengelola sekolah harus mengacu pada pola yang ditetapkan oleh Pemerintah Kabupaten/Kota karena sebagai syarat pencairan dana dari pemerintah dan dasar pertanggungjawaban anggaran sekolah.
3	Apakah pernah menghitung Besar Biaya per Peserta Didik per PK/KK	0	9	Belum ada pengelola sekolah yang pernah menghitung, kecuali satu sekolah namun penghitungan tidak diselesaikan karena dana sudah turun.
4	Apakah ada Perbedaan anggaran antar PK/KK	-	-	Belum ada yang membedakan anggaran antar program/kompetensi (PK/KK).
5	Lainnya	0	0	

Hasil wawancara tersebut menunjukkan bahwa: a) pengelola sekolah belum tahu bahwa ada Permendiknas No 69 Th 2009 sehingga belum menggunakannya dalam menyusun anggaran sekolah; b) pengelola SMK-BI belum menghitung biaya operasi (BO) per program atau kompetensi (PK/ KK) sebagai dasar penyusunan anggaran sekolah; c) anggaran sekolah disusun ber-

dasarkan arahan Dinas Pendidikan Kodya/ Kabupaten.

Kaji dokumen menghasilkan data bahwa ada dua pola pengelompokan kegiatan pada penyusunan RKAS, yaitu berorientasi pada: (1) pemenuhan Standar Nasional Pendidikan (SPN) dan (2) bidang garapan Wakil Kepala Sekolah/TU Sekolah disajikan pada Tabel 4 dan Tabel 5.

Tabel 4. Pengelompokan Kegiatan dan Anggaran Sekolah Berorientasi pada Pemenuhan Standar Nasional Pendidikan (SPN)

Program/Kegiatan dan Anggaran	
1	Standar Isi (SI)
1	Telaah/Analisis Standar Isi
2	Pengembangan Kurikulum
3	Pengembangan Silabus dan RPP
4	Dst
2	Standar Proses
1	Penyusunan Perangkat Administrasi Pembelajaran
2	Penyusunan Bahan Ajar
3	Tambahan Pelajaran (Les)
4	dst
3	Standar Kompetensi Kelulusan (SKL)
1	Telaah / Bedah SKL
2	Pengembangan SKL
3	Penyusunan KKM
4	dst
4	Standar Pendidik dan Tenaga Kependidikan
1	Kegiatan MKKS/KKKS/K3SK
2	Kegiatan KKG/MGMP
3	Pelatihan Penggunaan Media Pembelajaran
4	Dst
5	Standar Sarana dan Prasarana
1	Pemeliharaan Lingkungan Sekolah
2	Penataan Taman Sekolah
3	Pengadaan Alat-Alat Kebersihan
4	dst
6	Standar Pengelolaan
1	Kegiatan Home Visit
2	Penelusuran Tamatan
3	Penerimaan Peserta Didik Baru (PPDB)
4	dst
7	Standar Pembiayaan Pendidikan
1	Pemberian Beasiswa (Retrivel/Rapus/Prestasi)
2	Penghargaan Guru Berprestasi
3	Partnership
4	Dst
8	Standar Penilaian Pendidikan
1	Ulangan Harian
2	Ulangan Umum Tengah Semester (UTS)
3	Ulangan Umum Akhir Semester (UAS)
4	Dst

Tabel 5. Pengelompokan Kegiatan dan Anggaran Berorientasi pada Bidang Garapan Wakil Kepala Sekolah/TU Sekolah.

No	Bidang Kegiatan	No	Program	Volume
1	WK I: SARPRAS	1	Rapat	6 kegiatan
		2	Transport/Kesejaht.	11 kegiatan
		3	Pengadaan Sarpras	49 kegiatan
		4	Dst	
2	WK II: KURIKULUM			21 kegiatan
3	WK III: KESISWAAN	1	Osis	19 kegiatan
		2	Pramuka	5 kegiatan
		3	Keagamaan	4 kegiatan
		4	UKS	6 kegiatan
		5	Dst	
4	WK IV: HUMAS			11 kegiatan
5	WMM (Wakil Manajemen Mutu)	1	Awarenes	3 kegiatan
		2	Audit Internal	4 kegiatan
		3	Sertifikasi	9 kegiatan
		4	Dst	
6	SUB BAG T.U.	1	Rapat/Konsumsi	5 kegiatan
		2	Gaji PNS	
		3	Daya & Jasa	23 butir
		4	Dst	

RKAS kemudian dikaji dan ditransfer menjadi komponen biaya untuk menghitung besar BO non personalia seperti pola penghitungan pada Permendiknas No 69 Tahun 2009. Transfer menghasilkan 14 jenis komponen biaya sekolah disajikan pada Tabel 6.

Empat belas komponen BO sekolah tersebut mencakup 10 komponen biaya pada Permendiknas dan empat jenis komponen lain, yaitu: (a) penyusunan dan *Benchmark* kurikulum dan bahan ajar dg OECD, (b) kemitraan dengan industri dan *sister school*, (c) pengelolaan, dan (d) kegiatan lain, seperti disajikan Tabel 6.

Tabel 6. Komponen Biaya Operasi Sekolah

Jenis Komponen
1. Pengadaan alat tulis sekolah
2. Biaya daya dan jasa
3. Biaya pemeliharaan dan perbaikan ringan
4. Biaya transportasi dan perjalanan dinas
5. Biaya konsumsi
6. Biaya asuransi
7. Pembina kegiatan ekstra kurikuler
8. Bahan dan alat habis pakai
9. Biaya Prakerin dan Uji kompetensi
10. Biaya evaluasi hasil belajar dan pelaporan
11. Biaya penyusunan dan <i>Benchmark</i> kurikulum dengan OECD
12. Biaya kemitraan dengan industri dan <i>sister school</i>
13. Biaya pengelolaan
14. Biaya kegiatan lain

Besar dana tersedia per sumber dana per peserta didik diketahui dengan mengkaji dokumen. Kaji dokumen RKAS juga menghasilkan data besar dana per sekolah per

sumber dana kemudian diolah menjadi per peserta didik (siswa) disajikan pada Tabel 7 untuk SMK Non-Teknik dan Tabel 9 untuk SMK Teknik sebagai berikut.

Tabel 7. Besar Dana per Sumber per Sekolah per Siswa pada SMK Non-Teknik

Sumber Dana	Besar Dana per Sekolah per Siswa per Tahun (Rp)					
	SMKN A	SMKN B	SMKN C	SMKN D	SMKN E	Rerata
A Total per Siswa	6.994.097	6.497.311	7.018.490	9.547.881	-	7.514.445
B Diknas Kodya/ Kabupaten						
1 Gaji & tunjangan PNS pendidik & tenaga kependidikan lain	4.408.327	3.443.218	3.872.743	3.800.046	Tidak ada data	3.881.084
2 Selain gaji PNS	37.714	506.339	627.782	764.106	920.549	571.298
C Diknas Prop	126.330	210.549	59.493	154.709	135.685	137.353
D APBN Pusat						
1 INVEST	0	0	0	0	0	
2 NON-INVEST	79.115	127.559	241.475	137.395	226.475	162.404
B.2. + C + D2	243.159	844.448	928.749	1.056.210	1.282.710	871.055
E Sponsor/UPJ						-
F Komite Sekolah	2.342.611	2.209.646	1.288.248	3.635.414	2.709.041	2.436.992
G Nonpersonalia Total per Siswa	2.585.770	3.054.093	3.145.747	5.747.834	3.991.751	3.705.039

Tabel 8. Persentase Dana per Sumber per Sekolah per Siswa SMK Non-Teknik

Sumber Dana	Persentase per Sekolah per Siswa per Tahun					
	SMKN A	SMKN B	SMKN C	SMKN D	SMKN E	Rerata
A Total per Siswa	100%	100%	100%	100%	Tidak ada data	100%
B Diknas Kodya/Kab.						
1 Gaji & tunj. PNS	63,0%	53,0%	55,18%	39,80%	Tidak ada data	52,75%
2 Selain gaji PNS	0,54%	7,79%	8,94%	8,00%	Tidak dihitung	6,32%
C Diknas Prop	1,81%	3,24%	0,85%	1,62%	Tidak dihitung	1,88%
D APBN Pusat						
1 INVEST	-	-	-	-		
2 Non-INVEST	1,13%	1,96%	3,44%	1,44%	Tidak dihitung	1,99%
B.2. + C + D2	3,48%	13,00%	13,23%	11,06%	Tidak dihitung	10,19%
E Sponsor/UPJ						
F Komite Sekolah	33,49%	34,01%	18,36%	38,08%	Tidak dihitung	30,98%
G Nonpersonalia Total per Siswa	36,97%	47,01%	44,82%	60,20%	Tidak dihitung	47,25%

Tabel 9. Besar Dana per Sumber per Sekolah per Siswa pada SMK Teknik

Sumber	Besar Dana per Sekolah per Siswa per Tahun (Rp)				
	SMKN A	SMKN B	SMKN C	SMKN D	Rerata
A Total per Siswa	10.806.805	7.326.606	7.370.993	10.905.482	9.102.472
B Diknas Kodya/ Kabupaten					
1 Gaji & tunjangan PNS	4.622.795	4.116.344	5.110.672	4.668.550	4.629.591
2 Selain gaji PNS	177.917		433.372	635.845	415.711
C Diknas Prop	30.711		52.550	206.875	96.712
D APBN Pusat					
1 INVEST	2.059.139	-	-	1.632.396	1.845.767
2 Non INVEST	123.198		12.550	176.626	104.125
B.2. + C + D2	331.825	1.347.278	498.473	1.019.346	799.231
E Sponsor/UPJ			235.588		
F Komite Sekolah	3.793.046	1.862.983	1.526.260	3.585.190	2.691.870
G Nonpersonalia Total per Siswa	6.184.010	3.210.262	2.260.320	6.236.932	4.472.881

Tabel 10. Persentase Dana per Sumber per Sekolah per Siswa pada SMK Teknik

Sumber	Persentase per Sekolah per Siswa per Tahun				
	SMKN A	SMKN B	SMKN C	SMKN D	Rerata
A Total per Siswa	100%	100%	100%	100%	100%
B Diknas Kodya/Kab.					
1 Gaji & tunj. PNS	42,78%	56,18%	69,33%	42,81%	52,78%
2 Selain gaji PNS	1,65%		5,88%	5,83%	3,34%
C Diknas Prop	0,28%		0,71%	1,90%	0,72%
D APBN Pusat					
1 INVEST	19,05%			14,97%	17,01%
2 Non-INVEST	1,14%		0,17%	1,62%	0,73%
B.2. + C + D2	3,1%	18,39%	6,76%	9,35%	9,39%
E Sponsor/UPJ			3,2%		3,2%
F Komite Sekolah	35,10%	25,43%	20,71%	32,88%	28,53%
G Nonpersonalia Total per Siswa	57,22%	43,82%	30,67%	57,19%	47,22%

Distribusi dana untuk biaya operasi non-gaji per komponen diketahui dengan mengkaji dokumen. Distribusi dana untuk biaya operasi sekolah non-gaji per kompo-

nen disajikan dalam Tabel 11 sampai Tabel 14 sebagai berikut. Salah satu komponen pengeluaran tersebut adalah untuk bahan dan alat habis pakai (BHAP).

Tabel 11. Besar Biaya Operasi Non-Gaji per Komponen per Sekolah per Siswa per Tahun pada SMK Kelompok Non-Teknik

No	Komponen	Besar Biaya per Sekolah per Siswa per Tahun (Rp)					
		SMKN A	SMKN B	SMKN C	SMKN D	SMKN E	Rerata
1	Alat Tulis Sekolah	43.956	48.840	79.189	81.066	-	63.263
2	Daya dan Jasa	65.934	77.604	100.724	183.799	-	107.015
3	Pemeliharaan & perbaikan ringan	118.095	404.094	192.906	237.409	-	238.126
4	Transportasi	80.987	28.816	35.477	56.159	-	50.360
5	Konsumsi	27.473	13.187	57.047	-	-	24.426
6	Asuransi	-	-	-	-	-	-
7	Ekstra Kurikuler	174.780	247.911	356.511	526.069	-	326.318
8	BAHP	96.492	61.050	150.621	584.477	380.041	254.536
9	Prakerin/Uji Kom.	101.260	125.379	85.538	317.108	25.410	130.939
10	Evaluasi & Pelpr.	132.001	260.632	155.554	185.142	19.115	183.332
11	Kurikulum	403.024	375.569	47.052	302.007	-	281.913
12	Kemitraan	46.740	74.481	55.432	228.171	-	101.206
13	Kegiatan lain	113.626	41.852	125.395	111.643	-	98.129
14	Pengelolaan	72.161	309.234	95.723	207.792	-	171.228
	Berbagai kegiatan					2.303.431	
	Total Biaya Operasi	1.476.530	2.068.649	1.537.167	3.020.841	2.727.997	2.166.237

Tabel 12. Persentase Biaya Operasi Non-Gaji per Komponen per Sekolah per Siswa per Tahun pada SMK Kelompok Nok-Teknik

No	Total	Dana per Sekolah per Siswa per Tahun (%)					
		SMKN A	SMKN B	SMKN C	SMKN D	SMKN E	Rerata
1	Alat Tulis Sekolah	3,0%	2,4%	5,2%	2,7%		3,33%
2	Daya dan Jasa	4,5%	3,8%	6,6%	6,1%		5,25%
3	Pemeliharaan & perbaikan ringan	8,0%	19,5%	12,5%	7,9%		11,98%
4	Transportasi	5,5%	1,4%	2,3%	1,9%		2,78%
5	Konsumsi	1,9%	0,6%	3,7%			1,55%
6	Asuransi		0,0%				0,00%
7	Ekstra Kurikuler	11,8%	12,0%	23,2%	17,4%		16,10%
8	BAHP	6,5%	3,0%	9,8%	19,3%	13,9%	10,50%
9	Prakerin/Uji Komp.	6,9%	6,1%	5,6%	10,5%	0,9%	6,00%
10	Evaluasi & Pelaporn	8,9%	12,6%	10,1%	6,1%	0,7%	7,68%
11	Kurikulum	27,3%	18,2%	3,1%	10,0%		14,65%
12	Kemitraan	3,2%	3,6%	3,6%	7,6%		4,50%
13	Kegiatan lain	7,7%	2,0%	8,2%	3,7%		5,40%
14	Pengelolaan	4,9%	14,9%	6,2%	6,9%		8,23%
	Berbagai kegiatan					84,4%	
	Total Biaya Operasi	100%	100%	100%	100%	100%	100%

Tabel 13. Besar Biaya Operasi Non-Gaji per Komponen per Sekolah per Siswa per Tahun pada SMK Kelompok Teknik

No	Komponen	Dana per Sekolah per Siswa per Tahun(Rp)				
		SMKN A	SMKN B	SMKN C	SMKN D	Rerata
1	Alat Tulis Sekolah	97.949	73.534	27.057	87.936	71.619
2	Daya dan Jasa	122.008	98.011	83.077	85.513	97.152
3	Pemeliharaan & perbaikan ringan	186.472	323.888	97.952	192.818	200.283
4	Transportasi	75.646	111.360	87.759	39.870	78.659
5	Konsumsi	33.433	47.541	38.096	-	39.690
6	Asuransi	-	-	5.796	-	5.796
7	Ekstra Kurikuler	310.251	131.906	227.737	521.321	297.804
8	BAHP	445.824	380.436	309.119	586.380	430.440
9	Prakerin/Uji Komp.	264.482	113.260	84.467	528.679	247.722
10	Evaluasi & Pelaporan	304.845	101.393	181.258	99.976	171.868
11	Kurikulum	341.742	269.793	315.770	548.093	368.850
12	Kemitraan	61.093	32.804	38.640	80.986	53.381
13	Kegiatan lain	76.074	49.724	96.882	279.899	125.645
14	Pengelolaan	190.385	70.787	35.982	152.560	112.429
Total Biaya Operasi		2.510.204	1.804.437	1.629.592	3.204.032	2.287.66

Tabel 14. Persentase Biaya Operasi Non-Gaji per Komponen per Sekolah per Siswa per Tahun pada SMK Kelompok Teknik

No	Komponen	Dana per Sekolah per Siswa per Tahun (%)				
		SMKN A	SMKN B	SMKN C	SMKN D	Rerata
1	Alat Tulis Sekolah	3,9%	4,1%	1,7%	5%	3,7%
2	Daya dan Jasa	4,9%	5,4%	5,1%		3,9%
3	Pemeliharaan & perbaikan ringan	7,4%	17,9%	7,4%		8,2%
4	Transportasi	3,0%	6,2%	5,4%	1%	4,0%
5	Konsumsi	1,3%	2,6%	0,5%		1,1%
6	Asuransi			0,4%		0,1%
7	Ekstra Kurikuler	12,4%	7,3%	14,3%	17%	12,8%
8	BAHP	17,8%	21,1%	18,9%	20%	19,4%
9	Prakerin/Uji Komp.	10,5%	6,3%	5,18%	18%	9,9%
10	Evaluasi & Pelaporan	12,1%	5,6%	11,1%	3%	8,1%
11	Kurikulum	13,6%	15,0%	19,4%	18%	16,6%
12	Kemitraan	2,4%	1,8%	2,4%	3%	2,3%
13	Kegiatan lain	3,0%	2,8%	6,2%	9%	5,3%
14	Pengelolaan	7,6%	3,9%	2,2%	5%	4,7%
Total Biaya Operasi		100%	100%	100%	100%	100%

## Pembahasan

Panduan SMK-BI mengamanatkan sekolah untuk menghasilkan lulusan yang lebih baik dari lulusan SSN dengan “Menerapkan standar kelulusan lebih tinggi dari Standar Kompetensi Lulusan” berarti memiliki tujuan dengan kriteria berbeda (unik). Menurut Poston (2011), ada konsekuensi dari penetapan tujuan unik, yaitu (a) uraian program dan kegiatan berbeda (unik) dan (b) anggaran disusun berbasis kinerja (*performance based budgeting*) atau pencapaian tujuan dan satuan biaya (*unit cost*) berdasarkan program tersusun. Permendiknas No 69 Tahun 2009 mengamantakan satuan biaya SMK berbeda antar Program Keahlian (PK) dengan indikator besar biaya untuk komponen bahan dan alat habis pakai (BAHP).

Namun, hasil telaah dokumen RKAS SMK-BI di DIY menemukan bahwa satuan BO nonpersonalia tidak berbasis BO per PK dan tidak berorientasi pada tercapainya kompetensi lulusan lebih baik dari SMK SSN seperti tertulis pada panduan SMK-BI. RKAS yang berorientasi pada Standar Nasional Pendidikan (SNP), periksa Tabel 4, memuat kegiatan peningkatan mutu lulusan melalui “*Benchmark* kurikulum dengan kurikulum sekolah dari OECD”, namun belum menyebutkan kriteria “mutu lulusan di atas Standar Kompetensi Lulusan” sebagai target program dan nampak program untuk mencapainya. RKAS yang berorientasi pada bidang garap wakil kepala sekolah, periksa Tabel 5, juga belum memuat indikasi tersebut.

Besar BO nonpersonalia SMK-BI di DIY (Tabel 7 dan Tabel 9) hampir sama dengan angka pada SMK SSN (Permendiknas No 69 Tahun 2009). Besar BO nonpersonalia untuk SMK-BI kelompok bidang non-teknik adalah: rerata Rp2.166.237,00 (terendah Rp1.476.530,00 dan tertinggi Rp3.020.841,00), sedangkan SMK SSN adalah terendah Rp1.860.000,00 dan tertinggi Rp2.050.000,00 per siswa per tahun. Jadi, besar BO personalia SMK-BI tertinggi sedikit lebih tinggi, tetapi yang terendah lebih rendah dari SMK SSN.

Besar BO nonpersonalia SMK-BI kelompok bidang teknik (Tabel 9) adalah: rerata Rp2.287.066,00 (terendah Rp1.629.592,00 dan tertinggi Rp3.204.032,00), sedangkan SMK SSN adalah terendah Rp1.870.000,00 dan tertinggi Rp2.510.000,00 per siswa per tahun. Jadi, besar BO nonpersonalia SMK-BI yang tertinggi sudah lebih tinggi, tetapi yang terendah masih lebih rendah dari SMK SSN.

Persentase BO nonpersonalia per komponen per sekolah per siswa per tahun (Tabel 8 dan Tabel 10) menunjukkan bahwa persentase biaya komponen BAHP pada SMK-BI di DIY kelompok bidang non-teknik adalah 10,50% dan untuk kelompok bidang teknik adalah 19,40%. Persentase tersebut tersebut sedikit lebih kecil dari persentase yang diamanatkan pada standar biaya SMK SSN, yaitu terendah 10% dan tertinggi 20% untuk SMK kelompok bidang non-teknik dan terendah 10% dan tertinggi 30% untuk kelompok bidang teknik.

Dilihat dari sumber dan besar dana per siswa per tahun (Tabel 7 dan Tabel 9) pada butir B2+C+D2), dapat diketahui bahwa persentase dana dari Pemerintah (Pusat, propinsi, dan kabupaten/kota) relatif kecil dibanding kebutuhan, yaitu untuk kelompok bidang non-teknikyaitu sebesar Rp871.055,00 (15,30%) dan untuk kelompok bidang teknik justru lebih kecil yaitu sebesar Rp799,231,00 (9,40%). Dengan demikian, data tersebut memperkuat pendapat pengelola dan guru SMK-BI bahwa sekolah belum merasakan adanya peningkatan dana setelah berstatus sebagai SMK-BI.

Bila dilacak lebih rinci nampak pemerintah kabupaten/kota tertentu hanya mampu menyediakan dana BO nonpersonalia sebesar Rp178.000,00 per siswa per tahun. Jumlah dana tersebut tidak cukup untuk biaya BAHP, yaitu sebesar Rp254.536,00 per siswa per tahun untuk SMK kelompok bidang non-teknik dan Rp436.440,00 per siswa per tahun untuk kelompok bidang teknik (Tabel 14 dan Tabel 16 butir nomor 8). Akibatnya, komite sekolah (orang tua peserta didik) harus menanggung BO nonpersonalia penyelenggaraan

SMK-BI sebesar 90,60% untuk kelompok bidang teknik atau 84,70% untuk non-teknik.

Kondisi demikian menunjukkan bahwa Pemerintah kurang atau tidak konsekuen dalam menyelenggarakan program peningkatan mutu lulusan seperti SMK-BI dan sejenisnya dengan tidak menyediakan dana yang memadai, khususnya untuk peserta didik dari kelompok ekonomi lemah. Akibatnya, timbul polemik tentang mahalannya biaya SMK-BI.

Besar BO nonpersonalia penyelenggaraan SMK-BI lebih besar dari Standar Biaya SMK SSN merupakan kondisi yang tidak perlu dipersoalkan bila dilihat dari adanya empat komponen pengeluaran tambahan, yaitu (1) *benchmark* kurikulum dan bahan ajar dg OECD, (2) kemitraan dengan industri dan *sister school*, (3) kegiatan lain, dan (4) pengelolaan. Masalah yang perlu dicari solusinya adalah proporsi dana dari Pemerintah, orang tua peserta didik, dan pihak lain yang berkepentingan.

## Penutup

Berdasarkan penelitian dapat disimpulkan bahwa biaya operasi (BO) nonpersonalia penyelenggaraan SMK-BI/RSBI di DIY adalah sebagai berikut: (a) pengelola belum pernah menghitung besar BO nonpersonalia per Program Keahlian (PK) karena pengelola sekolah belum tahu bahwa ada Permendiknas No 69 Tahun 2009 tentang Standar Biaya SSN dan anggaran sekolah disusun sesuai dengan arahan Dinas Pendidikan Kabupaten/Kota yang belum mengakomodasi perbedaan kebutuhan BAHF antar-PK; (b) rerata besar BO nonpersonalia per satuan pendidikan (bukan per PK) adalah sebesar Rp2.166.237,00 untuk kelompok bidang non-teknik dan sebesar Rp2.287.066,00 untuk kelompok bidang teknik, berarti hampir sama dengan BO nonpersonalia pada SMK SSN; c) dana untuk BO penyelenggaraan SMK-BI/RSBI dari Pemerintah relatif kecil, yaitu 10% s.d. 16% sehingga tidak cukup untuk pengadaan BAHF yang mencapai 10,5% untuk kelompok non-teknik dan 19,4% untuk kelompok teknik. Akibatnya,

komite sekolah (orang tua siswa) harus menanggung sebagian besar (84% sampai dengan 90%) BO nonpersonalia, suatu beban yang cukup berat bagi orang dari golongan ekonomi lemah.

Berdasarkan hasil penelitian, dapat disarankan sebagai berikut: (a) pemerintah perlu lebih aktif melakukan sosialisasi Permendiknas No 69 Tahun 2009 tentang Standar Biaya SSN dan menghitung biaya program peningkatan mutu sekolah seperti SMK-BI atau sejenisnya; (b) pemerintah mengalokasikan dana penyelenggaraan program peningkatan mutu sekolah seperti SMK-BI sesuai dengan kebutuhan diikuti dengan panduan penerapan anggaran berbasis kinerja untuk meningkatkan akuntabilitas penggunaan dana.

## Daftar Pustaka

- Abbas Ghazali (2010). *Ekonomi Pendidikan*. Jakarta: Lemlit UIN Syarif Hidayatullah.
- Achieve, Inc. (2007). *Creating a World-Class Education System in Ohio*. [www.achieve.org](http://www.achieve.org)
- Amat Jaedun. (2010). Pencapaian IKKT pada Operasi SMK RSBI di DIY. *Jurnal Pendidikan Teknologi dan Kejuruan Vol 19 No 2 Th 2010*.
- Badan Penelitian dan Pengembangan Depdiknas. (2007). *Pedoman Penjaminan Mutu Sekolah/Madrasah Bertaraf Internasional pada Jenjang Pendidikan Dasar dan Menengah*. Jakarta: Depdiknas.
- Bray, M. (2002). *The Cost and Financing of Education: Trend and Policy Implication*. Manila: ADB
- Brock, W., Marshall, R., and Tucker, M. (2009). *10 Steps to World-Class Schools*. [www.washingtonpost.com](http://www.washingtonpost.com)
- Daljono. (2011). *Akuntansi Biaya*. Semarang: Badan Penerbit Universitas Diponegoro.
- Chalker, D.M. dan Haynes, R.M. (1995) *World-class schools: new standards for education*. Lancaster: Technomic.

- Depdiknas. (2002). *Kegiatan belajar mengajar kurikulum berbasis kompetensi*. Jakarta: Puskur Balitbang.
- Gasskov, V. (2000). *Managing Vocational Training Systems*. Geneva: ILO.
- Hoeckel, K, (2008). *Cost and Benfits in Vocational Education ang Training*. Edu/edpc/cert
- Klein, S., Hoachlauder,G., Beltrnena, R., Beal, T. (2000). *Funding of Vocational Education in Ohio: Results of National Survey of State*. Berkeley: MPR Associates
- Klein, S. 2001. *Financing Vocational Education: A State Policy Maker's Guide*. Berkeley: MPR Associates
- Law Song Seng. (2007). *Vocational Technical Education and Economic Development- The Singapore Experience*. <https://www.ite.edu.sg>.
- Lewin, K. And Caillods, F. (2001). *Financing Secondary Education in Developing Countries: Strategies for Sustainable Growth*. Paris: IIEP.
- Lunenberg and Ornstein. (2000). *Educational Administrattion*. Belmont: Wadsworth.
- Nolker, H. 1983. *Pendidikan Kejuruan: Pengajaran, Kurikulum, Perencanaan*. Jakarta: Gramedia.
- Mun C. Tsang. 1997. Cost of Vocational education. *International Journal of Manpower*. Vol 18, No 1/2, pp. 63-89. University Press.
- Peano, S., Vergel de Dios, B., Atchoarena, D., and Mendoza, U. (2008). *Invesment in Technical Vocational Education (TVET) in Philippines*. Paris: IIEP UNESCO
- Poston, Jr. W.K. 2011. *School Budgeting*. Thousand Oaks: Sage.
- Psacharopoulos, G ang Woodhall, M. (1985). *Education for Development: An Analysis for Investment Choices*. New York: Oxford University Press.
- Rauner, F. And Maclean, R. (Eds). (2008). *Hanbook of Tecnical and Vocational Education and Training Research*. Bremen: UNESCO.
- Shapiro, H. S. and Purpel, D. E. (2003) by American Educational Studies Association. From "Funding Accountability: States,Courts, and Public Responsibility," *Educational Studies*, Vol. 34, No. 3, pp. 317–336.
- State Departments of Education. (2002). *Vocational Education*. *Education.state university.com*.
- Stewart (2010). Singapore: Innovation in Technical Education. <http://www.apec knowledgebank.org>
- Swiss Agency for Development and Cooperation. (2005). *Costing and Pricing of VET Products and Services*. Berne: SDC.
- Tilak, J.B.G. dalam Keeves, J.P. and Watanabe, R. (2002), *Vocational Education and Training in Asia*. New Delhi: National Institute of Educational Planning.
- Yin Cheong Chen. (1996). *School Efectiveness & School-based Management*. London: Falmer Press.

## **MODEL EVALUASI REFLEKTIF KURIKULUM PENDIDIKAN BAHASA INGGRIS DALAM PENGEMBANGAN KARAKTER BANGSA**

<sup>1)</sup>Monika Gultom, <sup>2)</sup>Sumarno, <sup>3)</sup>Suwarsih Madya,

<sup>1)</sup>Universitas Cenderawasih Jayapura, <sup>2,3)</sup>Universitas Negeri Yogyakarta  
<sup>1)</sup>monikagultom27@gmail.com, <sup>3)</sup>suwarsihm@yahoo.com

### **Abstrak**

Penelitian ini bertujuan untuk mengembangkan model evaluasi reflektif kurikulum rumpun mata kuliah keahlian pendidikan bahasa Inggris untuk memotret penuangan nilai-nilai karakter bangsa di dalam kelas. Penelitian pengembangan ini menggunakan pendekatan campuran dalam lima tahap, yaitu studi awal, pendefinisian, perancangan, peragaan, dan pengembangan. Penetapan konstruk dilakukan melalui penilaian pakar, dengan subyek penelitian dosen dan mahasiswa di Program Studi Pendidikan Bahasa Inggris yang ada di DIY. Penentuan koefisien reliabilitas instrumen evaluasi menggunakan program SmartPLS Versi 2.0.M3 dan uji kelayakan model dengan program GeSCA. Hasil penelitian menyimpulkan: (1) model evaluasi yang dikembangkan didukung oleh lima instrumen evaluasi; (2) validitas, reliabilitas, dan kepraktisan instrumen evaluasi telah teruji melalui kegiatan FGD, teknik Delphi, dan analisis faktor konfirmatori non-parametrik; (3) koefisien reliabilitas komposit kelima instrumen adalah sebesar 0,93; 0,91; 0,95; 0,95; dan 0,86; (4) hasil uji kelayakan model menunjukkan model didukung oleh data ditunjukkan oleh nilai GFI 0,99 dan SRMR 0,06; 5) Pedoman penggunaan model dalam bentuk panduan evaluasi.

**Kata kunci:** *pengembangan model evaluasi kurikulum, karakter bangsa*

## **MODEL OF REFLECTIVE CURRICULUM EVALUATION OF ENGLISH LANGUAGE EDUCATION IN DEVELOPING NATIONAL CHARACTER**

<sup>1)</sup>Monika Gultom, <sup>2)</sup>Sumarno, <sup>3)</sup>Suwarsih Madya

<sup>1)</sup>Universitas Cenderawasih Jayapura, <sup>2,3)</sup>Universitas Negeri Yogyakarta  
<sup>1)</sup>monikagultom27@gmail.com, <sup>3)</sup>suwarsihm@yahoo.com

### **Abstract**

This study aimed to develop a model of reflective curriculum evaluation at the English Education Study Program that portrayed the delivery of the values of national character in the classroom. This development study used mixed approach carried out in five phases: initial study, defining, designing, demonstrating, and developing. The constructs of instruments were developed based on expert judgments; the subjects were lecturers and students of English Language Education Study Program in four universities in Yogyakarta Special Territory. The composite reliability coefficient of the instruments was analyzed by SEM program of SmartPLS Version 2.0.M3 and test of fit model by GeSCA. The result of the study concluded: 1) the developed model evaluation was supported by five instruments of evaluation; 2) validity, reliability, and its implementation had been verified through FGD, Delphi techniques, and non-parametric confirmatory factor analysis; 3) the composite reliability coefficient of the five instruments was 0.93, 0.91, 0.95, 0.95, and 0.86; 4) the result of the test of fit model indicated that model was supported by the data with GFI 0.99 and SRMR 0.06; 5) The guidelines of using the developed model consisted of background, rational, components of assessment, guidelines, and application sample.

**Keywords:** *developing model of curriculum evaluation, national character*

## **Pendahuluan**

Kemajuan yang telah dicapai bangsa Indonesia dalam hal pembangunan fisik harus diimbangi dengan pembangunan non-fisik, termasuk di antaranya membangun karakter dan jati diri bangsa agar menjadi bangsa yang kokoh dan berpendirian teguh (Pemerintah RI, 2010: p.16). Pentingnya pembangunan karakter dan jati diri bangsa telah diamanatkan dalam UUD 1945, yaitu agar Pemerintah Indonesia mengusahakan dan menyelenggarakan satu sistem pendidikan nasional yang meningkatkan keimanan, ketakwaan kepada Tuhan Yang Maha Esa, dan akhlak mulia serta mencerdaskan kehidupan bangsa.

Amanat UUD 1945 ini berimplikasi pada pengembangan kurikulum untuk selalu mengacu pada standar nasional pendidikan untuk mewujudkan tujuan pendidikan nasional dengan memperhatikan beberapa aspek penting yang diantaranya adalah persatuan nasional dan nilai-nilai kebangsaan. Amanat UUD 1945 ini juga merupakan tantangan bagi pendidikan bahasa asing karena kental dengan pengenalan dan penguasaan terhadap sejarah, kebudayaan, karakteristik, pranata sosial budaya dan nilai-nilai yang dimiliki penutur asli bahasa target. Sementara, PP Nomor 19 Tahun 2005 mensyaratkan terpenuhinya standar isi dan proses terkait dengan pelaksanaan pembelajaran pada satuan pendidikan yang merupakan elemen penting dalam kurikulum sebagai pedoman penyelenggaraan pembelajaran bermuatan karakter bangsa.

Kontribusi besar pendidikan karakter dalam proses pembelajaran selaras dengan aspek filosofis, ideologis, normatif, historis, dan sosiokultural yang melatarbelakangi tercetusnya Kebijakan Nasional Pembangunan Karakter Bangsa tahun 2010-2025 (Pemerintah RI, 2010:, pp.1-3). Secara filosofis, pembangunan karakter bangsa bersifat esensial dalam proses berbangsa karena hanya bangsa yang memiliki karakter dan jati diri yang kuat yang akan bertahan; secara ideologis, pembangunan karakter merupakan pengejawantahan ideologi Pancasila dalam kehidupan berbangsa dan bernegara;

secara normatif, pembangunan karakter bangsa merupakan wujud nyata langkah mencapai tujuan Negara; secara historis, pembangunan karakter bangsa merupakan dinamika inti proses kebangsaan yang terjadi tanpa henti dalam kurun sejarah; dan secara sosiokultural, pembangunan karakter bangsa merupakan suatu keharusan dari suatu bangsa yang multikultur.

UU Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional dan Ketetapan Kementerian Pendidikan Nasional Badan Penelitian dan Pengembangan Pusat Kurikulum (2010, pp.9-10) mendudukkan pentingnya pendidikan di bidang budaya dan karakter bangsa. Nilai semangat kebangsaan dan cinta tanah air dideskripsikan sebagai cara berpikir, bertindak, dan berwawasan yang menempatkan kepentingan bangsa dan Negara di atas kepentingan diri dan kelompoknya, dan cara berpikir, bersikap, dan berbuat yang menunjukkan kesetiaan, kepedulian, dan penghargaan yang tinggi terhadap bahasa nasional, lingkungan fisik, sosial, budaya, ekonomi dan politik bangsa.

Meskipun banyak ahli menyatakan pentingnya membentuk karakter bangsa melalui pendidikan, misalnya Ki Hadjar Dewantara yang menekankan pentingnya membentuk karakter bangsa melalui pendidikan atau pengajaran yang selaras dengan kehidupan bangsa (Ki Hadjar Dewantara, 2004), dalam praktiknya belum sepenuhnya terlaksana. Beberapa penelitian menunjukkan banyak program pendidikan guru yang secara tidak sungguh-sungguh mempersiapkan calon guru untuk menjadi guru yang tidak hanya menguasai ilmu pengetahuan dan keterampilan mengajar tetapi juga bagaimana menjadi pendidik nilai-nilai moral (Narvaez & Lapsley, 2006; Berkowitz & Bier, 2004, p.72).

Terlebih lagi, beberapa penelitian lain menunjukkan bahwa kurikulum bidang pendidikan memberikan lebih sedikit pengajaran tentang aspek-aspek moral kepada mahasiswa dibandingkan dengan kurikulum pada pembelajaran pada disiplin ilmu lainnya (Lampe, 1994; McNeel, 1994 dalam Leming, 2000, pp.421-2), padahal, kurikulum

dengan penekanan pembelajaran bermuatan karakter terbukti memberikan kontribusi yang lebih besar ketimbang variabel lainnya. Dengan kata lain, muatan kurikulum menunjukkan kesiapan dan keseriusan suatu lembaga pendidikan dalam membentuk karakter bangsa peserta didik melalui pendidikan. Muatan kurikulum menunjukkan bagaimana dan seberapa jauh lembaga pendidikan mempersiapkan peserta didik memegang peranan penting pengembangan karakter (Bath, Smith, Stein & Swann, 2004), terutama dalam menjawab tantangan zaman, perubahan kondisi, pergantian kebijakan, dan juga dalam merespon kebutuhan pasar lokal.

Sementara itu, belum banyak dilakukan evaluasi kurikulum tentang muatan nilai-nilai karakter bangsa dalam konteks pembelajaran, mengingat evaluasi kurikulum dalam konteks pembelajaran dapat membantu mengetahui apa yang sebenarnya terjadi di dalam kelas (Martone & Sireci, 2009, p.1332). Sebagai faktor penentu, evaluasi terhadap kurikulum berarti merujuk pada apa yang sebenarnya terjadi di dalam kelas, termasuk perencanaan pembelajaran dan implementasi kurikulum di dalam kelas (Richards, 2001). Evaluasi kurikulum kebanyakan berpegang pada hasil pengukuran skor prestasi belajar siswa dalam menentukan keberhasilan suatu kurikulum, dengan asumsi yang dipegang adalah berdasarkan kepada efisiensi, efektivitas biaya, kemampuan siswa, dan rata-rata perolehan nilai siswa (Apple & Beyer, 1983, p.425). Mengingat institusi pendidikan seperti sekolah atau universitas bukanlah suatu entitas yang terisolasi, atau terpisah dari masalah-masalah ekonomi, sosial, dan budaya, maka penelitian yang dilaporkan disini difokuskan pada pengembangan suatu model evaluasi kurikulum yang merefleksikan muatan karakter bangsa.

Berdasarkan penelitian prasarvei yang dilakukan pada awal tahun 2011 yang lalu, beberapa lembaga pendidikan bahasa Inggris di Provinsi DIY menunjukkan adanya komitmen pada pengembangan karakter dengan variasi model pendekatan pembel-

ajaran yang beragam. Rancangan model evaluasi pada penelitian ini merupakan kombinasi dari model evaluasi CIPP (Worthen & Sanders, 1984) dan kerangka *Logic Model* (Taylor-Powell, Jones & Henert, 2003), dan dikembangkan untuk dapat melihat kekuatan dari masing-masing model pendekatan pembelajaran dalam Prodi pendidikan bahasa Inggris tersebut.

Pendidikan bahasa Inggris memiliki karakteristik yang sangat kental dengan pengenalan dan penguasaan akan sejarah, kebudayaan, karakteristik, pranata sosial budaya dan nilai-nilai yang dimiliki oleh penutur asli. Peluang dan tantangan yang dimiliki lembaga pendidikan bahasa Inggris ini dapat dijawab dengan adanya suatu alat evaluasi yang mampu meneropong dan mengontrol bagaimana penanaman nilai-nilai karakter bangsa dalam kerangka pengembangan pendidikan nilai budaya dan karakter bangsa dituangkan dalam pembelajaran bahasa Inggris tumbuh di dalam kesadaran peserta didik.

Lembaga pendidikan, dalam hal ini Prodi PBI, mengemban tugas dan tanggung jawab yang besar dalam menghasilkan calon guru yang tidak hanya memiliki intelektualitas yang tinggi, terlebih penting lagi memiliki karakter yang utuh dengan integritas yang tinggi dalam mendidik generasi penerus bangsa yang berkarakter. Calon guru yang dididik dengan penanaman nilai-nilai karakter bangsa tersebut merupakan ujung tombak pencapaian cita-cita luhur bangsa sehingga perlu dibekali dengan penghayatan dan kesadaran akan tugas dan fungsi utamanya sebagai pendidik.

Sebagai perwujudan cita-cita luhur bangsa Indonesia yang tercantum dalam Pembukaan UUD 1945, pengimplementasian pembangunan karakter bangsa dalam kurikulum dan pembelajaran di kelas perlu diberi dukungan sepenuhnya, terutama untuk menjawab tantangan arus globalisasi. Pendidikan bahasa Inggris memiliki potensi yang besar dalam mewujudkan cita-cita luhur bangsa melalui pembangunan karakter peserta didik yang utuh dan juga nasionalis yang tertuang dalam kurikulum.

Kurikulum adalah rangkaian perencanaan pendidikan yang mencakup pengalaman seorang anak dalam lembaga pendidikan, keseluruhan pengalaman di kelas yang direncanakan dan diperankan dosen, keseluruhan pengalaman yang disediakan untuk mahasiswa sehingga mereka dapat memperoleh keterampilan dan pengetahuan dalam berbagai konteks pembelajaran (Ornstein & Huskins, 2009, pp.10-11; Parkay & Anctil, 2010, p.2; Marsh, 2009, p.7). Kurikulum dalam pembelajaran bahasa adalah rangkaian proses perancangan, perbaikan, pelaksanaan, dan evaluasi program pembelajaran bahasa (Richards, 2001, p.2). Dalam hal ini, kurikulum bersifat spesifik pada apa yang terjadi di dalam kelas dimana dosen dan mahasiswa menjadi kurikulum dari pembelajaran tersebut.

Kurikulum dalam pengertian di atas disebut dengan silabus dan dipahami sebagai spesifikasi dari isi dan urutan pembelajaran yang akan diajarkan, instrumen yang membantu pengembang silabus untuk mencapai kesesuaian antara apa yang dibutuhkan dan tujuan belajar peserta didik sebagai makhluk sosial dan sebagai individu, dan dengan semua kegiatan belajar yang terjadi di dalam kelas (Nunan, 2004, p.3). Dalam hal ini, proses penyusunan dan pengembangan kurikulum dalam pembelajaran bahasa adalah hasil olah pikir seorang pengembang kurikulum berdasarkan pemahamannya tentang pendidikan dan peserta didik yang akan menerima pembelajaran (Jackson, 1992, p.21), memegang peranan penting dalam menentukan penugasan muatan karakter bangsa dalam kurikulum pembelajaran yang diselenggarakannya.

Karakter bangsa adalah kualitas perilaku kolektif kebangsaan yang khas-baik yang tercermin dalam kesadaran, pemahaman, rasa, karsa, dan perilaku berbangsa dan bernegara sebagai hasil olah pikir, olah hati, olah rasa dan karsa, serta olah raga seseorang atau sekelompok orang (Pemerintah RI, 2010, p.7). Karakter bangsa dalam pandangan Anthropologi adalah tata nilai budaya dan keyakinan yang mengejawantah dalam kebudayaan masyarakat dan meman-

carkan ciri-ciri khas keluar sehingga dapat ditanggapi oleh orang luar sebagai kepribadian masyarakat tersebut (Saifuddin & Karim, 2008, p.8). Karakter bangsa sebagai nasionalisme adalah keyakinan dari suatu komunitas yang memiliki pengalaman yang berbeda tetapi disatukan dalam satu rezim politik (Meadow, 2008). Dengan demikian, karakter bangsa berarti suatu keyakinan yang terbentuk secara sosial, yang mempererat dan membentuk cara berkomunikasi suatu komunitas yang membentuk dan menghasilkan proposisi-proposisi dan hal-hal simbolis bersama yang menjadi identitas nasional komunitas tersebut.

Karakter bangsa sebagai identitas nasional dimaknai sebagai suatu komunitas masyarakat yang hidup dalam suatu wilayah/teritori ataupun dari luar wilayah/teritori suatu Negara, sebagai bagian dari anggota komunitas politik yang memiliki budaya dan ideologi yang sama dan juga kesamaan keturunan, etnik, ikatan persaudaraan, saling berbagi dalam hal budaya dan ikatan-ikatan sosial, meliputi semua hal yang secara hukum dimiliki seorang warga Negara diluar etnik atau budaya tertentu (Parmenter, 1999, p.455). Dalam pandangan ini, karakter bangsa adalah identitas nasional yang terbentuk tidak hanya berdasarkan batas-batas geografi maupun kategori-kategori etnik semata, tetapi lebih menitikberatkan pada ikatan saling memberi dan berbagi dalam hal budaya dan ikatan-ikatan sosial.

Dengan demikian, karakter bangsa sebagai identitas nasional dapat dipahami sebagai bentuk kesadaran atau sikap nasionalisme yang menyatukan suatu bangsa secara politik, kewilayahan, sejarah, dan ideologis. Meskipun batasan tentang nasionalisme ini memiliki pengaruh langsung terhadap kelas-kelas pembelajaran bahasa asing, bahkan beberapa ahli berpendapat ideologi nasionalisme mengganggu keefektifan pembelajaran bahasa asing (Meadows, 2008, p.118), sekolah atau lembaga pendidikan mempunyai tanggung jawab membentuk moral masyarakat atau karakter kebangsaan yang sesuai dengan yang dikehendaki masyarakatnya (Sari & Doğanay, 2009). Dalam hal ini, mo-

ral masyarakat yang sesuai dengan kehendak masyarakatnya adalah nilai-nilai moral yang merupakan kepribadian bangsa tersebut.

Nilai-nilai Pancasila, sebagai kepribadian bangsa Indonesia, yang terkandung dalam ke-lima silanya merupakan cerminan kepribadian bangsa yaitu karakter keIndonesiaan yang merupakan jati diri bangsa yang telah ditetapkan dalam ketetapan MPR RI No. VII/MPR/1978 adalah karakter yang religius, manusiawi, bersatu, demokratis, dan adil (Winarno, 2010, p.10). Karakter bangsa Indonesia yang khas baik tercermin dalam kesadaran, pemahaman, rasa, karsa dan perilaku berbangsa dan bernegara berdasarkan nilai-nilai Pancasila, norma UUD 1945, prinsip *Bhinneka Tunggal Ika*, dan komitmen terhadap NKRI (Pemerintah RI, 2010, p.7).

Keempat pilar kebangsaan yang merupakan dasar nilai, norma, prinsip, dan komitmen hidup berbangsa dan bernegara ini harus menjadi haluan pembangunan bangsa Indonesia melalui proses sosialisasi, pendidikan dan pembelajaran, pembudayaan, dan kerja sama seluruh komponen bangsa dan Negara secara koheren. Karakter bangsa Indonesia yang terbentuk dari karakter-karakter individu adalah karakter yang berlandaskan falsafah Pancasila yang dijiwai dari ke-lima sila Pancasila secara utuh dijabarkan sebagai karakter pribadi bangsa Indonesia sebagai karakter berKetuhanan Yang Maha Esa, karakter kemanusiaan, karakter kebangsaan, karakter kerakyatan, dan karakter berkeadilan sosial. Fokus penelitian yang dilaporkan di sini difokuskan pada evaluasi muatan karakter kebangsaan sebagai karakter pribadi bangsa Indonesia yang merupakan pencerminan sila ketiga Pancasila.

Karakter kebangsaan memiliki beberapa indikator pencerminan karakter bangsa Indonesia dari sila ketiga Pancasila "Persatuan Indonesia" yaitu: menempatkan persatuan, kesatuan, kepentingan dan keselamatan bangsa di atas kepentingan pribadi atau golongan, rela berkorban demi kepentingan bangsa dan Negara, bangga sebagai bangsa Indonesia yang bertanah air Indonesia serta menjunjung tinggi bahasa Indonesia, dan

memajukan pergaulan demi persatuan dan kesatuan bangsa yang ber-*Bhinneka Tunggal Ika*.

Karakter kebangsaan yang terbentuk dengan baik di sekolah secara langsung dan tidak langsung merupakan upaya pembentukan sikap dan kesadaran nasionalisme suatu bangsa dalam menghadapi masa depan yang penuh dengan tantangan. Tantangan masa depan yang diwarnai dengan pesatnya kemajuan teknologi komunikasi dan informasi, interaksi antarbangsa, antarkebudayaan, dan antar peradaban semakin mencair, terbuka di dunia virtual yang nyaris tidak mengenal batas ruang dan waktu (Sumarno, 2011, p.74), dapat berdampak pada terkikisnya identitas diri dan pudarnya karakter bangsa. Ancaman pudarnya karakter bangsa ini menjadikan penyusunan suatu model evaluasi kurikulum amat strategis dalam kaitannya dengan pemanfaatan informasi yang diperoleh dari hasil evaluasi terhadap kurikulum tersebut.

Informasi yang diperoleh dari data hasil evaluasi digunakan untuk tiga macam keputusan, yaitu menentukan perlu tidaknya perubahan materi dan metode pembelajaran, perencanaan pembelajaran, menginformasikan kemajuan peserta didik, dan menilai efektifitas suatu program atau seorang pengajar (Davis, 1980, p.25). Dalam hal ini, evaluasi kurikulum dipandang sebagai suatu proses pengumpulan dan penggunaan informasi yang akan digunakan untuk membuat keputusan terhadap suatu program pendidikan, dan memberikan rekomendasi yang diperlukan dalam pengambilan keputusan.

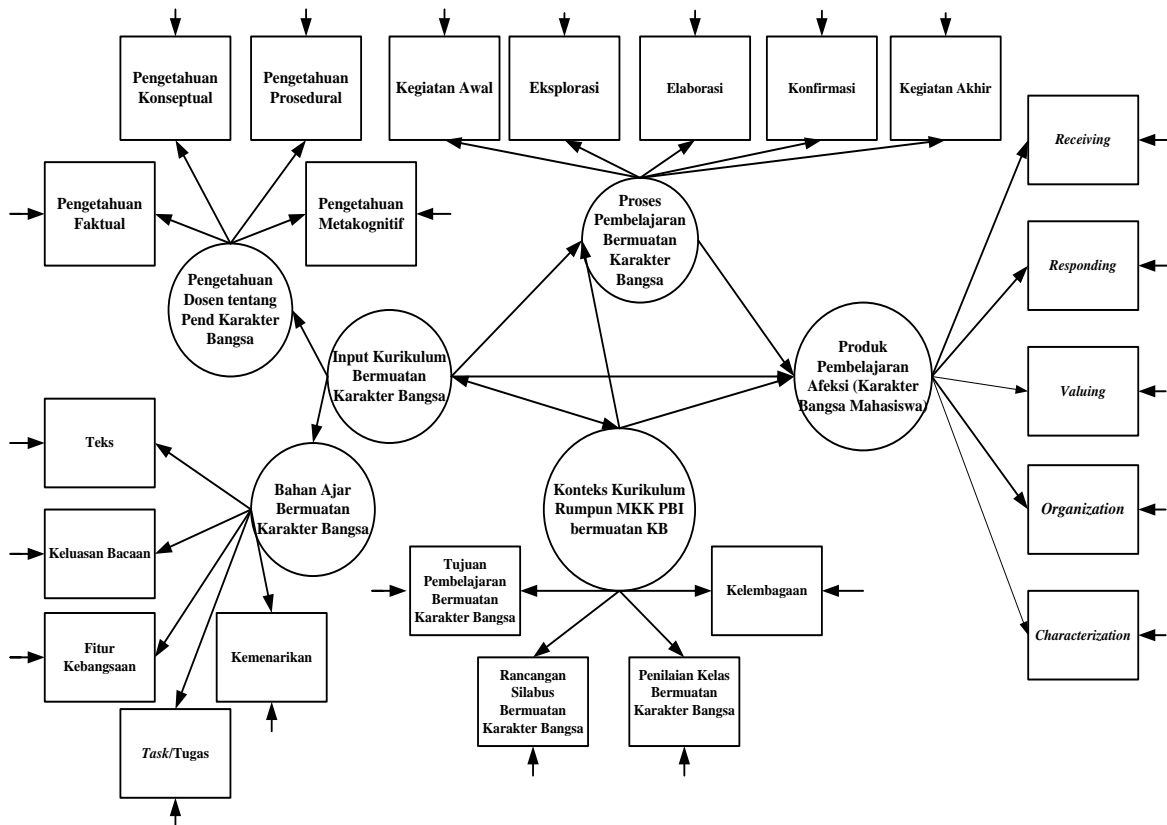
Proses pengumpulan dan penggunaan informasi untuk mengambil keputusan dalam evaluasi kurikulum dilakukan untuk menjawab pertanyaan tentang seberapa baik pelaksanaan suatu kurikulum termasuk upaya yang dilakukan dalam mencapai tujuan (Beane, Toepfer & Alessi, 1986, p.265). Karena itu, evaluasi harus dilakukan secara berkesinambungan pada seluruh aspek kurikulum dengan memperhatikan pertanyaan tentang tujuan, elemen-elemen perencanaan awal apakah sesuai dengan kebutuhan, minat, dan karakteristik peserta didik, kualiti-

tas interaksi antara peserta didik dan guru, performa guru dan peserta didik selama PBM, dan tercapainya tujuan yang telah ditetapkan (Davis, 1980, p.23). Dengan demikian, peranan guru/dosen dalam melakukan evaluasi kurikulum menjadi sangat strategis sebagai pihak yang dapat mengobservasi, mendeskripsikan, dan menganalisis interaksi yang terjadi di dalam kelas.

Mengingat pentingnya peranan guru/dosen dalam melakukan evaluasi kurikulum, tiap lembaga pendidikan termasuk pendidikan bahasa Inggris perlu senantiasa melakukan evaluasi kurikulum secara internal demi memastikan penuangan muatan karakter bangsa dalam kurikulum pembelajaran yang diberikan. Untuk itu, evaluasi kurikulum yang mampu menangkap adanya penuangan

karakter bangsa dalam perencanaan dan implementasi kurikulum diperlukan sebagai alat kontrol dan bahan penyedia masukan bagi pimpinan lembaga pendidikan dan juga bagi staf pengajar bahasa asing itu sendiri seperti yang akan dilaporkan dalam penelitian ini.

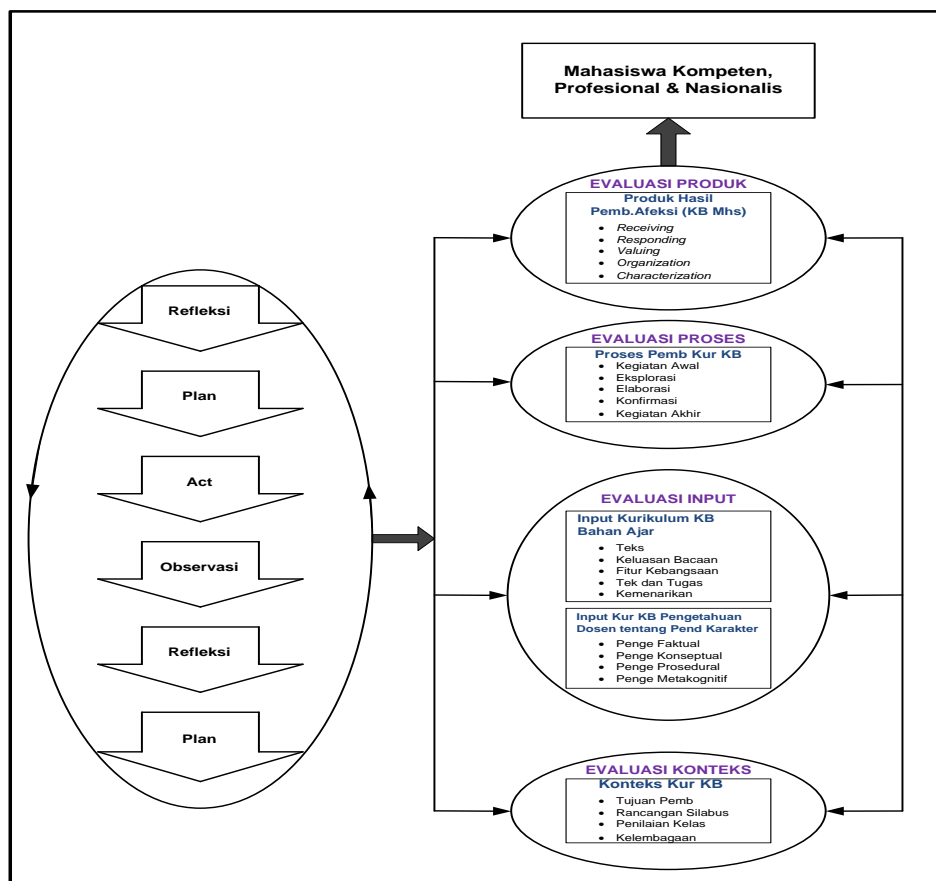
Model evaluasi reflektif kurikulum pendidikan bahasa Inggris dalam pengembangan karakter bangsa yang dikembangkan dalam penelitian ini memuat elemen-elemen evaluasi kurikulum, yaitu: muatan karakter bangsa dalam konteks kurikulum, pengetahuan dosen tentang pendidikan karakter bangsa, muatan karakter bangsa dalam bahan ajar, muatan karakter bangsa dalam proses pembelajaran, dan evaluasi produk pembelajaran afeksi. Hubungan konseptual model evaluasi disajikan pada Gambar 1.



Gambar 1. Kerangka Konseptual Model Evaluasi Reflektif Kurikulum MKKPBIKB

Prosedur kegiatan evaluasi yang dikembangkan merupakan hasil modifikasi model evaluasi CIPP dari Stufflebeam dan kawan-kawan yang dipadu dengan komponen-komponen evaluasi dalam *Logic Model* yang disajikan pada Gambar 2. Terkait de-

ngan hal tersebut, penelitian ini bertujuan untuk mengembangkan model evaluasi reflektif kurikulum rumpun mata kuliah keahlian pendidikan bahasa Inggris untuk memotret penuangan nilai-nilai karakter bangsa di dalam kelas.



Gambar 2. Prosedur Kegiatan Evaluasi Reflektif Kurikulum Rumpun MKK PBI dalam Pengembangan Karakter Bangsa

## Metode Penelitian

### Jenis Penelitian

Penelitian ini menggunakan pendekatan kualitatif dan kuantitatif (campuran). Pendekatan ini digunakan untuk membantu menjelaskan hal-hal yang tidak dapat dijelaskan dengan pendekatan kuantitatif semata, seperti menjelaskan apabila terdapat bias, *underestimated* atau *overestimated* penilaian oleh responden. Pengembangan model ini mengadaptasi tahapan pengembangan model spiral dari Cennamo & Kalk (2005) dengan tahapan seperti berikut: (1) penelitian, (2) pendefinisian, (3) perancangan, (4) peragaan, dan (5) pengembangan.

Model dikembangkan mengadaptasi model spiral dimana tiap fase diuraikan sebagai berikut: fase pertama terdiri dari kegiatan penelitian kajian pustaka dan hasil-hasil penelitian, penggalan informasi mengenai penerapan pendidikan karakter dalam kurikulum pembelajaran, pengumpulan

data tentang upaya Prodi PBI dalam penentuan muatan karakter bangsa melalui kegiatan prasurvei, tahap kedua dilakukan penentuan lingkup kegiatan, produk pengembangan yang akan dihasilkan, subjek uji, populasi, instrumen dan teknik analisis data yang akan digunakan dan jadwal kegiatan. tahap ketiga dilakukan perencanaan rancangan model, pola dasar, alur kerja, dan instrumen evaluasi reflektif kurikulum MKK PBI dalam pengembangan karakter bangsa, tahap keempat dilakukan kegiatan pengujian produk yang merupakan kegiatan pengujian kelompok terbatas yang dilanjutkan dengan kegiatan revisi dan analisis sampai akhirnya dihasilkan model tentatif evaluasi reflektif kurikulum MKK PBI dalam pengembangan karakter bangsa, dan tahap kelima adalah tahap lanjutan dari tahap peragaan (*demonstrate*) yaitu berupa kegiatan pengujian lapangan utama dan kegiatan revisi/modifikasi model.

Waktu dan Tempat Penelitian

Tahapan penelitian terdiri atas tahap penelitian dan tahap pengembangan yang dilaksanakan di empat Prodi Pendidikan bahasa Inggris yang ada di DIY, yaitu Prodi Pendidikan Bahasa Inggris Universitas Sanata Dharma (USD), Universitas Sarjanawiyata Tamansiswa (UST), Universitas Negeri Yogyakarta (UNY), dan Universitas Ahmad Dahlan (UAD) selama kurang lebih 1 tahun.

Subjek Penelitian

Penelitian ini merupakan penelitian yang pemilihan sampelnya menggunakan metode *purposive sampling* dengan subjek penelitian melibatkan responden sebanyak 5 orang dosen dan 70 orang mahasiswa pada kegiatan uji coba pertama (kelompok terbatas) dan sebanyak responden 20 orang dosen dan 291 orang mahasiswa dari 4 Prodi Pendidikan Bahasa Inggris yang ada di DIY, yaitu USD, UST, UNY, dan UAD.

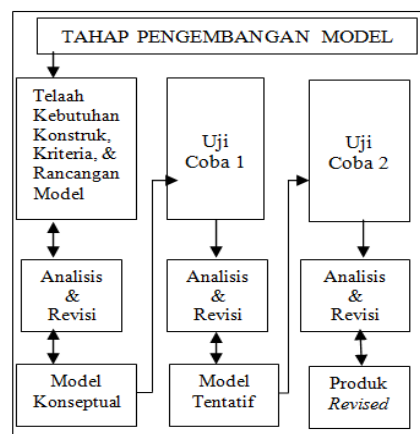
Prosedur

Prosedur yang dilakukan dalam penelitian pengembangan ini terdiri dari tahap penelitian dan tahap pengembangan model, sehingga dengan mengadaptasi prosedur pengembangan dari model spiral Cennamo & Kalk (2005) prosedur penelitian ini terdiri dari 5 tahap, yaitu: tahap penelitian, pendefinisian, perancangan, peragaan, dan pengembangan.

Pada tahap penelitian dilakukan kegiatan penelitian kajian pustaka dan hasil-hasil penelitian terkait berbagai teori tentang kurikulum, karakter bangsa, model evaluasi, dan model evaluasi kurikulum, sehingga dapat ditentukan komponen-komponen model evaluasi yang dikembangkan. Selanjutnya, pada tahap ini juga dilakukan kegiatan penggalan informasi mengenai penerapan pendidikan karakter dan upaya menuangkan muatan karakter bangsa dalam proses pembelajaran dengan melakukan observasi keempat program studi pendidikan bahasa Inggris di DIY, serta wawancara dengan Ketua Program Studi, Kepala dan Sekretaris Penjaminan Mutu, dan beberapa dosen di program studi pendidikan bahasa Inggris tersebut.

Untuk melengkapi data tentang masing-masing program studi, peneliti mengumpulkan dan mencermati dokumen program studi berupa buku kurikulum, pedoman model pembelajaran, buku karya pemikiran akademik sivitas akademik, pedoman evaluasi pembelajaran, uraian tentang kurikulum, sejarah, tujuan pendidikan, visi dan misi, *course outline*, rubrik penilaian, bahan kuliah, materi-materi khusus, pedoman akademik 2012/2013 Fakultas Keguruan dan Ilmu Pendidikan, dan juga refleksi konsep pendidikan (Ki Hadjar Dewantara, Kemuhammadiyah and an Pedagogi Ignasian).

Pada tahap pengembangan model dilakukan kegiatan penelitian yang mengadaptasi model pengembangan Cennamo & Kalk (2005) terdiri dari 5 tahap kegiatan, yaitu pendefinisian, perancangan, peragaan, pengembangan, dan penyajian. Pada tahap pendefinisian dilakukan kegiatan telaah kebutuhan, konstruk, dan kriteria model evaluasi, dilanjutkan dengan perancangan model, alur kerja dan instrumen evaluasi. Pada tahap selanjutnya dilakukan kegiatan peragaan yaitu berupa kegiatan pengujian lapangan pada kelompok terbatas dan dilanjutkan dengan revisi dan analisis, diteruskan dengan kegiatan pengembangan model pada kelompok yang lebih besar, revisi dan analisis, hingga menghasilkan model yang telah direvisi. Kegiatan selanjutnya adalah kegiatan lanjutan berupa penyajian/sosialisasi dan pengumpulan saran, masukan, dan pengembangan lanjutan yang akan dilakukan dalam penelitian lanjutan. Prosedur pengembangan model disajikan pada Gambar 3.



Gambar 3. Tahapan Pengembangan Model

## Data, Instrumen, dan Teknik Pengumpulan Data

Jenis data yang akan diperoleh terdiri dari data kualitatif yang diperoleh dari hasil telaah pustaka, hasil wawancara, observasi kelas, dan studi dokumen terhadap Ketua Program Studi, dan dosen dari Prodi PBI USD, UST, UNY, dan UAD, dan data kuantitatif diperoleh dari hasil olah CVR (*content validity ration*) dari penilaian rekan sejawat, dari hasil penyebaran angket, data hasil uji model dari pengolahan data menggunakan analisis faktor konfirmatori (CFA) dengan program SmartPLS dan juga uji kecocokan model menggunakan program GeSCA.

Instrumen yang digunakan berdasarkan jenis data yang terdiri dari wawancara, diskusi terbuka dalam kegiatan FGD, teknik Delphy untuk kegiatan penilaian pakar, lembar pengamatan untuk memperoleh data observasi pelaksanaan pembelajaran di dalam kelas, dan angket terstruktur untuk informasi penilaian dosen dan mahasiswa tentang penuangan muatan karakter bangsa dalam kurikulum pembelajaran pendidikan bahasa Inggris.

## Teknik Analisis Data

Teknik analisis data kualitatif menggunakan teknik analisis deskriptif dalam menganalisis kajian pustaka dalam membangun teori terkait konstruk dan kriteria dalam model evaluasi yang dikembangkan, mendeskripsikan hasil penilaian pakar, dan deskripsi tentang hasil uji coba pada kelompok terbatas maupun uji coba lapangan yang lebih luas dengan menggunakan *cross-case analysis*.

Teknik analisis data kuantitatif menggunakan CVR (*content validity ration*) dalam melakukan uji validitas isi, analisis faktor konfirmatori (CFA) dalam menganalisis data hasil uji coba kelompok terbatas dan kelompok yang lebih luas dengan program SmartPLS, analisis reliabilitas komposit dan Alpha dalam melihat konsistensi internal butir instrumen, dan analisis uji kecocokan model menggunakan program GeSCA untuk melihat apakah model yang dikem-

bangkan telah sesuai atau cocok antara teori dan penerapannya di lapangan.

## Hasil Penelitian dan Pembahasan

Hasil penelitian terdiri dari hasil pengembangan dan hasil uji coba produk. Data yang dihasilkan pada hasil pengembangan terbagi atas hasil penelitian dan hasil pengembangan sedangkan hasil uji coba produk terbagi atas hasil validitas isi, seleksi butir dan reliabilitas instrumen, uji kecocokan model, dan kriteria penyekoran (kategorisasi) berdasarkan model yang dikembangkan.

### Hasil Pengembangan

#### *Hasil Penelitian*

Hasil penelitian menunjukkan beberapa permasalahan dalam penerapan pendidikan karakter bangsa dalam kurikulum pembelajaran pendidikan bahasa Inggris yang ada di Provinsi DIY adalah (1) konteks kurikulum MKK dituangkan masih secara implisit sehingga perlu diperiksa lebih lanjut muatan karakter bangsa dalam kurikulum MKK yang diselenggarakan, (2) pengetahuan dosen tentang pendidikan karakter di keempat Prodi dipandang sudah tinggi, tetapi pelatihan tentang penuangan nilai-nilai karakter tersebut belum terprogram secara rutin, (3) belum ada mekanisme yang dapat digunakan untuk memeriksa muatan karakter bangsa dalam bahan ajar yang diberikan, (4) belum ada wadah yang mengakomodir informasi tentang muatan karakter bangsa dalam proses pembelajaran baik itu dari dosen maupun mahasiswa, dan (5) penelusuran tentang produk pembelajaran afeksi (karakter bangsa mahasiswa) masih belum banyak disentuh demi memastikan internalisasi nilai-nilai kebangsaan dalam diri mahasiswa.

#### *Hasil Pengembangan*

Hasil pengembangan menunjukkan bahwa pada tahap pendefinisian (*defining*) dihasilkan rumusan komponen model evaluasi untuk mengidentifikasi muatan karakter bangsa dalam kurikulum MKK PBI yang terdiri atas empat aspek yaitu: (1) muatan karakter bangsa dalam konteks kurikulum

MKK PBI, (2) muatan karakter bangsa dalam input kurikulum MKK PBI, (3) muatan karakter bangsa dalam PBM, dan (4) produk pembelajaran afeksi (karakter bangsa mahasiswa).

Tahap perancangan (*designing*) menghasilkan kisi-kisi instrumen evaluasi muatan karakter bangsa dalam kurikulum MKK PBI dengan total jumlah 141 butir. Selanjutnya dilakukan validasi, analisis, dan revisi sebanyak 4 kali dan uji coba sebanyak 2 kali untuk mengetahui apakah indikator yang disusun dalam kisi-kisi sudah terwakili dalam instrumen yang dikembangkan.

Pada tahap validasi teman sejawat jumlah butir yang tidak dapat dipertahankan sebanyak 12 butir sehingga total jumlah butir menjadi 129 butir yang tersebar dalam 5 instrumen. Pada tahap validasi isi yang melibatkan 10 orang panelis, sebanyak 6 butir yang tidak dapat dipertahankan sehingga total butir menjadi 123 butir yang tersebar pada 5 instrumen. Pada tahap validasi pakar dihasilkan validasi dan revisi kejelasan tujuan evaluasi dan komponen evaluasi muatan karakter bangsa dalam kurikulum MKK PBI, kriteria penyekoran, dan instrumen evaluasi. Pada tahap validasi praktisi dihasilkan revisi kejelasan model evaluasi, instrumen evaluasi, dan panduan evaluasi berdasarkan saran dan masukan para praktisi.

Tahap peragaan (*demonstrating*) menghasilkan revisi hasil uji coba instrumen tahap pertama berdasarkan informasi empirik butir-butir yang termuat dalam instrumen evaluasi yang dikembangkan, menggunakan analisis faktor konfirmatori. Dukungan tiap *item* terhadap faktor-faktor dan reliabilitas komposit tiap instrumen disajikan pada Tabel 1.

Tahap pengembangan (*developing*) menghasilkan revisi analisis uji coba tahap kedua berdasarkan informasi empirik tentang butir-butir yang termuat dalam instrumen evaluasi menggunakan analisis faktor konfirmatori dengan menunjukkan dukungan tiap *item* terhadap faktor-faktor dan reliabilitas komposit tiap instrumen disajikan pada Tabel 2.

Tabel 1. Sebaran Butir Instrumen Uji Coba Tahap Pertama

Instrumen Evaluasi	Butir Terpakai	Rel komp	Jumlah Resp
Muatan KB dlm Konteks KurMKK PBI Pengetahuan Dosen ttg Pend Karakter	25	0,94	75
Muatan KB dlm Bahan Ajar	18	0,96	75
Muatan KB dlm PBM Produk Pemb Afeksi	34	0,92	75
	29	0,88	75
Jumlah	123		75

Tabel 2. Sebaran Butir Instrumen Uji Coba Tahap Kedua

Instrumen Evaluasi	Butir Terpakai	Rel komp	Jumlah Resp
Muatan KB dlm Konteks KurMKK PBI Pengetahuan Dosen ttg Pend Karakter	20	0,95	311
Muatan KB dlm Bahan Ajar	17	0,92	311
Muatan KB dlm PBM	17	0,95	311
Muatan KB dlm PBM Produk Pemb Afeksi	27	0,94	311
	23	0,88	311
Jumlah	106		311

Hasil Uji Coba Produk

Validitas Isi

Instrumen yang telah direvisi berdasarkan masukan teman sejawat selanjutnya dinilai kembali oleh 10 orang panelis yang memiliki keahlian dalam bidang evaluasi, pengukuran, pendidikan, statistik, dan psikologi, untuk melihat relevansi antara indikator dan butir. Masukan panelis terhadap kelima instrumen dianalisis menggunakan *content validity ratio* (CVR) dari Lawshe (1975) dengan rumus berikut:

$$CVR = \frac{n_e - N_2}{N_2}$$

Keterangan:

- $n_e$  : Jumlah panelis yang berpendapat butir relevan
- $N_2$  : Jumlah total panelis

Hasil analisis validitas isi menggunakan analisis CVR (*content validity ratio*) oleh

10 orang panelis menunjukkan bahwa terdapat total 6 butir tidak berkualitas karena tidak memenuhi nilai minimum CVR 0,62 sedangkan hasil analisis validitas instrumen dengan menggunakan CVI (*content validity index*) diperoleh indeks validitas isi kelima instrumen valid karena memenuhi nilai CVI > 0,51 (Lawshe, 1975, p.572), berikut ini:

Tabel 3. Hasil Analisis CVR Instrumen Evaluasi Reflektif Kurikulum MKKPBIKB

Instrumen Evaluasi	Butir Berkualitas	Butir Tdk Berkualitas
Muatan KB dlm Konteks Kur MKK PBI	25	2
Pengetahuan Dosen ttg Pend Karakter	17	1
Muatan KB dlm Bahan Ajar	18	-
Muatan KB dalam PBM	34	1
Produk Pemb Afeksi	29	2
Jumlah	123	6

Tabel 4. CVI Instrumen Evaluasi Reflektif Kurikulum MKKPBIKB

CVI Instrumen Evaluasi	
Muatan KB dlm konteks Kurikulum MKK PB	0,86
Pengetahuan Dosen ttg Pendidikan Karakter	0,94
Muatan KB dlm Bahan Ajar	0,94
Muatan KB dalam PBM	0,91
Produk Pembelajaran Afeksi	0,94

*Seleksi Butir dan Reliabilitas Instrumen*

Analisis data hasil uji coba instrumen evaluasi dilakukan dalam dua tahap, yakni uji coba pada kelompok terbatas dan uji coba pada kelompok yang lebih besar dengan melakukan analisis faktor konfirmatori menggunakan program SmartPLS Versi 2.03 M3 oleh Ringle, Wende & Will (2005). Hal ini bertujuan melihat dukungan tiap *item* terhadap faktor-faktor yang terdapat dalam tiap indikator atau kekuatan hubungan pada masing-masing indikator terhadap variabel laten yang terdapat dalam instrumen evaluasi yang dikembangkan. Berikut hasil seleksi butir kelima instrumen:

Tabel 5. Seleksi butir pada Uji Coba Instrumen Tahap Pertama

Instrumen Evaluasi	Butir Terpakai	Rel komp	Jumlah Resp
Muatan KB dlm Konteks Kur MKK PBI	20	0,95	75
Pengetahuan Dosen ttg Pend Karakter	17	0,92	75
Muatan KB dlm Bahan Ajar	17	0,95	75
Muatan KB dlm PBM	27	0,94	75
Produk Pemb Afeksi	23	0,88	75
Jumlah	106		75

Tabel 6. Seleksi butir pada Uji Coba Instrumen Tahap Kedua

Instrumen Evaluasi	Butir Terpakai	Rel komp	Jumlah Resp
Muatan KB dlm Konteks Kur MKK PBI	20	0,95	311
Pengetahuan Dosen ttg Pend Karakter	17	0,92	311
Muatan KB dlm Bahan Ajar	17	0,95	311
Muatan KB dlm PBM	25	0,95	311
Produk Pemb Afeksi	21	0,88	311
Jumlah	100		311

Untuk menetapkan butir-butir yang akan digunakan dilakukan analisis korelasi butir yang memiliki koefisien korelasi butir terhadap total sebesar  $\geq 0,30$  dengan muatan faktor  $\geq 0,5$  dan reliabilitas komposit > 0,70 seperti berikut:

Tabel 7. Penetapan Butir Setelah Seleksi Butir pada Uji Coba Instrumen Tahap Kedua

Instrumen Evaluasi	Butir Terpakai	Rel komp	Jumlah Resp
Muatan KB dlm Konteks Kur MKK PBI	19	0,95	311
Pengetahuan Dosen ttg Pend Karakter	17	0,92	311
Muatan KB dlm Bahan Ajar	17	0,95	311
Muatan KB dlm PBM	25	0,95	311
Produk Pemb Afeksi	14	0,85	311
Jumlah	106		311

Tabel 8. Reliabilitas Kelima Instrumen Evaluasi Reflektif Kurikulum MKKPBIKB

Instrumen Evaluasi	Reliabilitas Komposit
Muatan KB dlm konteks Kurikulum MKK PB	0,93
Pengetahuan Dosen ttg Pend Karakter	0,91
Muatan KB dlm Bahan Ajar	0,95
Muatan KB dalam PBM	0,95
Produk Pemb Afeksi	0,86

Uji Kecocokan Model

Sebelum dilakukan uji kecocokan model menggunakan model persamaan struktural, terlebih dahulu dilakukan uji kelayakan variabel dengan menggunakan Program PASW 18. Jika nilai Kaiser-Meyer-Olkin (KMO)  $\geq 0,50$  dan nilai signifikansi pada *Bartlett's Test of Sphericity*  $\leq 0,05$  maka variabel layak diuji lebih lanjut. Hasil uji kelayakan menunjukkan bahwa variable layak diuji lebih lanjut karena memiliki nilai KMO = 0,917 dan nilai signifikansi Bartlett's Test of Sphericity  $\leq 0,05$  seperti ditunjukkan pada Tabel 9.

Tabel 9. Nilai KMO dan Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	0,917
Bartlett's Test of Sphericity	
Approx. Chi-Square	3372,478
Df	253
Sig.	0,000

Selanjutnya, untuk menganalisis model persamaan struktural digunakan *software* program SmartPLS Versi 2.0 yang merupakan analisis model persamaan struktural berbasis komponen dari Ringle, Wende & Will (2005), sedangkan untuk uji kecocokan model digunakan *software* program *Generalized Structured Component Analysis* (GeSCA) yang dikembangkan oleh Heungsun Hwang (2011).

Untuk menguji fit model, digunakan pasangan hipotesis sebagai berikut:

$H_0$ : Model sesuai dengan/didukung data  
 $H_1$ : Model tidak sesuai dengan/didukung data

Kriteria ujinya adalah:

Terima  $H_0$  bila nilai GFI mendekati 1 atau  $\geq 0,98$  dan SRMR (*standardized root mean square residual*) mendekati 0 atau  $\leq 0,08$  merupakan indikasi kelayakan model yang dapat diterima

Terima  $H_1$  bila nilai GFI mendekati 0 dan SRMR (*standardized root mean square residual*) mendekati 1.

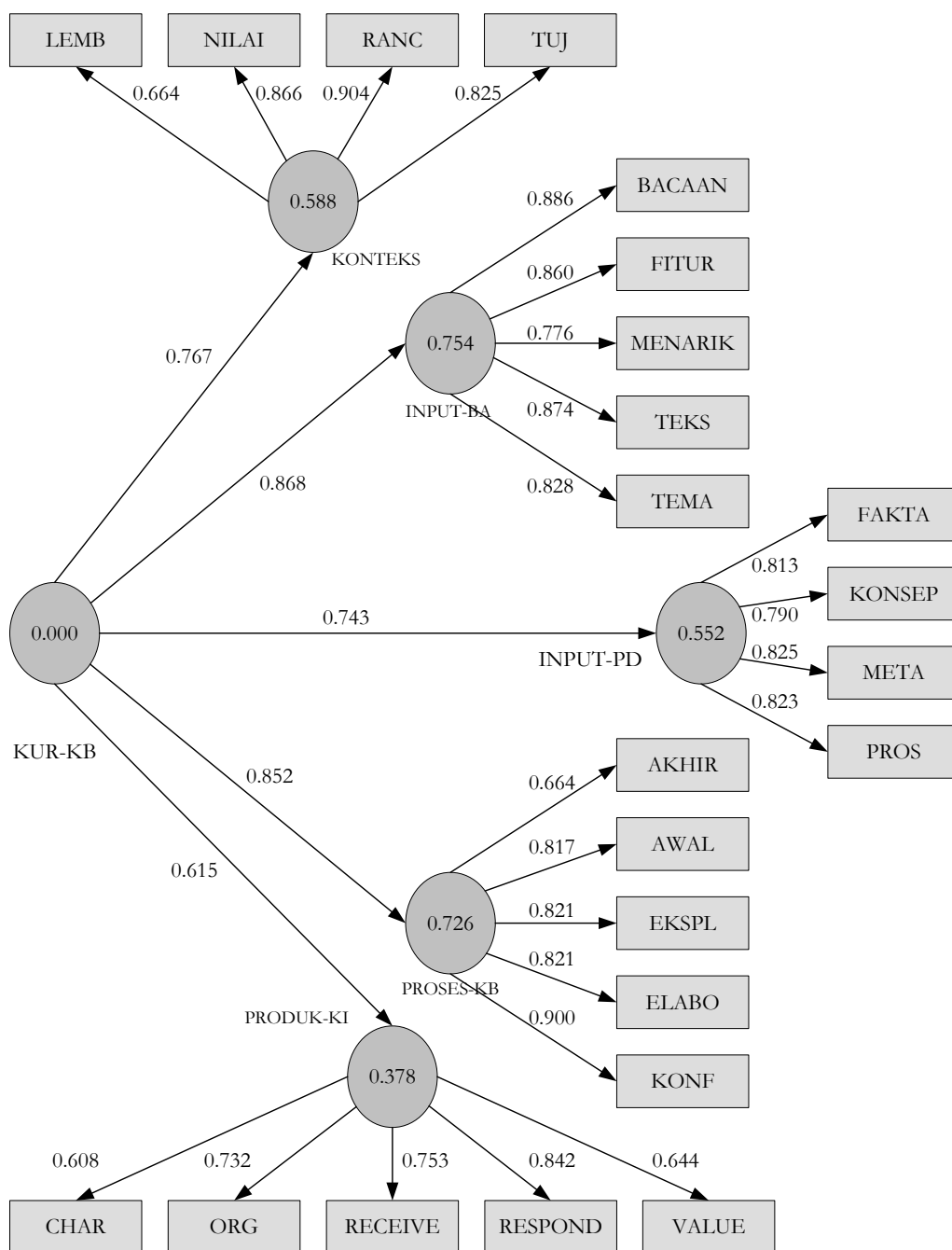
Hasil uji kelayakan model evaluasi reflektif kurikulum rumpun MKK pendidikan bahasa Inggris dalam pengembangan karakter bangsa disajikan pada Tabel 10.

Tabel 10. Uji Kecocokan Model Evaluasi Reflektif Kurikulum MKKPBIKB

Model Fit	
FIT	0,64
AFIT	0,64
GFI	0,99
SRMR	0,06
NPAR	97

Pada tabel di atas tampak bahwa nilai FIT sebesar 0,64 dan AFIT sebesar 0,64 menunjukkan bahwa model yang dispesifikasi mampu menjelaskan 64% varians pada data. Nilai GFI (*unweight least squares/Goodness of Fit*) sebesar 0,99 dan SRMR (*standardized root mean square residual*) sebesar 0,06 menunjukkan model fit yang baik karena nilai GFI mendekati 1 atau  $\geq 0,98$  dan SRMR mendekati 1 atau  $\leq 0,08$ . Dengan demikian dapat disimpulkan bahwa model teoritik yang dikembangkan didukung oleh data empirik.

Hasil analisis menunjukkan bahwa konstruk model EvalRefKurMKKPBIKB mengalami perubahan pada jumlah sub-variabel laten dari empat menjadi lima yaitu dengan 2 variabel input yaitu pengetahuan dosen dan bahan ajar seperti pada Gambar 4.



Gambar 1. Model Empirik Instrumen Evaluasi Reflektif Kurikulum MKKPBIKB

Konstruk kurikulum bermuatan karakter bangsa dibentuk oleh sub-konstruk konteks kurikulum MKK PBI bermuatan karakter bangsa, sub-konstruk input menghasilkan 2 dimensi yaitu bahan ajar bermuatan karakter bangsa dan pengetahuan dosen tentang pendidikan karakter bangsa, sub-konstruk muatan karakter bangsa dalam PBM dan sub-konstruk produk hasil pembelajaran afeksi (karakter bangsa mahasiswa).

Secara teoretik konstruk input hanya terdiri dari 1 dimensi, ternyata setelah diterapkan di lapangan menghasilkan 2 dimensi yaitu dimensi bahan ajar bermuatan karakter bangsa dan dimensi pengetahuan dosen tentang pendidikan karakter, sehingga konstruk kurikulum MKK PBI bermuatan karakter bangsa menjadi model C-I-I-P-P (*Context-Input-Input-Process-Product*).

*Kriteria Penyebaran (Kategorisasi) Model Evaluasi Reflektif Kurikulum Rumpun MKK PBI dalam Pengembangan Karakter Bangsa*

Kriteria penyeoran atau kategorisasi menggunakan kriteria pacuan norma (*norm reference*) dengan menguji signifikansi perbedaan antara *mean* skor/sampel dan *mean* skor teoretik/populasi dan mengkategorikan ke jenjang Rendah, Sedang, dan Tinggi (Azwar, 2012, pp.154-8). Penentuan skor dilakukan dengan menetapkan kategori Tengah/Sedang, dengan menghitung batas-bawah dan batas-atas internal skor-skor yang berbeda secara signifikan dari harga suatu *mean* populasi, menurut tingkat kepercayaan yang diinginkan dengan rumusan:

$$\mu - t_{(\alpha/2, n-1)} (S/\sqrt{n}) \leq X \leq \mu + t_{(\alpha/2, n-1)} (S/\sqrt{n})$$

- $\mu$  : *Mean* teoritis pada skala
- $t_{(\alpha/2, n-1)}$  : Harga t pada  $\alpha/2$  dan derajat kebebasan n-1
- S : Deviasi standar skor
- N : Banyaknya subyek

Hasil output *one-sample statistics* dari SPSS menunjukkan harga *Std. Error Means* tiap indikator dalam konstruk seperti pada indikator instrumen konsteks kurikulum MKK PBI bermuatan karakter bangsa disajikan pada Tabel 11.

Tabel 11. Harga *Std. Error Means* (SEM) instrumen konteks kurikulum MKKPBIKB

Indikator	N	Mean	SD	SER
TUJ	311	24,86	7,121	0,404
RANC	311	18,91	4,767	0,270
NILAI	311	12,24	3,285	0,186
LEMB	311	7,23	1,822	0,103

Tabel 12. Interval Nilai Tengah untuk Kategori Sedang Instrumen Evaluasi Konteks KurMKKPBIKB

Indikator	Kategori Sedang
Tujuan Pembelajaran	$23 \leq X \leq 25$
Rancangan Silabus	$17 \leq X \leq 19$
Penilaian Kelas	$11.6 \leq X \leq 12$
Kelembagaan	$6,7 \leq X \leq 7$

Dengan menggunakan rumusan internal dan harga SER pada Tabel 11 dan dengan melihat harga pada tabel Deviasi Normal (ukuran sampel melebihi 100) sebesar 1,960 diperoleh interval nilai tengah untuk kategori sedang pada instrumen konteks KurMKKPBIKB seperti disajikan pada Tabel 12.

Dengan demikian, diperoleh norma kategorisasi untuk tiap indikator instrumen konteks kurikulum MKKPBIKB disajikan pada Tabel 13.

Tabel 13. Norma Kategorisasi Instrumen Evaluasi Konteks KurMKKPBIKB

Indikator	Norma Kategorisasi	Kategori
Tujuan Pemb	$X < 23$	Rendah
	$23 \leq X \leq 25$	Sedang
	$25 < X$	Tinggi
Rancangan Silabus	$X < 17$	Rendah
	$17 \leq X \leq 19$	Sedang
	$19 < X$	Tinggi
Penilaian Kelas	$X < 11,6$	Rendah
	$11,6 \leq X \leq 12$	Sedang
	$12 < X$	Tinggi
Kelembagaan	$X < 6,7$	Rendah
	$6,7 \leq X \leq 7$	Sedang
	$7 < X$	Tinggi

Hasil evaluasi reflektif kurikulum rumpun MKK PBI secara keseluruhan dinilai menggunakan kriteria umum penilaian berdasarkan persentase untuk membedakan muatan KB pada kurikulum seperti berikut:

Tabel 14. Kriteria Umum Penilaian Muatan Karakter Bangsa pada KurMKKPBIKB

KRITERIA	Penilaian	KET
76 – 100 %	SK	Sangat Kuat Muatan KB
51 – 75 %	K	Kuat Muatan KB
26 – 50 %	L	Lemah Muatan KB
0 – 25 %	SL	Sangat Lemah Muatan KB

**Simpulan dan Saran**

Hasil penelitian dan pembahasan terhadap model evaluasi reflektif kurikulum rumpun MKK PBI dalam pengembangan karakter bangsa dapat disimpulkan: (1) model ini mencakup lima konstruk dan lima in-

strumen evaluasi reflektif kurikulum bermuatan karakter bangsa yang mencakup dimensi konteks kurikulum, dimensi pengetahuan, dimensi bahan ajar, dimensi proses pembelajaran, dan dimensi produk pembelajaran afeksi (karakter bangsa mahasiswa). Model pengukuran kurikulum bermuatan karakter bangsa yang disusun berdasarkan konstruk kurikulum bermuatan karakter bangsa fit (cocok) dengan model, yang berarti didukung data empirik dengan nilai GFI sebesar 0,99 sesuai kriteria mendekati 1 atau  $\geq 0,98$  dan nilai SRMR sebesar 0,06 sesuai kriteria mendekati 0 atau  $\leq 0,08$ ; (2) Semua instrumen evaluasi reflektif kurikulum MKK PBI dalam pengembangan karakter bangsa telah teruji validitas dan reliabilitasnya, yaitu memiliki muatan faktor  $> 0,50$  dan reliabilitas paling kecil 0,70 dengan rincian sebagai berikut: instrumen evaluasi muatan karakter bangsa dalam konteks kurikulum berjumlah 20 butir memiliki muatan faktor  $> 0,50$  dan reliabilitas sebesar  $0,92 \geq 0,70$ ; Instrumen evaluasi pengetahuan dosen tentang pendidikan karakter berjumlah 17 butir memiliki muatan faktor  $> 0,50$  dan reliabilitas sebesar  $0,90 \geq 0,70$ ; Instrumen evaluasi muatan karakter bangsa dalam bahan ajar berjumlah 17 butir memiliki muatan faktor  $> 0,50$  dan reliabilitas sebesar  $0,94 \geq 0,70$ ; Instrumen evaluasi muatan karakter bangsa dalam proses pembelajaran berjumlah 25 butir memiliki muatan faktor  $> 0,50$  dan reliabilitas sebesar  $0,95 \geq 0,70$ ; dan Instrumen evaluasi produk pembelajaran afeksi (karakter bangsa mahasiswa) berjumlah 21 butir memiliki muatan faktor  $> 0,50$  dan reliabilitas sebesar  $0,83 \geq 0,70$ ; (3) hasil pengembangan model evaluasi reflektif kurikulum MKK PBI dalam pengembangan karakter bangsa yang telah dibangun berdasarkan kajian pustaka, penilaian rekan sejawat, pakar, dan praktisi, dan uji coba empirik terdiri atas lima aspek/dimensi evaluasi dengan nama Konteks (*Context*), Pengetahuan Dosen (*Input*), Bahan Ajar (*Input*), Proses (*Process*), dan Produk (*Product*) yang dapat disingkat menjadi CIIPP (*Context-Input-Input-Process-Product*). Model evaluasi yang dikembangkan ini terdiri dari seperangkat instrumen evaluasi, skema model

evaluasi, panduan cara melakukan kegiatan evaluasi dan kriteria penyekorannya yang dapat digunakan untuk memfasilitasi proses refleksi diri dosen terhadap kurikulum MKK yang diselenggarakannya, karena memiliki model pengukuran yang fit (cocok) dengan model yang didukung data empirik dengan nilai GFI sebesar 0,99 sesuai kriteria mendekati 1 atau  $\geq 0,98$  dan nilai SRMR sebesar 0,06 sesuai kriteria mendekati 0 atau  $\leq 0,08$ .

Beberapa saran pemanfaatan produk evaluasi reflektif kurikulum rumpun MKK PBI adalah sebagai berikut: (1) model dapat dipergunakan oleh para dosen sebagai refleksi diri terhadap penguasaan muatan karakter bangsa dalam kurikulum MKK yang diselenggarakannya demi mem mempersiapkan calon guru bahasa Inggris yang tidak hanya kompeten dan profesional tetapi juga mempunyai kesadaran untuk membentuk karakter bangsa peserta didiknya kelak; (2) Ketua Program Studi dapat memanfaatkan hasil evaluasi sebagai dasar untuk merencanakan kegiatan internal yang dapat membantu mengingatkan dan memastikan penguasaan karakter bangsa dalam kurikulum pembelajaran yang diselenggarakan oleh tiap dosen secara rutin dan berkesinambungan seperti yang diamanatkan oleh UUD dan demi mewujudkan visi dan misi lembaga; (3) pimpinan perguruan tinggi dapat memanfaatkan hasil evaluasi untuk pengambilan keputusan terkait langkah-langkah yang harus dilakukan dalam memastikan penguasaan karakter bangsa dalam mempersiapkan lulusan yang memiliki selain kompetensi dan profesionalisme juga kesadaran dan karakter bangsa yang kuat dalam membangun bangsa dan Negara.

### Daftar Pustaka

- Apple, M.W. & Beyer, L.E. (1983). Social evaluation of curriculum. *American Educational Research Association*, 5 (4), 425-434.
- Azwar, S. (2012). *Penyusunan skala psikologis, Edisi 2*. Yogyakarta: PUSTAKA PELAJAR.

- Bath, D. Smith, C. Stein, S. & Swann, R. (2004). Beyond mapping and embedding graduate attributes: bringing together quality assurance and action learning to create a validated and living curriculum. *Higher Education Research & Development*, 23 (3), 313-328.
- Beane, J.A. Toepfer, C.F. Jr. & Alessi, S.j. (2004). *Curriculum planning and development* Boston: Allyn and Bacon, Inc.
- Berkowitz, M. W., & Bier, M.C. (2004). Research-based character education. *The ANNALS of the American Academy of Political and Social Science*, 591 (1), 72-85.
- Cennamo, K. & Kalk, D. (2005). *Real world instructional design*. Canada: Thomson Wadsworth.
- Davis, E. (1980). *Teachers as curriculum evaluators*. North Sydney: GEORGE ALLEN & UNWIN.
- Hamilton, D. (1977). Making sense of curriculum evaluation: continuities and discontinuities in an educational ideas. *Review of Research in Education*, 5, 318-347.
- Heungsun Hwang. (2011). *GeSCA user's manual*. Diakses tanggal 15 Juni 2013 dari [www.sem-gesca.org](http://www.sem-gesca.org)
- Jackson, P.W. (Ed.).(1992). *Handbook of research on curriculum: a project of the American Educational Research Association*. New York: MACMILLAN PUBLISHING COMPANY.
- Kementerian Pendidikan Nasional Badan Penelitian dan Pengembangan Pusat Kurikulum. (2010). *Bahan pelatihan penguatan metodologi pembelajaran berdasarkan nilai-nilai budaya untuk membentuk daya saing dan karakter bangsa: pengembangan pendidikan budaya dan karakter*. Diakses tanggal 27 April 2011 dari [http://www.puskur.net/files/1\\_%20Pendidikan%20Budaya%20dan%20Karakter%20Bangsa.pdf](http://www.puskur.net/files/1_%20Pendidikan%20Budaya%20dan%20Karakter%20Bangsa.pdf)
- Ketetapan MPR RI No. VII/MPR/1978.
- Ki Hadjar Dewantara. (2004). *Karya Ki Hadjar Dewantara bagian pertama: pendidikan*. Yogyakarta: Majelis Luhur Persatuan Taman Siswa.
- Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563-575.
- Leming, J.S. (2000). Tell me a story: an evaluation of literature-based character education program. *Journal of Moral Education*, 29 (4), 413-426.
- Marsh, C. J. (2009). *Key concepts for understanding curriculum*. London: Routledge.
- Martone, A. & Sireci, S.G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, 79, 1332-1361.
- Meadows, B. (2008). Co-constructing the 'Familiar Exotic' in second language learner discourse. *Texas Linguistic Forum*, 52, 117-130.
- Narvaez, D., & Lapsley, D.K. (2006). Teaching moral character: two strategies for teacher education. *In Press*,
- Nunan. (2004). *Task-based language teaching*. Cambridge: Cambridge Language Teaching Library
- Orstein, A.C. & Hunkins, F.P. (2009). *Curriculum: foundations, principles, and issues*. (5<sup>th</sup> ed.) Boston, MA: Pearson International Edition.
- Parkay, F.W., Hass, G.J., & Anctil, E.J. (2010). *Curriculum leadership: readings for developing quality educational programs*. PEARSON.
- Parmenter, L. (1999). Constructing national identity in a changing world: perspective in Japanese education. *British Journal of Sociology of Education*, 20 (4), 453-463.
- Pemerintah Republik Indonesia. (2010). *Kebijakan nasional pembangunan karakter bangsa tahun 2010-2025*. Diakses

- tanggal 11 Pebruari 2013 dari <http://www.google.com/#hl=en&tbo=d&output=search&scient=psy-ab&q=kebijakan+nasional+pembangunan+karakter+bangsa+tahun+2010-2025&gs>
- Richards, J. C. (2001). *Curriculum development in language teaching*. New York: Cambridge Language Education.
- Ringle, C.M. Wende, S. & Will, S. (2005). SmartPLS 2.0 (M3). Beta, Hamburg. Diakses tanggal 10 Juli 2013 dari <http://www.smartpls.de>
- Saifuddin, A.F., & Karim, M. (2008). *Refleksi karakter bangsa*. Jakarta: Penerbit Forum Kajian Antropologi Indonesia.
- Sari, M, & Doğanay, A. (2009). Hidden curriculum on gaining the value of respect for human dignity: a qualitative study in two elementary schools in Adana. *Educational Sciences: Theory & Practice*, 9 (2), 925-940.
- Sumarno. (2011). Peran pendidikan nonformal dan informal dalam pendidikan karakter bangsa. *Cakrawala Pendidikan Jurnal Ilmiah Pendidikan*, Mei, Th. XXX, Edisi Khusus Dies Natalis UNY, 73-84.
- Taylor-Powell, E., Jones, L., & Henert, E. (2003). *Enhancing program performance with logic models*. Diakses tanggal 1 Maret 2011 dari <http://www.uwex.edu/ces/Imcourse/>
- Winarno. (2010). *Implementasi Pancasila melalui pendidikan kewarganegaraan (civic education)*. Makalah disajikan dalam Seminar Pengalaman Indonesia dan Malaysia dalam hal pembinaan warga Negara yang cerdas dan baik, di Universitas Pendidikan Sultan Idris.
- Worthen, B.R., & Sanders, J.R. (1984). *Educational evaluation: theory and practice*. Worthington, OH: Charles A Jones Publishing Company.
- Undang-Undang RI Nomor 20, Tahun 2003, *tentang Sistem Pendidikan Nasional*.

## METODE STANDARD SETTING UNTUK UJIAN NASIONAL DI SEKOLAH DASAR

<sup>1)</sup>Sri Rejeki, <sup>2)</sup>Djemari Mardapi, <sup>3)</sup>Kumaidi

<sup>1)</sup>Universitas PGRI Yogyakarta, <sup>2)</sup>Universitas Negeri Yogyakarta,

<sup>3)</sup>Universitas Muhammadiyah Surakarta

<sup>1)</sup>sri.rjk10@yahoo.com, <sup>2)</sup>djemarimardapi@gmail.com, <sup>3)</sup>kuma\_426@yahoo.com

### Abstrak

Penelitian ini bertujuan mengetahui karakteristik hasil *Cut Score* UN di SD tahun 2009 berdasarkan pengembangan implementasi metode Nedelsky, Ebel, *Bookmark* dan hasil modifikasi metode Ebel-*Bookmark*; keyakinan panelis terhadap implementasi metode *Standard Setting* untuk menetapkan *Cut Score*; dan akurasi implementasi berdasarkan standar deviasi masing-masing metode. Sampel meliputi 10 orang panelis pada putaran 1, 16 guru pada putaran 2, 8 pakar dan 1650 hasil pekerjaan siswa. Prosedur dan analisis data mengikuti langkah empat metode yang ditetapkan. Analisis *item* menggunakan program ITEMAN dan Bilog MG dengan 1 Parameter Logistik. Berdasarkan hasil analisis data disimpulkan (1) Rerata *Cut Score* berdasarkan metode Nedelsky untuk mata pelajaran Bahasa Indonesia 28,44, Matematika 23,47, IPA 27. Metode Ebel untuk Bahasa Indonesia 80,75, Matematika 60,21, IPA 77,54. Metode *Bookmark* untuk Bahasa Indonesia sebesar 51,63, Matematika 51,33, IPA 59,16. Modifikasi Ebel-*Bookmark* untuk Bahasa Indonesia 81,27, Matematika 79,06 dan IPA 77,44; (2) keyakinan panelis dalam implementasi metode *Standard Setting* untuk metode *Bookmark* 81,25%, modifikasi Ebel-*Bookmark* 62,50%, metode Ebel 43,75% dan metode Nedelsky 43,75%; (3) Metode *Bookmark* lebih akurat dalam menetapkan *cut score* mata pelajaran Bahasa Indonesia dan Matematika, modifikasi metode Ebel-*Bookmark* lebih akurat untuk menetapkan *cut score* mata pelajaran IPA.

**Kata kunci:** *standard setting, ujian nasional, sekolah dasar*

## THE STANDARD SETTING METHOD FOR THE NATIONAL EXAMINATION IN THE ELEMENTARY SCHOOL

<sup>1)</sup>Sri Rejeki, <sup>2)</sup>Djemari Mardapi, <sup>3)</sup>Kumaidi

<sup>1)</sup>Universitas PGRI Yogyakarta, <sup>2)</sup>Universitas Negeri Yogyakarta,

<sup>3)</sup>Universitas Muhammadiyah Surakarta

<sup>1)</sup>sri.rjk10@yahoo.com, <sup>2)</sup>djemarimardapi@gmail.com, <sup>3)</sup>kuma\_426@yahoo.com

### Abstract

This study aims to investigate the characteristics results in Cut Score of UN elementary school in 2009 based on developing the implementation of Nedelsky, Ebel, Bookmarks methods, and modified Ebel method-Bookmark; panelists' confidence on the implementation of the methods for setting the Standard Setting Cut Score; and the implementation's accuracy based on the standard deviation of each method. The sample included 10 panelists in round 1, 16 teachers in round 2, 8 experts, and 1650 students' work. Procedures and analysis of data followed the steps of four methods specified. Item analysis used the program ITEMAN and Bilog MG with 1 Parameter Logistic. Based on the results of data analysis it is concluded 1). Cut Mean Score based on Nedelsky method for Indonesian is 28.44, Math is 23.47, and Science is 27. Ebel methods for Indonesian is 80.75, Mathematics is 60.21, and Science is 77.54. Bookmark method for Indonesian is 51.63, Mathematics is 51.33, Science is 59.16. Modified Ebel-Bookmark for Indonesian is 81.27, Mathematics is 79.06, and Science is 77.44; 2). panelists confidence in the implementation of the Standard Setting of Bookmark method is 81.25%, Ebel-modification Bookmark is 62.50%, Ebel methods is 43.75%, and methods Nedelsky is 43.75%; 3).The Bookmark method is more accurate in determining the cut score for Indonesian Language and Mathematics, modified Ebel-Bookmark method is more accurate to establish the cut score for the science subject.

**Keywords:** *Standard Setting, National Examination, elementary school*

## **Pendahuluan**

Undang-Undang Republik Indonesia Nomor 20 Tahun 2003, pada bab XVI pasal 57 sampai dengan 59 tentang Evaluasi menyatakan bahwa dalam rangka pengendalian mutu pendidikan secara nasional dilakukan evaluasi sebagai bentuk akuntabilitas penyelenggara pendidikan kepada pihak-pihak yang berkepentingan. Evaluasi dilakukan oleh lembaga yang mandiri secara berkala, menyeluruh, transparan, dan sistematis untuk menilai pencapaian standar nasional pendidikan dan proses pemantauan evaluasi tersebut harus dilakukan secara berkesinambungan, yang akhirnya akan dapat membenahi mutu pendidikan.

Evaluasi adalah bagian utama dalam proses belajar mengajar, evaluasi di sekolah tidak hanya mengevaluasi siswa tetapi juga kinerja guru dan sekolah, untuk menunjukkan efektivitas sekolah. Ini berarti asesmen dan evaluasi pada gilirannya merupakan bagian dari proses akuntabilitas. Secara tidak langsung perencanaan kegiatan asesmen dan evaluasi memerlukan aturan secara nasional untuk memenuhi tujuan pendidikan nasional dan lokal serta kebutuhan sekolah. Asesmen dan evaluasi diharapkan dapat memberikan informasi dan pertimbangan kelanjutan proses belajar mengajar baik untuk sekolah, daerah maupun pemerintah pusat.

Mengembangkan pembelajaran adalah tujuan utama sekolah. Asesmen merupakan kerangka kerja yang dapat menghasilkan dasar bagi perencanaan langkah-langkah pendidikan selanjutnya dalam merespon kebutuhan siswa. Penilaian pendidikan sebagai proses pengumpulan dan pengolahan informasi hasil evaluasi dalam rangka menentukan pencapaian hasil belajar siswa dan keberhasilan proses pembelajaran. Dalam penilaian pendidikan ada beberapa yang harus diperhatikan baik prinsip penilaian, standar penilaian, mekanisme, prosedur dan instrumen penilaian hasil peserta didik. Prinsip penilaian hasil peserta didik pada jenjang pendidikan dasar dan menengah berdasarkan penjelasan Peraturan

Menteri Pendidikan Nasional No. 20 Tahun 2007(2007, pp.7-8) adalah sah, objektif, adil, terpadu, terbuka, menyeluruh dan berkesinambungan, sistematis, beracuan kriteria dan akuntabel. Mekanisme dan prosedur penilaian hasil belajar dilaksanakan oleh pendidik, satuan pendidikan, dan pemerintah.

Ujian Nasional (UN)/Ujian Akhir Sekolah Berstandar Nasional (UASBN) di Indonesia (dahulu dengan istilah EBTANAS, UNAS, UAN) pada sejarah perkembangan dan implementasinya terdapat beberapa permasalahan. Permasalahan yang muncul adalah masalah administrasi, persepsi *stakeholders*, penetapan batas kelulusan, dan masalah geografis, pemahaman terhadap Kurikulum Tingkat Satuan Pendidikan dan adanya otonomi daerah yang berdampak pada otonomi pendidikan di Indonesia.

Administrasi Ujian Nasional selama ini memang berliku mulai penyusunan kisi-kisi UN, penyusunan soal baik tingkat nasional dan daerah, uji coba, penetapan soal untuk standar nasional dan daerah, cetak perangkat soal dan LJK, pengiriman, distribusi dan pendataan sesampainya di sekolah dan pengiriman kembali hasil pekerjaan peserta ujian. Permasalahan yang sering muncul dari administrasi UN adalah pada kerawanan cetak soal dan LJK, pengiriman dan distribusi soal. Permasalahan lain dari UN adalah persepsi *stakeholders* terhadap pelaksanaan dan hasil UN, seperti adanya pro-kontra Ujian Akhir Nasional dalam masyarakat, sosialisasi tentang batas kelulusan belum merata, masyarakat dalam memahami batas kelulusan belum seperti yang diharapkan dalam kebijakan pendidikan.

Penetapan batas kelulusan hasil UN di Indonesia banyak menuai kritikan karena batas kelulusan di Indonesia seperti di tingkat SMP dan SMA batas kelulusan masih sama, batas kelulusan di tingkat SD ditentukan oleh sekolah, belum dikembangkannya metode penetapan standar berdasarkan prosedur atau metode penetapan *standard setting* secara tunggal (nasional) ataupun tingkat daerah dan adanya bias

dalam penentuan batas kelulusan yang belum berbasis pada SKL (standar kompetensi lulusan) dan KKM (kriteria ketuntasan minimal) dan kurikulum yang berlaku. Di sisi lain, dalam penetapan standar kelulusan di Indonesia belum memanfaatkan metode *standard setting* yang berkembang dalam sistem pengujian dan penilaian.

Geografis Indonesia yang terdiri dari kepulauan juga menjadi permasalahan yang tak dapat dihindarkan dalam UN. Hal ini berkaitan erat dengan pengiriman dan distribusi soal, pemerataan fasilitas pendidikan, kualitas pendidikan yang belum merata dan jarak tempuh peserta ujian dengan sekolah.

Penilaian yang dilakukan seseorang atau suatu lembaga biasanya memiliki tujuan. Tujuan tersebut antara lain (a) menetapkan atau membuat keputusan seseorang peserta ujian lulus atau gagal, (b). menempatkan para peserta ujian pada level tertentu misalnya termasuk kategori dasar, menengah atau lanjut dalam suatu pencapaian kemampuan ketrampilan tertentu. Ranah/isi dan ketrampilan yang telah ditetapkan dalam kurikulum dan kategori kinerja untuk penafsiran skor ujian adalah konsep yang mendasar dalam sistem penilaian pendidikan yang bertujuan untuk menjelaskan apa yang diketahui dan bisa dilakukan oleh para siswa. Informasi hasil ujian juga bermanfaat untuk *stakeholders*.

Penetapan kriteria penilaian memerlukan standar tertentu yang disebut dengan *standard setting (setting performance standard)*. Penentuan standar pendidikan adalah penentuan nilai batas (*cut score*). Seseorang dikatakan sudah lulus/kompeten bila telah melewati nilai batas tersebut berupa nilai batas antara peserta didik yang sudah menguasai kompetensi tertentu dengan peserta didik yang belum menguasai kompetensi tertentu. Bila hal itu terjadi pada ujian nasional atau sekolah maka nilai batas berfungsi untuk memisahkan antara peserta didik yang lulus dan tidak lulus disebut batas kelulusan. Kegiatan penentuan batas kelulusan disebut *standard setting*.

Kata “standar” dapat digunakan dalam kaitannya dengan (1) muatan dan keterampilan yang dipandang perlu untuk dikuasai peserta ujian, dan (2) skor yang perlu mereka dapatkan untuk menunjukkan pengetahuan dan keterampilan yang relevan. Dalam konteks penilaian yang digunakan di lingkungan pendidikan, perbedaan harus dilakukan antara standar isi dan standar prestasi, karena ketidakjelasan mengenai perbedaan antara kedua konsep ini sering kali muncul di kalangan pengambil kebijakan, pendidik, dan masyarakat. Standar isi tercerminkan dalam kurikulum dan menetapkan apa yang semestinya diketahui dan bisa dikerjakan oleh peserta ujian. Standar isi memberikan arahan bagi pengajar mengenai apa yang perlu diajarkan. Standar prestasi dengan demikian dapat dipandang sebagai operasionalisasi standar isi bagi tes atau penilaian yang telah disusun untuk mengukur standar isi.

Secara umum metode *standard setting* diklasifikasikan dalam 3 kelompok (Cizek & Bunch, 2007, p.9) yakni *examine centered versus test centered; holistic models versus analytic models; norm referenced versus criterion referenced*. Menurut Jaeger (Cizek & Bunch, 2007, p.9) metode penentuan standar bisa diklasifikasikan menjadi dua kategori yaitu metode yang terpusat pada peserta ujian dan metode yang terpusat pada ujian. Kane (Cizek & Bunch, 2007, p.9) telah mengemukakan cara untuk mengklasifikasikan metode penentuan standar. Ia mengemukakan bahwa metode bisa diklasifikasikan sebagai model holistik yang berasumsi bahwa prestasi dan keterampilan sangat terintegrasi atau model analitis yang berasumsi bahwa prestasi bisa diukur menggunakan sampel kinerja. Ada tiga hal yang berpengaruh dalam implementasi metode *standard setting* dalam menentukan *cut score*: (1) terminologi mengenai standar, (2) peran panelis dalam menentukan standar prestasi, dan (3) perbedaan dalam metode penetapan standar antarjenis butir (Hambleton, 2006, p.436).

## Metode Penelitian

Populasi dalam penelitian ini adalah seluruh hasil UASBN siswa SD kelas VI tahun 2009 dan guru kelas VI SD di DIY. Sampel yang diambil meliputi guru SD sebagai panelis sejumlah 10 guru pada putaran 1, 16 guru pada putaran 2, 8 pakar dan 1650 hasil pekerjaan siswa. Penelitian ini dilakukan mengikuti prosedur dan analisis data dalam metode Nedelsky, Ebel, *Bookmark* dan modifikasi Ebel-*Bookmark*. Analisis *item* menggunakan program ITEMAN dan Bilog MG dengan 1 Parameter Logistik (1PL).

## Hasil dan Pembahasan

### Prosedur

#### *Prosedur Metode Nedelsky*

Pembahasan mengenai kelebihan dan kekurangan metode Nedelsky dalam implementasinya menunjukkan bahwa kelebihan dari metode ini pada saat digunakan, di satu sisi guru mengatakan metode ini lebih teliti karena per *item* ditelaah kunci jawaban yang benar dan kemungkinan jawaban yang salah, sehingga *cut score* yang dihasilkan akan lebih baik, tetapi guru mengalami kesulitan (dari hasil angket, guru/panelis ada yang mengatakan paling sulit) pada saat menetapkan soal yang dianggap benar dan diminta menetapkan distraktor yang ada, waktu yang digunakan untuk telaah soal dalam metode Nedelsky relatif lama. Permasalahan yang lain adalah guru merasa ditest saat mereka menelaah per *item* dan tidak diberi kunci jawaban untuk setiap mata pelajaran, sehingga metode ini di rasa sulit oleh panelis. Implementasi metode Nedelsky memperkenalkan kepada guru bagaimana kriteria batas kelulusan dapat ditetapkan berdasarkan telaah *item* soal ujian.

#### *Prosedur Metode Ebel*

Metode Ebel sebenarnya sudah merupakan perbaikan dari metode Nedelsky dan Angoff Ya/Tidak, yaitu panelisnya tidak hanya guru tapi juga ahli, namun panelis mengalami kesulitan saat harus menetapkan bobot kategori, karena harus memberikan

perkiraan berapa persen siswa dimungkinkan akan mampu mengerjakan soal tersebut dalam kategori relevansi dan tingkat kesulitan yang bervariasi. Panelis/guru merasa bahwa memberikan perkiraan dengan persentase akan menimbulkan bias, jika perkiraan tersebut terlalu tinggi atau terlalu rendah, sehingga akan berpengaruh pada *cut score*.

#### *Prosedur Metode Bookmark*

Prosedur *Bookmark* dirasa paling mudah bagi guru/panelis dalam menetapkan *cut score*, karena data sudah disiapkan terlebih dahulu, soal sudah tertata dari yang paling mudah ke paling sulit, panelis merasa lebih mudah menentukan dan memasukkan sampai halaman *Bookmark* ke berapa siswa dianggap mampu menjawab pertanyaan dengan benar sebagai penetapan *cut score*. Kelemahan implementasi prosedur *Bookmark* menurut para panelis adalah komposisi soal dari yang termudah ke yang paling sulit memungkinkan ada beberapa indikator dari kemampuan yang diujikan tidak termasuk dalam halaman *Bookmark* yang dipilih oleh panelis sebagai batas kemampuan siswa dalam menjawab soal.

#### *Prosedur Modifikasi Metode Ebel-Bookmark*

Kelebihan hasil modifikasi metode Ebel-*Bookmark* diartikan bahwa modifikasi metode ini memberikan kemudahan kepada guru dan ahli dalam menetapkan batas kelulusan untuk suatu kompetensi yang diharapkan, dan dapat menentukan level kinerja peserta ujian, dengan tiga cara. Pertama melihat pada relevansi dan tingkat kesulitan soal yang sudah terurut; kedua batas kelulusan berdasarkan pada tingkat kesulitan butir dan tingkat kemampuan siswa; ketiga penetapan batas kelulusan melibatkan ahli dan guru secara bersama.

Kelemahan dari metode ini adalah dalam mendeteksi kemampuan *examine* hanya dilihat dari jumlah butir yang mampu dijawab benar oleh *examine* dan belum menjelaskan kemampuan di standar kompetensi yang mana secara detail. Selama ini metode *standard setting* dan level kinerja yang ber-

kembang juga belum memberikan informasi yang demikian. Berdasarkan implementasi di lapangan dari prosedur modifikasi Ebel-Bookmark diperoleh informasi soal yang terurut dalam booklet (*OIB/Odered Item Booklet*) dari soal yang paling mudah ke soal yang paling sulit ditelaah oleh guru ada beberapa soal yang masuk kategori tingkat kesulitan “mudah” tetapi ternyata masuk kategori “sulit”, kategori “sedang” tetapi oleh panelis termasuk kategori “sulit”.

Batas Kelulusan

Metode penetapan batas kelulusan (*cut score*) (*cut score*) diklasifikasikan secara umum dalam 3 kelompok (Cizek & Bunch, 2007, p.9) yakni *examine centered versus test centered; holistic models versus analytic models; norm referenced versus criterion referenced*.

Mata Pelajaran Bahasa Indonesia

Berdasarkan implementasi 4 metode *standard setting* yang berdasar *test centered, cut score* untuk mata pelajaran Bahasa Indonesia disajikan pada Tabel 1.

Tabel 1. Perbandingan *Cut Score* Mata pelajaran Bahasa Indonesia ditinjau dari 4 Metode *Standard setting*

	Nedelsky	Ebel	Bookmark	Ebel-Bookmark
Putaran 1	30	80,227	47,553	
Putaran 2	26,875	81,226	55,714	82,135
Rerata	28,4375	80,7465	51,6335	

Mata Pelajaran Ilmu Pengetahuan Alam

Berdasarkan implementasi 4 metode *standard setting* yang berdasar *examine centered dan test centered* (Metode *Bookmark*) dan berdasarkan *test centered* (Metode Nedelsky, dan Ebel), *cut score* untuk mata pelajaran IPA dijelaskan pada Tabel 2.

Tabel 2. Perbandingan *Cut Score* Mata Pelajaran IPA ditinjau dari implementasi 4 Metode *Standard setting*

	Nedelsky	Ebel	Bookmark	Ebel-Bookmark
Putaran 1	29	70,854	54,42	
Putaran 2	25	74,54	63,84	77,4359
Rerata	27	77,54	59,155	

Perbandingan *cut score* antarmata pelajaran

Berdasarkan hasil analisis implementasi 5 metode *standard setting* yang dikembangkan diperoleh beberapa data yang dapat dijelaskan dalam beberapa aspek. Salah satunya adalah dengan membandingkan *cut score* 2 mata pelajaran yang diujikan (Mata Pelajaran Bahasa Indonesia dan Ilmu Pengetahuan Alam).

Tabel 3. Perbandingan *cut score* mata pelajaran USBN 2009

Putaran	Metode <i>Standard setting</i>	Mata Pelajaran	
		Bahasa Indonesia	IPA
Putaran 1	Nedelsky	30	29
	Ebel	80,227	70,854
	<i>Bookmark</i>	47,553	54,42
Putaran 2	Nedelsky	26,4375	25
	Ebel	81,266	74,54
	<i>Bookmark</i>	55,714	63,84
Rerata	Ebel- <i>Bookmark</i>	82,135	77,4359
	Nedelsky	28,4378	27
	Ebel	80,7465	77,54
	<i>Bookmark</i>	51,6335	59,155

Kemampuan lulusan

Berdasarkan hasil penelitian, kemampuan lulusan dapat dilihat dari beberapa aspek. Aspek itu antara lain kemampuan lulusan dilihat dari deskriptor implementasi tiap-tiap metode setiap putaran atau reratanya, perbedaan nilai yang diperoleh dalam

setiap mata pelajaran, persentase menjawab benar setiap mata pelajaran, jumlah kelulusan dari setiap mata pelajaran berdasarkan sampel penelitian, pendapat panelis dalam telaah soal.

*Kemampuan lulusan ditinjau dari deskriptor.*

Berdasarkan implementasi 4 metode *standard setting* dan modifikasi metode *Ebel-Bookmark* kemampuan lulusan dapat dilihat dari deskriptor masing-masing mata pelajaran. Deskriptor berdasarkan implementasi metode *standard setting* dapat dijelaskan sebagai berikut.

1) Metode Nedelsky

Dalam menetapkan deskriptor dilakukan analisis panelis terhadap soal UASBN berkaitan dengan SKL dan indikatornya. Deskriptor diambil berdasarkan jumlah panelis yang menyatakan bahwa suatu soal diperkirakan mampu dijawab benar oleh siswa minimal lebih dari separoh ( $1/2$ ) dari jumlah panelis ( $p > 5$ ) untuk putaran 1 dan ( $p > 8$ ) untuk putaran 2. Berikut ini Tabel deskriptor untuk putaran 1 dan 2 dengan metode Nedelsky.

a) Mata Pelajaran Bahasa Indonesia

Berdasarkan Tabel analisis panelis pada putaran pertama dan kedua kemampuan lulusan dalam ujian bahasa Indonesia berdasar SKL meliputi membaca dan menulis. Kemampuan lulusan dalam membaca meliputi menentukan isi bacaan (kecuali indikator menjawab pertanyaan mengapa), menentukan unsur intrinsik dongeng yang dapat dikerjakan siswa berdasar analisis hanya amanat dalam cerita dan membedakan watak tokoh dalam dongeng, menentukan isi laporan, menentukan isi tersurat dan tersirat dari rubrik yang dibaca, menentukan unsur intrinsik drama anak-anak, menentukan unsur intrinsik dan makna kalimat cerita anak-anak dan menentukan denah. Kemampuan lulusan di dalam menulis adalah kemampuan menulis di-

alog, mengisi formulir, menyusun kata majemuk, menyusun paragraf, menulis iklan menulis petunjuk pemakaian, menggunakan sinonim dan antonim, mendeskripsikan Gambar dengan kalimat yang tepat, dan mengurutkan Gambar seri.

b) Mata pelajaran IPA

Berdasarkan hasil diskusi antar-panelis diperoleh hampir semua indikator dalam SKL mata pelajaran IPA mampu dijawab oleh siswa, indikator kompetensi yang dimiliki oleh siswa berdasar Standar Kompetensi Lulusan untuk mata pelajaran IPA yang **tidak dapat dijawab** oleh peserta ujian menurut panelis adalah menjelaskan akibat dari penggunaan bahan tertentu, menjelaskan fungsi organ pada tumbuhan, mengidentifikasi bentuk energi, menjelaskan cara penghematan energi, menjelaskan dampak pemanasan global, terjadinya gerhana bulan/matahari.

2) Metode Ebel

Kemampuan lulusan berdasarkan hasil analisis panelis dalam Metode Ebel terdapat pada tingkat relevansi dan tingkat kesulitan soal yang dimungkinkan mampu dijawab benar oleh peserta ujian. Rerata *cut score* mata pelajaran Matematika sebesar 60,21 (menjawab benar 23,48) artinya siswa dinyatakan lulus jika nilai minimal yang dicapai dalam UASBN dalam mata pelajaran Matematika sebesar 6.02. Apabila diimplementasikan pada peserta ujian yang menjadi sampel penelitian (1650 peserta) maka yang dinyatakan lulus sebanyak 67,33% (1111 peserta) dan yang tidak lulus berdasarkan rerata *cut score* Matematika sebanyak 539 siswa (32,67 %). Rerata *cut score* mata pelajaran Bahasa Indonesia sebesar 80,75% (menjawab benar 37,14) artinya siswa dinyatakan lulus jika nilai minimal yang dicapai dalam UASBN dalam mata pelajaran Bahasa Indonesia sebesar 8,127. Apabila diimplementasikan pada peserta ujian yang menjadi sampel penelitian (1650

peserta) maka yang dinyatakan lulus sebanyak 60,36% (996 peserta) dan yang tidak lulus berdasarkan rerata *cut score* Bahasa Indonesia sebanyak 654 siswa (39,64 %).

Rerata *cut score* mata pelajaran IPA sebesar 72,695% (menjawab benar 28 dari 39 soal) artinya siswa dinyatakan lulus jika nilai minimal yang dicapai dalam UASBN dalam mata pelajaran IPA sebesar 7,18. Apabila diimplementasikan pada peserta ujian yang menjadi sampel penelitian (1650 peserta) maka yang dinyatakan lulus sebanyak 61,58% (1016 peserta) dan yang tidak lulus berdasarkan rerata *cut score* IPA sebanyak 1016 siswa (38,42 %).

### 3) Metode Bookmark

Berdasarkan hasil analisis panelis dalam metode *Bookmark* dapat diketahui kemampuan lulusan yang diasumsikan berdasarkan rerata *Bookmark* yang ditetapkan oleh masing-masing panelis dimana peserta dianggap mampu menyelesaikan soal. Tabel-Tabel di bawah ini menunjukkan kemampuan lulusan untuk implementasi metode *Bookmark*.

### 4) Modifikasi Ebel-Bookmark

Kemampuan lulusan dalam analisis modifikasi Ebel *Bookmark* dapat dilihat dari asumsi panelis tentang tingkat kesulitan yang dimungkinkan dialami oleh siswa dalam menjawab soal ujian, meskipun dalam booklet sudah terurut menurut tingkat kesulitan. Berdasarkan modifikasi Ebel-*Bookmark* maka untuk mata pelajaran Bahasa Indonesia peserta ujian diasumsikan mampu menjawab *item* soal sebesar 81,266%, Matematika 79.06% dan IPA 77,44 %.

*Kemampuan lulusan ditinjau dari perbedaan nilai UASBN tiap mata pelajaran.*

Berdasarkan implementasi 4 metode *standard setting* dan *cut score* dihasilkan, batas minimal nilai kelulusan untuk setiap mata pelajaran. Nilai minimal yang dapat digunakan untuk menetapkan kelulusan *examine* dalam mata pelajaran bahasa Indonesia ber-

dasarkan hasil rerata dalam metode Nedelsky sebesar 6,183, metode Ebel 8,075, Metode *Bookmark* 5,164 dan modifikasi Ebel-*Bookmark* sebesar 8,127. Mata pelajaran Matematika metode Nedelsky sebesar 6,02, metode Ebel 6,021, metode *Bookmark* 5,133 dan Modifikasi Ebel *Bookmark* 7,906. Mata Pelajaran IPA untuk metode Nedelsky 6,923, metode Ebel 7,754, metode *Bookmark* 5,916 dan modifikasi Ebel-*Bookmark* 7,744. Dengan demikian, dapat ditemukan bahwa metode *Bookmark* memberikan peluang peserta ujian yang lulus akan lebih banyak, karena nilai minimal kelulusan lebih rendah dari metode yang lain untuk semua mata pelajaran yang di ujikan.

*Tingkat Keyakinan/Kepercayaan Guru terhadap metode standard setting yang diimplementasikan*

Berdasarkan hasil angket kepada para panelis dan analisis berdasarkan standar deviasi masing-masing setelah selesai pelaksanaan implementasi metode *standard setting* diperoleh hasil sebagai berikut.

- 1) Keyakinan/kepercayaan panelis dalam implementasi *standard setting* metode *Bookmark* (81,25%), metode Angoff Ya/Tidak (68,75%), modifikasi Ebel-*Bookmark* (62,50%), metode Ebel (43,75%) dan metode Nedelsky (43,75%).
- 2) Penetapan batas kelulusan (*cut score*) akan lebih baik dilaksanakan dengan metode *standard setting* karena dengan anggapan bahwa dengan metode *standard setting* akan menemukan *cut score* yang baik dan kredibel, meskipun pada akhirnya akan ditetapkan dengan kebijakan namun hendaknya berdasarkan pada hasil *cut score* dengan metode *standard setting*;
- 3) Para panelis percaya bahwa dengan metode *standard setting* penetapan standard kelulusan tidak akan mengurangi pro kontra masalah UN;
- 4) Berdasarkan putaran pertama dan kedua serta uji coba terbatas lima metode *standard setting* para panelis menyatakan bahwa metode Angoff Ya/Tidak paling mudah di implementasikan, namun peluang kelulusan peserta ujian rendah, metode *Bookmark* dikatakan mudah diimplementasikan

dan memberikan peluang paling besar kepada peserta ujian untuk lulus UN, modifikasi Ebel-Bookmark mudah dilaksanakan namun membutuhkan waktu dalam menganalisis kembali Booklet yang sudah terurut dalam memasukkan ke dalam kriteria relevansi dan tingkat kesulitan maupun dalam memberikan skor harapan, metode Nedelsky dan Ebel dirasa sulit oleh panelis karena harus menganalisis satu-pesatu *item* sehingga mereka para panelis merasa seperti diuji;

- 5) Perlu adanya pelatihan tentang metode *standard setting* kepada guru-guru khususnya guru SD, karena batas kelulusan (*cut score*) di SD ditetapkan oleh sekolah.

### Rekomendasi Hasil

Maksud rekomendasi dalam penelitian ini adalah memberikan masukan dan sebagai bahan pertimbangan berdasarkan hasil penelitian tentang *standard setting*. Dalam berbagai hasil penelitian tentang *standard setting* yang menggunakan beberapa metode *standard setting* akan berusaha untuk memberikan rekomendasi, metode *standard setting* yang mudah dan dapat dimanfaatkan untuk menetapkan *cut score* dengan akurat serta diimplementasikan oleh para guru ataupun pemangku kepentingan.

Dalam penelitian ini ada lima metode *standard setting* yang digunakan untuk menetapkan *cut score*. Dasar penetapan rekomendasi adalah besarnya standar deviasi dari masing-masing metode *standard setting*, dengan asumsi semakin kecil standar deviasi maka metode *standard setting* tersebut semakin kecil tingkat kesalahan dalam menetapkan *cut score*.

Penetapan standar deviasi dengan mencari rerata standar deviasi antarpanelis pada putaran kedua. Dipilihnya putaran kedua karena panelis jumlahnya lebih banyak, 68,75% panelis telah mengikuti putaran pertama. Hasil Analisis standar deviasi diperoleh bahwa baik untuk mata pelajaran Bahasa Indonesia, IPA dan Matematika metode *Bookmark* lebih akurat dalam menentukan *cut score* selengkapnya disajikan pada Tabel 4.

Tabel 4. Rangkuman Standard Deviasi 4 Metode *Standard Setting*

	Nedelsky	Ebel	<i>Bookmark</i>	Modifikasi Ebel- <i>Bookmark</i>
Bahasa Indonesia	12,45	6,097	1,774	6,779
Matematika	9,136	5,471	1,736	4,725
IPA	18,97	5,283	4,288	4,247

Berdasarkan pendapat para panelis dan hasil perhitungan standar deviasi antarpanelis di setiap metode yang diimplementasikan dan dikembangkan dan *cut score* yang diperoleh menunjukkan bahwa metode *Bookmark* merupakan metode yang mudah dilaksanakan dan memberikan peluang lulus yang paling besar untuk mata pelajaran Bahasa Indonesia dan Matematika diantara metode yang lain. Khusus untuk mata pelajaran IPA dengan modifikasi metode Ebel-*Bookmark*.

Adanya perbedaan dalam mengidentifikasi metode yang dapat diimplementasikan dalam penetapan *cut score* yaitu panelis mengidentifikasi dengan mengurutkan metode mana yang mudah diimplementasikan tanpa melihat mata pelajaran yang diujikan, sedangkan dalam penetapan standar deviasi didasarkan pada hasil rerata standar deviasi antarpanelis disetiap implementasi metode *standard setting* dan berdasarkan identifikasi untuk setiap mata pelajaran .

### Simpulan

Berdasarkan hasil analisis data dapat disimpulkan bahwa: (1) rerata *Cut Score* berdasarkan metode Nedelsky untuk mata pelajaran Bahasa Indonesia sebesar 28,44, Matematika 23,47, IPA 27. Rerata *Cut Score* metode Ebel untuk mata pelajaran Bahasa Indonesia sebesar 80,75, Matematika 60,21, IPA 77,54. Rerata *Cut Score* metode *Bookmark* untuk mata pelajaran Bahasa Indonesia sebesar 51,63, Matematika 51,33, IPA 59,16. *Cut Score* modifikasi Ebel-*Bookmark* untuk mata pelajaran Bahasa Indonesia sebesar 81,27, Matematika 79,06 dan IPA 77,44; (2) keyakinan/kepercayaan panelis dalam implementasi metode *Standard Setting*

untuk metode *Bookmark* sebesar 81,25%, modifikasi Ebel-*Bookmark* 62,50%, metode Ebel 43,75% dan metode Nedelsky 43,75%; (3) metode *Bookmark* lebih akurat dibandingkan dengan metode yang lain untuk menetapkan *cut score* mata pelajaran Bahasa Indonesia dan Matematika dalam UN SD 2009, modifikasi metode Ebel-*Bookmark* lebih akurat dibandingkan dengan metode lain untuk menetapkan *cut score* mata pelajaran IPA dalam UN SD 2009.

#### Keterbatasan Penelitian

Penelitian ini dilaksanakan terbatas di wi-layah Propinsi Daerah Istimewa Yogyakarta dengan jumlah panelis yang terbatas mewakili 5 daerah kabupaten dan kota di DIY. Pendekatan penelitian dengan model pe-ngembangan ini terbatas pada empat implementasi metode yang sudah ada dan satu modifikasi metode baik dalam rubrik maupun analisisnya. Kemampuan peserta ujian belum dapat terdeteksi secara lengkap, baru berdasarkan pada analisis *item* dan deskriptor.

#### Saran

Pendidikan dasar meliputi SD dan SMP, di tingkat sekolah dasar apabila semua siswa yang menempuh UN akan dinyatakan lulus maka untuk melanjutkan ke jenjang SMP maka diperlukan kategori kinerja/level

kinerja tertentu seperti *basic*, *advance* dan *provicient* sehingga dapat menempatkan siswa pada sekolah lanjutan pertama dengan lebih tepat sesuai kompetensi yang dimilikinya.

#### Daftar Pustaka

- Angoff Ya/Tidak,W.H. (1971). *Scales, norms, and equivalent scores*. Thorndike (Ed), *Educational Meauserment (2<sup>nd</sup> ed)*. Washington, DC: American Council on Education
- Badan Standar Nasional Pendidikan.(2007). *Peraturan Menteri Pendidikan Nasional Nomor 20 Tabun 2007 tentang Standar Penilaian Pendidikan*. Jakarta: BSNP
- Cizek, J Gregory, Bunch Michael B. (2007).*Standard Setting: A Guide to Estlabishing and Evaluating Performance standards on Tests*. London: Sage Publications.
- Hambleton, Ronald K & Pitoniak, Mary J. (2006). Setting Performance Standards dalam Brennan, L. Robert. *Educational Measurement (Fourth Edition)*, American Council on Education .p. 433-470.
- Depdiknas. (2003). *Peraturan Pemerintah No. 20 Tabun 2003, tentang Undang-Undang Sistem Pendidikan Nasional*.

## PENGEMBANGAN INSTRUMEN EVALUASI UJI KOMPETENSI KEAHLIAN (UKK) ADMINISTRASI PERKANTORAN DI SMK

<sup>1)</sup>Suranto, <sup>2)</sup>Muhyadi, <sup>3)</sup>Djemari Mardapi

<sup>1,2)</sup>Fakultas Ekonomi Universitas Negeri Yogyakarta,

<sup>3)</sup>Fakultas Teknik Universitas Negeri Yogyakarta

<sup>1)</sup>suranto@uny.ac.id, <sup>2)</sup>muhyadi@uny.ac.id, <sup>3)</sup>djemarimardapi@gmail.com

### Abstrak

Penelitian ini bertujuan mengembangkan instrumen yang cocok untuk melakukan evaluasi penyelenggaraan kegiatan UKK administrasi perkantoran di Sekolah Menengah Kejuruan (SMK). Penelitian pengembangan ini mencakup empat tahapan utama, yaitu tahap pendahuluan, tahap perencanaan dan pengembangan, tahap uji coba evaluasi dan revisi, serta tahap implementasi. Teknik pengumpulan data penelitian ini menggunakan teknik *focus group discussion*, angket/kuesioner, wawancara, observasi, dan studi dokumen. Analisis data dengan responden siswa menggunakan program *Lisrel 8.51* dan data dengan responden guru dianalisis menggunakan *SPSS 17.00 for Windows*. Hasil penelitian ini sebagai berikut: (1) komponen penyelenggaraan UKK mencakup: (a) kolaborasi sekolah dengan asosiasi profesi dan DU/DI; (b) kinerja asesor; (c) sarana prasarana penunjang UKK; (d) sikap siswa terhadap UKK; (e) informasi capaian kompetensi siswa; dan (f) pengakuan legal asosiasi profesi dan DU/DI. (2) hasil pengujian menunjukkan: (a) instrumen dengan responden siswa berdasarkan data uji coba pada tahap implementasi, seluruh instrumen valid, reliabel, dan memenuhi syarat sebagai model yang fit; (b) instrumen dengan responden guru pada tahap implementasi menunjukkan seluruh butir instrumen memiliki nilai validitas > 0,30 dan memenuhi kriteria  $KMO > 0,50$  serta koefisien reliabilitas  $\alpha > 0,70$ .

**Kata kunci:** pengembangan, instrumen evaluasi, uji kompetensi keahlian

## DEVELOPING AN EVALUATION INSTRUMENTS OF THE OFFICE ADMINISTRATION EXPERTISE COMPETENCY TEST IN VOCATIONAL HIGH SCHOOLS

<sup>1)</sup>Suranto, <sup>2)</sup>Muhyadi, <sup>3)</sup>Djemari Mardapi

<sup>1,2)</sup>Fakultas Ekonomi Universitas Negeri Yogyakarta,

<sup>3)</sup>Fakultas Teknik Universitas Negeri Yogyakarta

<sup>1)</sup>suranto@uny.ac.id, <sup>2)</sup>muhyadi@uny.ac.id, <sup>3)</sup>djemarimardapi@gmail.com

### Abstract

This study aims to develop an evaluation instruments which is appropriate to evaluate the implementation of the Expertise Competency Test (ECT) for the Office Administration Expertise in vocational high schools (VHSs). This developmental research includes four major stages, namely the preliminary stage, the planning and development stage, the trial stage for evaluation and revision, as well as the implementation stage. For the data collection techniques, this study used *focus group discussion*, questionnaire, interview, observation, and document study. The data analysis for the students' responds was carried out using *Lisrel Program* version 8.51., while that for the teachers' responds was carried out using *SPSS version 17.00 for Windows*. The results of the study are as follows. 1) the components of the ECT activities include: (a) the collaboration of the schools and professional associations and the Business Sector/Industrial Sector (BS/IS); (b) the assessors' performances; (c) infrastructure facilities supporting the ECT; (d) students' attitudes towards the ECT; (e) the information about the students' attainments, and (f) the legal recognition from professional associations and the Business Sector/Industrial Sector or the BS/IS. 2) The study showed that: (a) based on the tryout data in the implementation stage, all items in the instrument with student respondents were valid, reliable, and qualified as a fit model; (b) on the implementation stage, all items in the instrument with teacher respondents had the validity value > 0.3 and, therefore, met the criteria of  $KMO > 0.5$  and the coefficient of reliability  $\alpha > 0.70$ .

**Keywords:** development, evaluation instrument, expertise competency test

## Pendahuluan

Sekolah Menengah Kejuruan (SMK) dirancang untuk menyiapkan peserta didik atau lulusan yang siap memasuki dunia kerja dan memiliki kompetensi di bidang kejuruan. Menurut Rupert Evans (Djojonegoro, 1998, p.33) pendidikan kejuruan ialah bagian dari sistem pendidikan yang mempersiapkan seseorang agar lebih mampu bekerja pada satu bidang pekerjaan. Salah satu upaya untuk menghasilkan lulusan yang memiliki kompetensi unggul, pemerintah melaksanakan kegiatan Uji Kompetensi Keahlian (UKK) dengan menerapkan sistem pengujian kompetensi atau "*competency test*" yang dilakukan untuk komponen mata pelajaran produktif dalam bentuk ujian praktik dan tertulis secara komprehensif-integratif. Mata pelajaran produktif yang diujikan pada UKK Administrasi Perkantoran mencakup: mengetik kecepatan, membuat surat, menangani arsip, mengelola kas kecil, membuat agenda kegiatan pimpinan, mengelola data dan informasi, menerima dan melakukan panggilan telepon, presentasi, menangani perjalanan bisnis, dan membuat notula rapat.

Pada hakikatnya UKK Administrasi Perkantoran adalah kegiatan yang bertujuan mengumpulkan data untuk mengukur penguasaan kompetensi siswa di bidang administrasi perkantoran. Kegiatan UKK terdiri dari dua kategori, yakni: (1) Uji Teori Kejuruan (UTK) dimaksudkan untuk menguji aspek teori-teori yang berbasis pada kompetensi keahlian administrasi perkantoran, nilai diakumulasikan dengan nilai rapor; (2) Uji Praktik Kejuruan (UPK) ialah mata uji untuk aspek praktik program keahlian yang melibatkan pihak eksternal (Dunia Kerja dan Industri). Nilai UPK menjadi dasar penerbitan sertifikat kompetensi.

Pelibatan secara aktif pihak eksternal sebagai *asesor* merupakan komitmen sekolah untuk dievaluasi oleh pihak eksternal guna mengkondisikan agar cakupan pembelajaran di SMK mengarah kepada kompetensi yang dituntut oleh dunia kerja. Selain itu, kegiatan UKK dengan melibatkan pihak eksternal, pada hakikatnya adalah sebuah

pengakuan dari dunia kerja terhadap kompetensi siswa SMK. Menurut Uwe Schippers & Patriana (Sigalingging, 2009, p.4) seorang lulusan SMK dianggap memiliki kompetensi bila ia telah memiliki kualifikasi kejuruan (kompetensi profesi), kualifikasi kejuruan penunjang, kualifikasi interdisipliner, kualifikasi operasional, dan kualifikasi kepribadian (personalitas).

Kegiatan UKK memiliki makna penting bagi sekolah kejuruan yang berusaha mempersiapkan lulusannya memasuki dunia kerja. Dalam Surat Edaran Direktorat Jenderal Manajemen Pendidikan Dasar dan Menengah Nomor: 351/C.C5/MN/2007 (Surat Edaran, 2007) menguatkan posisi pentingnya UKK bagi siswa SMK, karena nilai kelulusan UKK menjadi pertimbangan kelulusan peserta ujian nasional SMK.

Direktorat Pembinaan SMK (2011, pp.2-4) menguraikan komponen-komponen yang berpengaruh terhadap kualitas penyelenggaraan ujian praktik kejuruan, mencakup: (1) tempat penyelenggaraan ujian praktik kejuruan, (2) kerja sama penyelenggara tingkat satuan pendidikan dengan dunia usaha/industri/asosiasi profesi/institusi mitra, (3) sarana prasarana bahan, peralatan, dan alat penunjang ujian praktik kejuruan, (4) penguji atau asesor dari dunia usaha/industri/asosiasi profesi/guru sesuai kriteria yang telah ditetapkan, (5) peserta uji atau siswa, (6) hasil ujian praktik kejuruan berupa informasi capaian kompetensi siswa, dan (7) sertifikat yang ditandatangani oleh pihak eksternal.

Kolaborasi atau kerja sama penyelenggara UKK pada satuan pendidikan dengan dunia usaha/industri/asosiasi profesi/institusi mitra, pada hakikatnya pihak sekolah ingin memperoleh informasi dari dunia kerja mengenai standar kompetensi kerja yang dibutuhkan oleh pasar kerja. Berdasarkan standar kompetensi tersebut, maka sekolah menyusun kurikulum yang relevan. Selanjutnya kurikulum diimplementasikan dalam proses pembelajaran. Untuk mengetahui keberhasilan pembelajaran, dilakukan uji kompetensi. Dengan demikian, materi uji kompetensi juga harus sesuai dengan cakup-

an SKL serta materi pembelajaran. Dengan demikian kolaborasi dalam menyiapkan perangkat UKK, antara sekolah dengan asosiasi profesi dan DU/DI yang diwakili oleh para asesor eksternal, perlu mempertimbangkan SKL pada sekolah yang bersangkutan.

Kualifikasi asesor sesuai dengan tujuan UKK juga perlu diperhatikan secara sungguh-sungguh. Asesor dapat berasal dari asosiasi profesi, maupun dari pengguna lulusan, yaitu DU/DI. Hal ini disebabkan, secara rasional, pihak-pihak yang paling mengetahui perkembangan kompetensi terkini yang dibutuhkan oleh dunia kerja adalah para praktisi di lapangan, yaitu asosiasi profesi dan DU/DI. Begitu pula kesesuaian antara ketersediaan sarana dan prasarana UKK dengan kebutuhan, juga menentukan kualitas penyelenggaraan kegiatan UKK.

Produk penyelenggaraan UKK ialah informasi capaian kompetensi yang menjadi dasar pengambilan keputusan kelulusan siswa. Apabila siswa lulus, maka menjadi dasar bagi asesor eksternal dari asosiasi profesi atau DU/DI untuk menerbitkan sertifikat kompetensi, dan menjadi pertimbangan bagi sekolah terkait kelulusan siswa dari SMK untuk mendapatkan ijazah.

Idealnya, setiap kegiatan pendidikan perlu dievaluasi untuk mengetahui tingkat keberhasilan pelaksanaan kegiatan tersebut sesuai dengan tujuan yang diharapkan. Berangkat dari permasalahan tersebut, dirasa penting adanya pengembangan instrumen evaluasi penyelenggaraan kegiatan UKK. Hal ini disebabkan, kegiatan UKK adalah sebuah aktivitas pendidikan yang sangat strategis untuk mengungkap capaian kompetensi siswa. Diperlukan evaluasi untuk mengetahui tingkat keberhasilan penyelenggaraan kegiatan UKK tersebut, sehingga menghasilkan informasi yang dapat digunakan oleh pimpinan sekolah untuk memperbaiki pelaksanaan UKK pada masa-masa mendatang. Secara rasional, ketersediaan instrumen evaluasi akan memotivasi pimpinan sekolah untuk melakukan evaluasi.

Instrumen evaluasi yang dikembangkan dalam penelitian ini adalah instrumen

yang digunakan untuk mengevaluasi penyelenggaraan kegiatan UKK Administrasi Perkantoran di SMK. Salah satu alasan yang melatarbelakangi pengembangan instrumen evaluasi ini adalah adanya tuntutan peningkatan kualitas penyelenggaraan UKK. Bahwa dunia usaha dan industri menuntut agar penyelenggaraan UKK dapat ditingkatkan kualitasnya sehingga mampu menilai kesesuaian capaian kompetensi siswa SMK dengan tuntutan dunia kerja. Agar diperoleh informasi yang berguna sebagai pertimbangan peningkatan kualitas penyelenggaraan UKK, maka pengembangan instrumen evaluasi dipandang sebagai upaya strategis.

Dalam beberapa literatur (Cullingford, 1997; Djemari Mardapi, September 2000; Ebel & Frisbie, 1986; Johnson & Christensen, 2008), dapat dirangkum beberapa keuntungan dari evaluasi yang diterapkan untuk berbagai kegiatan dan program pendidikan, yakni terciptanya sebuah metode untuk menilai apakah kegiatan yang diselenggarakan tersebut telah sesuai dengan keinginan semula, dapat digunakan untuk melihat kembali apakah suatu kegiatan telah dapat dilaksanakan sesuai dengan perencanaan dan mencapai hasil sesuai yang diharapkan, selanjutnya informasi yang diperoleh dari evaluasi digunakan untuk menentukan alternatif yang tepat dalam mengambil sebuah keputusan.

Evaluasi merupakan salah satu rangkaian kegiatan untuk meningkatkan kualitas, kinerja, maupun produktivitas suatu lembaga dalam melaksanakan kegiatan. Melalui evaluasi akan diperoleh informasi tentang sejauhmana suatu kegiatan dapat dilaksanakan, selanjutnya informasi ini digunakan untuk perbaikan kegiatan tersebut. Menurut Royse, Thyer, dan Padgett (2010, p. 12) evaluasi merupakan bagian dari proses managerial yang diterapkan untuk memperoleh informasi sebagai acuan dalam pembuatan keputusan. Evaluasi bukan merupakan hal baru dalam kehidupan organisasi sebab evaluasi senantiasa mengiringi dinamika kehidupan organisasi. Fitzpatrick, Sanders, dan Worthen (2011, p. 5) mengatakan, "*evaluation is not a new concept. In fact, people have been*

*evaluating, or examining and judging things, since the beginning of human history*".

Faktor yang cukup dominan dalam menentukan keberhasilan suatu kegiatan pendidikan adalah kualitas komponen-komponen yang ada pada program pendidikan itu sendiri. Menurut Cox (2006, p. 8), "*the quality of an educational program is compared of three elements, materials (and equipment), activities, and people*". Berdasarkan pendapat tersebut dapat dipahami bahwa kualitas program pendidikan tergantung pada sarana dan prasarana pembelajaran, aktivitas tenaga kependidikan dan siswa dalam kegiatan pendidikan, serta semua personal yang terlibat dalam kegiatan itu.

Berbagai penelitian dan evaluasi tentang kegiatan pendidikan menguatkan pandangan tentang pentingnya informasi tingkat keberhasilan penyelenggaraan suatu kegiatan. Penelitian tersebut diantaranya dilakukan oleh Mizikaci (2007, p.12), menyimpulkan bahwa evaluasi untuk mengetahui atau mendeteksi kualitas sebuah program pendidikan, berawal dari asumsi-asumsi tentang "kualitas" suatu program itu sendiri, meliputi: (a) *quality as excellent*: kualitas sebagai hal yang baik; (b) *quality as zero errors*: kualitas dengan eror nol; (c) *quality as fitness for purposes*: kualitas sebagai pemenuhan kebutuhan; (d) *quality as transformation*: kualitas sebagai perubahan; (e) *quality as threshold*: kualitas sebagai ambang batas; (f) *quality as value for money*: kualitas sebagai nilai untuk memperoleh uang; (g) *quality as enhancement or improvement*: kualitas sebagai peningkatan atau perbaikan.

Selanjutnya penelitian yang dilakukan oleh Crawford (2006, p. 28) menunjukkan adanya komitmen, bahwa untuk mengetahui berbagai hasil yang terkait dengan kegiatan, tidak cukup dengan mencermati hasil kegiatan itu. Langkah yang tidak kalah pentingnya adalah mengumpulkan informasi untuk menentukan atau mengetahui bahwa hasil itu benar-benar karena suatu kegiatan, dan bukan karena faktor lain di luar kegiatan. Sanders & Sullins (2006, p. 1) menjelaskan, "*program evaluation is the process of systematically determining the quality of a program and*

*how it can be improved*". Esensi dari definisi tersebut bahwa evaluasi program adalah proses sistematis menentukan kualitas program dan upaya meningkatkan kualitas program tersebut. Terkait dengan hal-hal tersebut dalam penelitian ini bertujuan mengembangkan instrumen yang cocok untuk melakukan evaluasi penyelenggaraan kegiatan UKK administrasi perkantoran di Sekolah Menengah Kejuruan (SMK).

## Metode Penelitian

Penelitian ini menggunakan metode Penelitian Pengembangan (*Research and Development*), ialah metode penelitian yang digunakan untuk menghasilkan produk tertentu, dan menguji keefektifan produk tersebut. Dalam penelitian ini produk yang dikembangkan ialah instrumen evaluasi penyelenggaraan kegiatan UKK administrasi perkantoran di SMK. Model pengembangan yang digunakan sebagai kajian pada penelitian dan pengembangan ini mengacu pada model Borg dan Gall dengan dilakukan teknik memodifikasi dari 10 langkah penelitian menjadi empat tahapan penelitian, yakni tahap pendahuluan, tahap perencanaan dan pengembangan, tahap uji coba, dan tahap implementasi.

Modifikasi langkah-langkah penelitian tersebut dirancang dan dilaksanakan untuk dapat mengembangkan instrumen evaluasi penyelenggaraan kegiatan UKK administrasi perkantoran di SMK. Disamping itu, modifikasi langkah-langkah penelitian juga menyesuaikan langkah-langkah pengembangan instrumen yang dikemukakan oleh Mardapi (2005, pp.16-21), yaitu: (1) berdasarkan teori-teori tentang konsep dari variabel yang hendak diukur, dirumuskan konstruk variabel, (2) dikembangkan dimensi dan indikator, (3) membuat kisi-kisi instrumen, (4) menetapkan besaran atau parameter, (5) menulis butir-butir instrumen, (6) proses validasi, (7) revisi, (8) uji coba di lapangan.

Subjek uji coba atau responden yang terlibat dalam penelitian ini terdiri dari pakar evaluasi, kepala sekolah atau wakil kepala sekolah sebagai representasi dari pimpinan sekolah, guru administrasi perkantoran

an, dan siswa peserta kegiatan UKK KK Administrasi Perkantoran.

Instrumen yang digunakan dalam pengumpulan data, juga sekaligus merupakan instrumen evaluasi yang dikembangkan yaitu inventori dalam bentuk *rating scale* untuk penyelenggaraan UKK Administrasi Perkantoran di SMK.

Analisis data uji coba lapangan bertujuan untuk memperoleh bukti validitas konstruk dan reliabilitas instrumen, yaitu menentukan kesesuaian antara konstruk teori dengan hasil pengukuran di lapangan. Analisis uji coba lapangan dilakukan dengan teknik analisis faktor konfirmatori. Instrumen pengumpul data dengan responden siswa dianalisis dengan *Second Order Confirmatory Factor Analysis (CFA)* program *Lisrel 8.70*, sedangkan instrumen pengumpul data dengan responden guru dianalisis dengan *Exploratory Factor Analysis (EFA)* menggunakan *SPSS 17,00 for windows*. Penentuan validitas butir instrumen dengan responden siswa menggunakan CFA didasarkan pada kriteria nilai muatan faktor  $>0,3$  dan nilai  $t > 1,96$  untuk masing-masing butir instrumen. Hal ini sesuai dengan pendapat Hair (2010, p.119) "*factor loadings  $\pm 0,3$  to  $0,4$  are minimally acceptable*."

Selain untuk menganalisis validitas butir instrumen, *Lisrel* digunakan juga untuk menguji kecocokan model pengukuran (fit model). Model instrumen yang dikembangkan dinyatakan cocok dengan data lapangan apabila sudah terpenuhi dua kriteria dari tiga kriteria yang menjadi ukuran kecocokan yaitu *Root Mean Square Error of Approximation (RMSEA) < 0,08*; *Chi-Square* yang diperoleh dari pengujian memiliki probabilitas lebih besar dari 0,05 ( $p > 0,05$ ); dan *Goodness of Fit Index (GFI) > 0,90*.

Validitas instrumen dengan responden guru dianalisis dengan analisis faktor menggunakan *SPSS 17 for windows*. Uji validitas dilakukan menggunakan *Exploratory Faktor Analysis (EFA)*. Selain itu, *EFA* digunakan untuk mengetahui apakah suatu konstruk dapat dijelaskan oleh indikator-indikatornya. Apabila indikator-indikator dapat membentuk konstruk atau variabel, maka ditunjukkan dengan nilai *loading factor* yang tinggi ( $>$

0,3) berarti pengukuran sudah sesuai dengan data, dan nilai *Kaiser Meyer Olkin (KMO) > 0,5* (Ghozali, 2005, pp.29-31). Untuk mengetahui validitas instrumen ditentukan dengan nilai muatan faktor. Butir-butir yang terdapat dalam setiap faktor harus memiliki muatan lebih besar 0,3 (Nurosis, 1986, p.12). Untuk menetapkan reliabilitas konstruk, digunakan formula *Cronbach's Alpha* yaitu sekurang-kurangnya 0,7.

## Hasil Penelitian dan Pembahasan

### Hasil Penelitian Tahap Pendahuluan

Data yang disajikan pada hasil penelitian tahap pendahuluan ini, diperoleh melalui proses pengumpulan informasi dan studi pendahuluan yang meliputi: observasi awal di berbagai SMK untuk identifikasi masalah riil di lapangan, kajian pustaka, kajian peraturan, dan kajian hasil penelitian terdahulu. Tahapan ini mempunyai dua tujuan utama yaitu: (1) menemukan dan menetapkan variasi jenis SMK, terkait dengan visi misi sekolah, sistem pendidikan, sistem pengelolaan, serta sarana dan prasarana yang tersedia; dan (2) menemukan dan menetapkan komponen-komponen penyelenggaraan kegiatan UKK yang menjadi penentu kualitas penyelenggaraan UKK Administrasi Perkantoran di SMK yang menjadi dasar penjabaran menjadi konstruk teoritik dan indikator instrumen evaluasi IE-UKK.

Hasil studi pendahuluan dengan observasi awal di berbagai SMK dan divalidasi melalui *Focus Group Discussion (FGD)* pimpinan sekolah dan para guru, ditetapkan empat variasi jenis SMK, yaitu (1) SMK pemerintah (SMK Negeri), (2) SMK berbasis agama Islam, (3) SMK berbasis agama non-Islam, dan (4) SMK swasta berbasis yayasan sosial/pendidikan.

Berdasarkan studi literatur *Petunjuk teknis pelaksanaan ujian kompetensi keahlian SMK tahun pelajaran 2010/2011* (Direktorat Pembinaan SMK, 2011, pp.2-4) dapat ditemukan komponen-komponen yang berpengaruh terhadap kualitas penyelenggaraan ujian praktik kejuruan, mencakup: (1) tempat penyelenggaraan ujian praktik kejuruan, (2) kerja sama penyelenggara tingkat satuan

pendidikan dengan dunia usaha/industri/asosiasi profesi/institusi mitra, (3) sarana prasarana bahan, peralatan, dan alat penunjang ujian praktik kejuruan, (4) penguji atau asesor dari dunia usaha/industri/asosiasi profesi/guru sesuai kriteria yang telah ditetapkan, (5) peserta uji atau siswa, (6) pengorganisasian UKK, (7) hasil ujian praktik kejuruan berupa informasi capaian kompetensi siswa, dan (8) sertifikat yang ditandatangani oleh pihak eksternal.

Berdasarkan penilaian dan validasi melalui *Focus Group Discussion (FGD)* pimpinan sekolah dan para guru, ditetapkan komponen-komponen penyelenggaraan kegiatan UKK yang menjadi penentu kualitas penyelenggaraan UKK Administrasi Perkantoran di SMK yang menjadi dasar penjabaran menjadi konstruk teoritik dan indikator instrumen IE-UKK, mencakup: (1) kolaborasi sekolah dengan asosiasi profesi atau DU/DI; (2) kinerja asesor; (3) ketersediaan sarana dan prasarana penunjang UKK; (4) sikap siswa terhadap kegiatan UKK; (5) informasi capaian kompetensi siswa peserta uji kompetensi; dan (6) pengakuan legal dari pihak eksternal (asosiasi profesi atau DU/DI).

### Hasil Uji Coba

#### Keterbacaan Instrumen

Uji keterbacaan instrumen didasarkan pada hasil validasi (penilaian) dari para ahli evaluasi, pemakai model evaluasi (kepala sekolah), guru kompetensi administrasi perkantoran, serta praktisi yang memberi masukan-masukan dalam rangka perbaikan instrumen. Penilaian terhadap keterbacaan instrumen diarahkan pada aspek: (1) kejelasan petunjuk instrumen, (2) kejelasan indikator kolaborasi sekolah dengan asosiasi profesi dan DU/DI, (3) kejelasan indikator kinerja asesor, (4) kejelasan indikator sarana dan prasarana penunjang UKK, (5) kejelasan indikator sikap siswa terhadap UKK, (6) kejelasan indikator ketersediaan informasi capaian kompetensi siswa, (7) kejelasan indikator pengakuan legal asosiasi profesi, (8) penggunaan bahasa Indonesia baku, (9) perumusan pernyataan yang komunikatif, (10) bentuk huruf, (11) ukuran huruf, dan (12) tanda baca. Adapun hasil penilaian dari tiga tahap uji coba terhadap keterbacaan instrumen disajikan pada Tabel 1.

Tabel 1. Rekapitulasi Hasil Uji Keterbacaan Instrumen Evaluasi UKK (IE-UKK)

Tahap	Jumlah responden	Rerata Skor												Rerata Total
		1	2	3	4	5	6	7	8	9	10	11	12	
1	20	3,80	3,80	3,85	3,80	3,40	4,00	3,90	3,90	3,40	4,00	3,90	3,90	3,81
2	54	4,00	3,90	4,03	3,90	4,01	4,01	3,90	3,90	3,40	4,00	3,90	3,90	3,91
3	65	4,03	3,95	3,93	3,93	4,01	4,01	4,01	3,98	4,45	3,85	3,97	3,90	3,92

Penilaian keterbacaan instrumen menggunakan skala 5, yaitu skor minimal 1 dan skor maksimal 5. Tabel 1 menunjukkan rerata skor total pada uji coba ke-3 adalah 3,92. Apabila dikonsultasikan dengan pedoman konversi data kuantitatif ke kualitatif, berada pada rerata skor > 3,4 – 4,2 masuk dalam klasifikasi layak atau baik. Hal ini memberi indikasi bahwa tingkat keterbacaan instrumen yang dikembangkan ini dapat diklasifikasikan layak atau baik.

#### Hasil Pengujian Instrumen dengan Responden Siswa

Instrumen pengumpul data dengan responden siswa dianalisis dengan *Second Order Confirmatory Factor Analysis (CFA)*

program *Lisrel 8.70*. Imam Ghozali (2005, p.157) mengatakan bahwa *second order confirmatory* digunakan apabila satu faktor laten memiliki beberapa indikator, dan indikator-indikator tersebut tidak langsung dapat diukur karena memerlukan indikator lagi. Penentuan validitas butir instrumen dengan responden siswa menggunakan CFA didasarkan pada kriteria nilai muatan faktor > 0,3 dan nilai  $t > 1,96$  untuk masing-masing butir instrumen. Apabila nilai alpha lebih besar dari 5% dan nilai  $t$  kurang dari  $t_{tabel}$  yaitu sebesar 1,96 (pada diagram *second order* menunjukkan warna merah) berarti nomor butir yang bersangkutan masuk dalam kate-

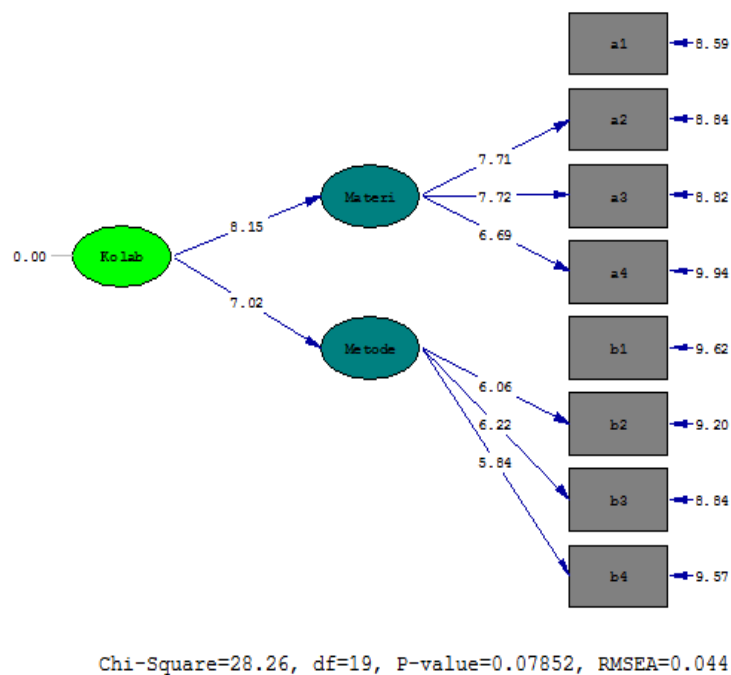
gori tidak valid dan harus di-drop. Sebaliknya apabila nilai  $t$  lebih besar dari  $t_{\text{tabel}}$  sebesar 1,96 (pada diagram *second order* menunjukkan warna hitam), maka nomor butir tersebut termasuk dalam kategori valid.

Selain untuk menganalisis validitas butir instrumen, *Lisrel* digunakan juga untuk menguji kecocokan model pengukuran (fit model). Sebuah model dikatakan cocok (fit) apabila model teoritis atau model hipotetik yang dikembangkan pada penelitian cocok (fit) didukung oleh data empirik (Imam Ghozali, 2005, pp.29-31).

Model hipotetik yang diuji secara empiris dalam penelitian ini adalah instrumen evaluasi uji kompetensi keahlian administrasi perkantoran, yang terdiri dari: (1) kolaborasi antara sekolah dengan asosiasi profesi dan DU/DI; (2) kinerja asesor; (3) sarana dan prasarana penunjang UKK; (4) sikap siswa terhadap UKK; (4) ketersediaan informasi capaian kompetensi siswa; dan (5) pengakuan legal asosiasi profesi atau DU/DI. Model instrumen yang dikembangkan dinyatakan cocok dengan data lapangan apabila sudah terpenuhi dua kriteria dari tiga kriteria yang menjadi ukuran kecocokan yaitu *Root Mean Square Error of Approximation*

(*RMSEA*) < 0,08; *Chi-Square* yang diperoleh dari pengujian memiliki probabilitas lebih besar dari 0,05 ( $p > 0,05$ ); dan *Goodness of Fit Index (GFI)* > 0,90.

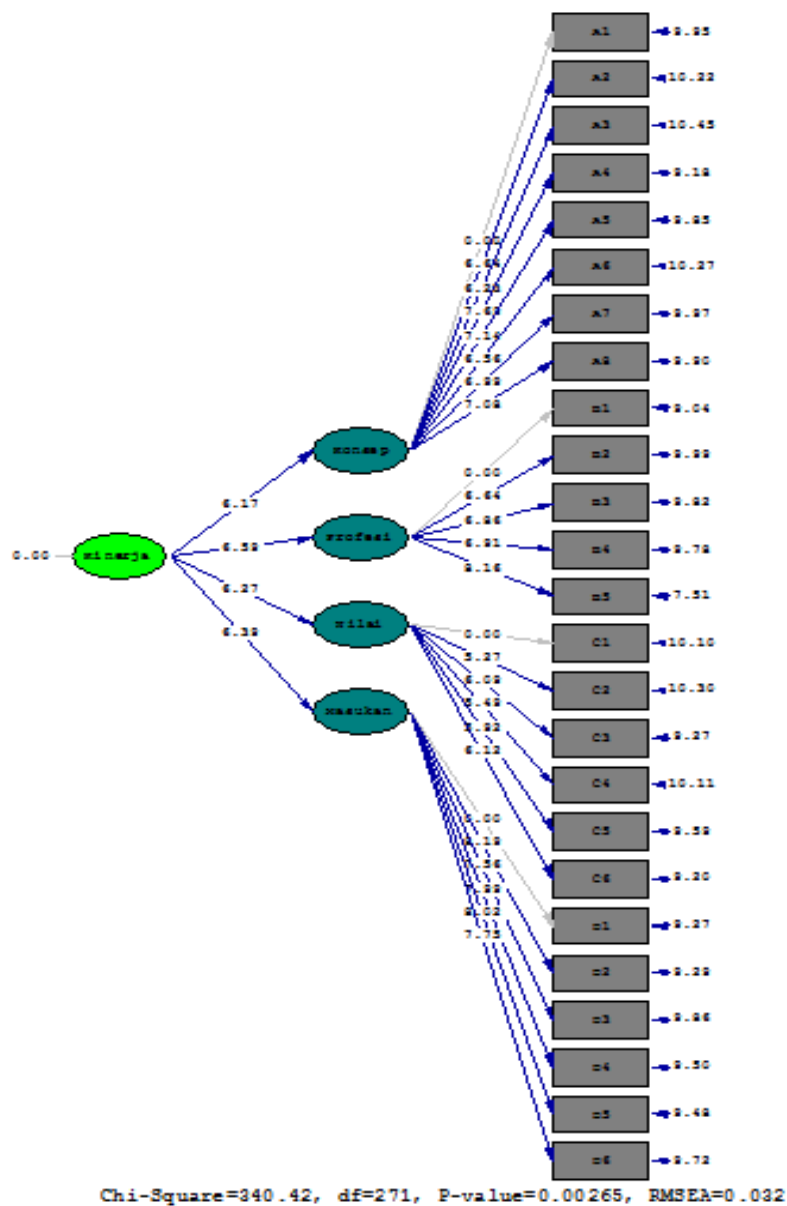
Hasil pengujian pada uji coba tahap operasional instrumen kolaborasi sekolah dengan asosiasi profesi dan DU/DI, menunjukkan bahwa untuk semua nomor butir instrumen kualitas kolaborasi sekolah dengan asosiasi profesi dan DU/DI memiliki muatan faktor di atas 0,3 dan nilai  $t$  lebih besar dari 1,96 sehingga semua butir instrumen dalam kategori valid. Koefisien Alpha menunjukkan untuk aspek 1  $\alpha = 0,799$  dan aspek 2  $\alpha = 0,749$  ( $> 0,7$ ) sehingga instrumen pada kedua aspek tersebut reliabel. Pengujian instrumen dengan *Second Order CFA* menunjukkan bahwa dilihat dari tiga kriteria, yaitu *Chi-Square* dan  $p$  value, *RMSEA*, dan *GFI* maka model masuk dalam kategori fit, karena nilai *Chi-Square* = 28,26  $p$ -value = 0,078 > 0,05 *RMSEA* = 0,067 < 0,08 dan *GFI* = 0,96 > 0,90. Hal tersebut menunjukkan bahwa model pengukuran teoritis instrumen kualitas kolaborasi sekolah dengan asosiasi profesi atau DU/DI cocok (*fit*) dengan model yang diperoleh dari data empirik.



Gambar 1. Hasil Uji Coba Model Pengukuran (Konstruk) Kolaborasi Sekolah dengan DU/DI

Konstruk variabel kinerja asesor terdiri atas empat faktor yaitu aspek penguasaan konsep UKK, keterkaitan profesi dengan materi yang diujikan, kemampuan penilaian, dan kemampuan memberikan masukan kepada sekolah. Hasil pengujian menunjukkan bahwa dari 25 butir instrumen kinerja asesor seluruhnya valid karena memiliki nilai muatan faktor di atas 0,3 dan nilai  $t$  lebih besar dari 1,96. Dilihat dari koefisien Alpha, untuk ke empat aspek semuanya diperoleh angka lebih besar 0,7 sehingga instrumen kinerja asesor dikategorikan reliabel.

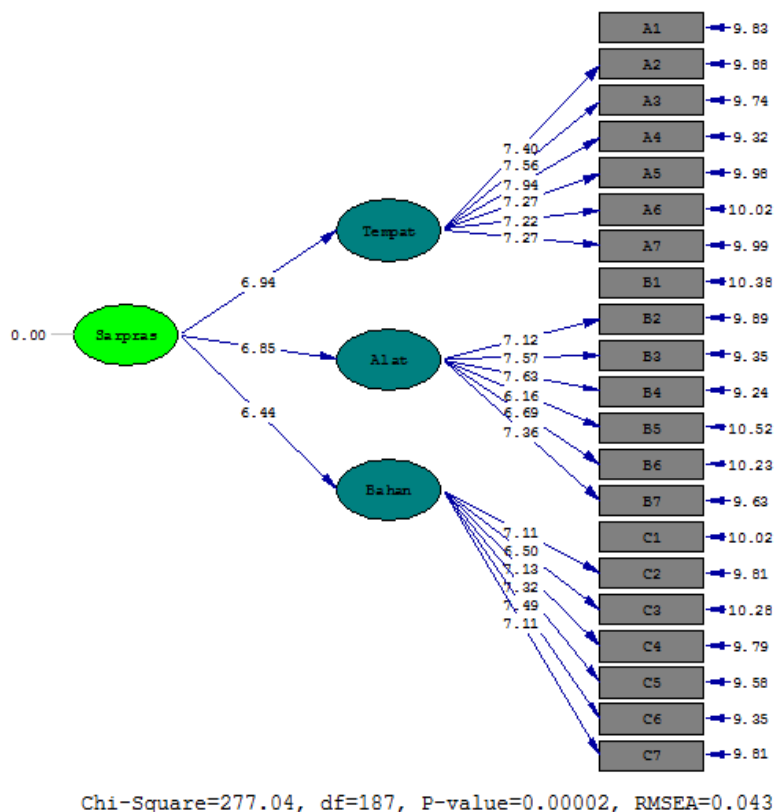
Dilihat dari kriteria kecocokan model pengukuran hipotetik dengan data empirik, menunjukkan nilai  $RMSEA= 0,032 (< 0,80)$  memenuhi syarat sebagai fit model, dan nilai  $GFI=0,91 (>0,90)$  juga telah memenuhi syarat sebagai fit model. Berdasarkan pengujian kecocokan model pengukuran tersebut dapat disimpulkan bahwa instrumen pengukuran kinerja asesor tersebut sudah dapat digunakan sebagai model pengukuran yang cocok untuk mengumpulkan data tentang kinerja asesor UKK.



Gambar 2. Hasil Uji Coba Model Pengukuran (Konstruk) Kinerja Asesor

Hasil pengujian instrumen sarana dan prasarana penunjang UKK, menunjukkan bahwa semua nomor butir instrumen sarana dan prasarana penunjang UKK memiliki muatan faktor di atas 0,3 dan nilai  $t$  lebih besar dari 1,96 sehingga semua butir instrumen dalam kategori valid. Aspek tempat atau ruang ujian beserta kelengkapannya  $\alpha = 0,797$ , peralatan dan mesin kantor  $\alpha = 0,795$ , jumlah dan kualitas bahan  $\alpha = 0,793$  sehingga semua aspek instrumen dalam kategori reliabel.

Dari hasil pengujian ini dapat dikatakan bahwa dari 21 butir instrumen seluruhnya valid karena memiliki nilai muatan faktor  $> 0,3$  dan nilai  $t > 1,96$ . Hasil pengujian menunjukkan nilai  $RMSEA = 0,043$  ( $< 0,80$ ) dan nilai  $GFI = 0,91$  ( $> 0,90$ ), artinya dua dari tiga kriteria sebagai model yang fit terpenuhi, maka instrumen tersebut sudah dapat digunakan sebagai model pengukuran yang cocok untuk mengumpulkan data tentang sarana dan prasarana penunjang UKK.

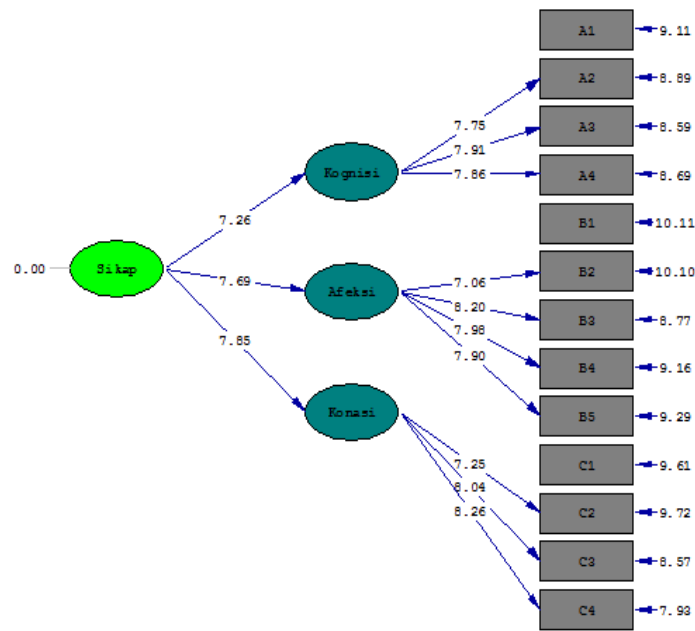


Gambar 3. Hasil Uji Coba Konstruk Sarana dan Prasarana Penunjang UKK

Hasil pengujian instrumen sikap siswa terhadap UKK, menunjukkan bahwa semua butir instrumen sikap siswa terhadap UKK memiliki muatan faktor di atas 0,3 dan nilai  $t$  lebih besar dari 1,96 sehingga semua butir instrumen dalam kategori valid. Pengujian reliabilitas menunjukkan bahwa pada aspek kognisi  $\alpha=0,735$ , aspek afeksi  $\alpha=0,757$ , dan aspek konasi  $\alpha=0,737$ , oleh karena  $\alpha$  lebih besar 0,7 instrumen pada ketiga aspek tersebut dikategorikan reliabel Hasil pengujian ini, memberikan indikasi dari 13 butir instru-

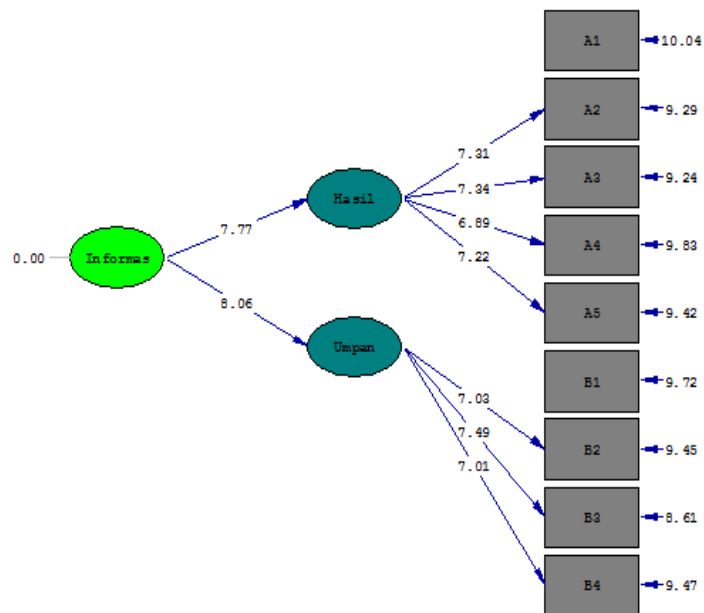
men seluruhnya valid karena memiliki nilai muatan faktor di atas 0,3 dan nilai  $t$  lebih besar dari 1,96.

Pengujian model *Second Order CFA* menunjukkan bahwa dilihat dari ukuran  $RMS- EA$  dan  $GFI$  maka model masuk dalam kategori fit, karena nilai  $RMSEA = 0,052 < 0,08$  dan  $GFI=0,91 > 0,90$ . Hal tersebut menunjukkan bahwa model pengukuran teoretis instrumen sikap siswa terhadap UKK cocok (*fit*) dengan model yang diperoleh dari data empirik.



Chi-Square=106.53, df=63, P-value=0.00051, RMSEA=0.052

Gambar 4. Hasil Uji Coba Model Pengukuran (Konstruk) Sikap Siswa



Chi-Square=37.66, df=26, P-value=0.06516, RMSEA=0.042

Gambar 5. Hasil Uji Coba Konstruk Ketersediaan Informasi Capaian Kompetensi Siswa

Konstruk variabel informasi capaian kompetensi siswa mencakup dua aspek, yaitu: ketersediaan informasi hasil UKK dan umpan balik dari asesor untuk meningkatkan kompetensi siswa. Hasil pengujian instrumen ketersediaan informasi capaian

kompetensi siswa, menunjukkan bahwa semua butir instrumen informasi capaian kompetensi siswa memiliki muatan faktor di atas 0,3 dan nilai t lebih besar dari 1,96 sehingga semua butir valid. Analisis reliabilitas pada aspek ketersediaan informasi hasil

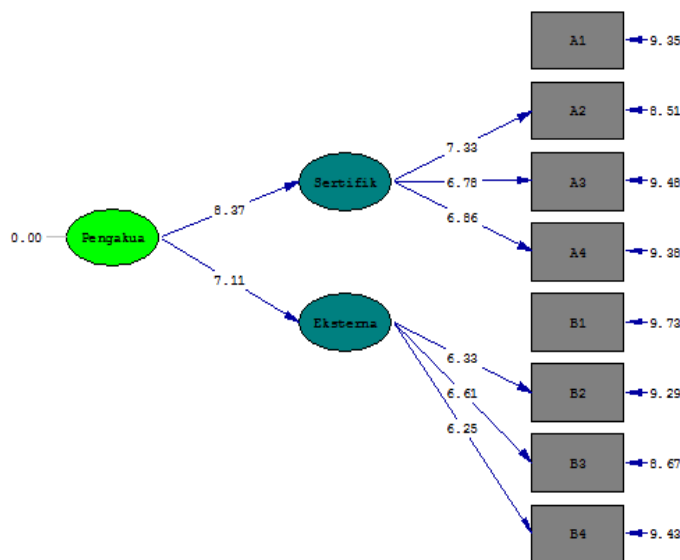
UKK  $\alpha=0,781$  dan aspek umpan balik dari asesor untuk meningkatkan kompetensi siswa  $\alpha=0,724$ , oleh karena  $\alpha$  lebih besar 0,7 maka instrumen pada ke dua aspek dikategorikan reliabel. Hasil pengujian ini mengindikasikan dari 9 butir instrumen informasi capaian kompetensi siswa seluruhnya valid karena memiliki nilai muatan faktor di atas 0,3 dan nilai  $t$  lebih besar dari 1,96.

Pengujian model *Second Order CFA* menunjukkan bahwa dilihat dari ukuran  $p$  value, RMSEA dan GFI maka model masuk dalam kategori fit, karena nilai  $p$ -value =  $0,06516 > 0,05$ , RMSEA =  $0,042 < 0,08$  dan GFI =  $0,97 > 0,90$ . Hal tersebut menunjukkan bahwa model pengukuran teoritis instrumen informasi capaian kompetensi siswa cocok (*fit*) dengan model yang diperoleh dari data empirik.

Instrumen pengakuan legal asosiasi profesi atau DU/DI mencakup dua aspek, yaitu penerbitan sertifikat kompetensi dan pengakuan pihak eksternal. Hasil pengujian instrumen pengakuan legal asosiasi profesi

dan DU/DI memperlihatkan, bahwa semua butir instrumen memiliki muatan faktor di atas 0,3 dan nilai  $t$  lebih besar dari 1,96 sehingga semua butir instrumen dalam kategori valid. Reliabilitas instrumen didasarkan pada kriteria  $\alpha > 0,7$  sedangkan hasil analisis menunjukkan aspek 1 mempunyai  $\alpha=0,708$  dan aspek 2  $\alpha = 0,777$  sehingga instrumen pada kedua aspek tersebut reliabel.

Pengujian model *Second Order CFA* menunjukkan bahwa dilihat dari ukuran RMSEA dan GFI maka model masuk dalam kategori fit, karena nilai RMSEA =  $0,054 < 0,08$  dan GFI =  $0,97 > 0,90$ . Hal tersebut menunjukkan bahwa model pengukuran teoritis instrumen pengakuan legal asosiasi profesi atau DU/DI cocok (*fit*) dengan model yang diperoleh dari data empirik. Oleh karena dua kriteria sebagai model yang fit terpenuhi, maka instrumen tersebut sudah dapat digunakan sebagai model pengukuran yang cocok untuk mengumpulkan data tentang pengakuan legal pihak eksternal, yaitu asosiasi profesi atau DU/DI.



Chi-Square=32.93, df=19, P-value=0.02447, RMSEA=0.054

Gambar 6. Hasil Uji Coba Konstruk Pengakuan Legal Asosiasi Profesi dan DU/DI

*Hasil Pengujian Instrumen dengan Responden Guru*

Instrumen pengumpul data dengan responden guru dianalisis menggunakan *Exploratory Factor Analysis (EFA)* dengan bantuan program *SPSS 17,00 for windows*,

Analisis dengan *EFA* digunakan untuk mengetahui apakah konsep yang telah disusun berdasarkan kajian teori dan telah diuji secara empirik, didukung data atau tidak. Teknik ini juga dapat digunakan untuk mengestimasi validitas instrumen.

Analisis *EFA* ini digunakan untuk mengetahui apakah suatu konstruk dapat dijelaskan oleh indikator-indikatornya. Apabila indikator-indikator dapat membentuk konstruk atau variabel, maka ditunjukkan dengan nilai *loading factor* yang tinggi ( $>0,3$ ) berarti pengukuran sudah sesuai dengan data, dan nilai *Kaiser Meyer Olkin (KMO)*  $> 0,5$  (Imam Ghazali, 2005, pp.29-31). Untuk mengetahui validitas instrumen ditentukan dengan nilai muatan faktor. Butir-butir yang terdapat dalam setiap faktor harus memiliki muatan lebih besar  $0,3$  (Nurosis, 1986, p.123).

Hasil pengujian instrumen kolaborasi sekolah dengan asosiasi profesi dan DU/DI menunjukkan besaran nilai *Bartlett's Tes of Sphericity* adalah 228,050 pada signifikansi 0,000. Hal ini berarti pada aspek ini ada korelasi yang signifikan antarvariabel, dan hasil perhitungan *KMO* sebesar 0,670 sehingga kecukupan sampel terpenuhi.

Aspek kolaborasi sekolah dengan asosiasi profesi dan DU/DI pada uji coba tahap ketiga ini, dikembangkan dua faktor. Setelah dilakukan analisis, kedua faktor tersebut memiliki nilai *eigen* lebih dari satu, sehingga terpenuhi syarat sebagai faktor yang bisa dikembangkan.

Berdasarkan *output rotated factor matrix* ada dua faktor kolaborasi sekolah dengan asosiasi profesi dan DU/DI, yaitu faktor 1 penyusunan cakupan materi UKK dan faktor 2 kolaborasi penetapan metode penilaian. Selanjutnya disajikan sebaran butir-butir instrumen kolaborasi sekolah dengan asosiasi profesi dan DU/DI. Selanjutnya hasil uji coba ketiga terdapat delapan butir pernyataan untuk instrumen kolaborasi sekolah dengan asosiasi profesi dan DU/DI, karena satu butir nomor A5 didrop pada uji coba kedua. Kedelapan butir pernyataan tersebut masuk dalam faktor 1 dan 2 dengan *total varian explained* sebesar 61,631%. Hal ini menunjukkan kedua faktor yang dikembangkan mampu mengukur variabel kolaborasi asosiasi profesi dan DU/DI sebesar 61,631%.

Dilihat dari muatan faktor mengindikasikan untuk semua nomor butir instrumen kolaborasi sekolah dengan asosiasi

profesi dan DU/DI pada uji coba ketiga ini memiliki muatan faktor di atas  $0,3$  sehingga semua butir masuk kategori valid. Untuk reliabilitas, instrumen pada aspek penyusunan cakupan materi UKK  $\alpha=0,891$  dan aspek kolaborasi penetapan metode penilaian  $\alpha=0,818$  (lebih besar dari  $0,7$ ) sehingga disimpulkan instrumen pada kedua aspek tersebut reliabel.

Hasil pengujian instrumen kinerja asesor menunjukkan besaran nilai *Bartlett's Tes of Sphericity* adalah 937,417 pada signifikansi 0,000. Hal ini berarti pada aspek kinerja asesor ini ada korelasi yang signifikan antarvariabel. Hasil perhitungan *KMO* sebesar 0,689 (lebih besar dari  $0,5$ ) sehingga kecukupan sampel terpenuhi.

Berdasarkan sebaran butir, terdapat 27 butir pernyataan atau pertanyaan untuk instrumen kinerja asesor, dan masuk dalam faktor 1, 2, 3, dan 4 dengan *total varian explained* sebesar 56,970%. Hal ini menunjukkan ke empat faktor yang dikembangkan mampu mengukur variabel kinerja asesor sebesar 56,970%.

Selanjutnya, dapat diketahui bahwa semua butir instrumen memiliki nilai validitas  $> 0,3$  dan memenuhi kriteria *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* dengan muatan sebesar 0,689 ( $>0,5$ ). Reliabilitas pada aspek 1 mempunyai  $\alpha=0,885$ , aspek 2  $\alpha = 0,953$ , aspek 3  $\alpha=0,837$ , dan aspek 4  $\alpha = 0,847$  ( $> 0,7$ ).

Instrumen sarana dan prasarana penunjang UKK mencakup tiga aspek, yaitu: tempat atau ruang ujian beserta kelengkapannya, peralatan dan mesin kantor, jumlah dan kualitas bahan. Hasil perhitungan menunjukkan besaran nilai *Bartlett's Tes of Sphericity* adalah 665,608 pada signifikansi 0,000 yang berarti pada aspek sarana dan prasarana penunjang UKK ini ada korelasi yang signifikan antarvariabel, dan hasil perhitungan *KMO* sebesar 0,810 ( $>0,5$ ) sehingga kecukupan sampel terpenuhi.

Berdasarkan sebaran butir, terdapat 17 butir pernyataan atau pertanyaan untuk instrumen sarana dan prasarana penunjang UKK, dan masuk dalam faktor 1, 2, dan 3 dengan *total varian explained* sebesar 67,217%.

Hal ini menunjukkan ke tiga faktor yang dikembangkan mampu mengukur variabel sarana dan prasarana penunjang UKK sebesar 67,217%.

Selanjutnya, dapat diketahui bahwa semua butir instrumen memiliki nilai validitas  $>0,3$  dan memenuhi kriteria *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* dengan muatan sebesar 0,810 ( $>0,5$ ). Untuk reliabilitas pada aspek 1 mempunyai  $\alpha=0,943$ , aspek 2  $\alpha=0,869$ , dan aspek 3  $\alpha=0,916$  ( $>0,7$ ).

Hasil pengujian instrumen informasi capaian kompetensi siswa menunjukkan besaran nilai *Bartlett's Test of Sphericity* adalah 255,997 pada signifikansi 0,000. Hal ini berarti pada aspek ketersediaan informasi capaian kompetensi siswa ini ada korelasi yang signifikan antarvariabel, dan hasil perhitungan KMO sebesar 0,744 sehingga kecukupan sampel terpenuhi.

Analisis faktor aspek ketersediaan informasi capaian kompetensi menghasilkan dua faktor yang memiliki nilai eigen lebih dari satu. terdapat 9 butir pernyataan atau pertanyaan untuk instrumen ketersediaan informasi capaian kompetensi siswa, dan masuk dalam faktor 1 dan 2 dengan *total varian explained* sebesar 60,835%. Hal ini menunjukkan kedua faktor yang dikembangkan mampu mengukur variabel ketersediaan informasi capaian kompetensi siswa sebesar 60,835%.

Selanjutnya, dari hasil pengujian dapat diketahui bahwa semua butir instrumen memiliki nilai validitas  $>0,3$  dan memenuhi kriteria *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* dengan muatan sebesar 0,744 ( $>0,5$ ). Untuk reliabilitas pada aspek 1 mempunyai  $\alpha=0,902$  dan aspek 2  $\alpha=0,753$  ( $>0,7$ ).

Hasil pengujian instrumen pengakuan legal asosiasi profesi dan DU/DI menunjukkan besaran nilai *Bartlett's Test of Sphericity* adalah 182 pada signifikansi 0,000 yang berarti pada aspek pengakuan legal asosiasi profesi dan DU/DI ini ada korelasi yang signifikan antarvariabel, dan hasil perhitungan KMO sebesar 0,747 sehingga kecukupan sampel terpenuhi.

Berdasarkan *output rotated factor matrix*, ada dua faktor aspek ketersediaan informasi capaian kompetensi siswa, yaitu faktor 1 pengakuan pihak eksternal, dan faktor 2 penerbitan sertifikat kompetensi. Selanjutnya dilihat dari sebaran butir-butir instrumen pengakuan legal asosiasi profesi dan DU/DI menjelaskan adanya tujuh butir pernyataan atau pertanyaan untuk instrumen pengakuan legal asosiasi profesi dan DU/DI dan masuk dalam faktor 1 dan 2 dengan *total varian explained* sebesar 65,724%. Hal ini menunjukkan kedua faktor yang dikembangkan mampu mengukur variabel pengakuan legal asosiasi profesi dan DU/DI sebesar 65,724%. Hasil pengujian validitas dan reliabilitas menunjukkan bahwa semua butir instrumen memiliki nilai validitas  $>0,3$  dan memenuhi kriteria *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* dengan muatan sebesar 0,747 ( $>0,5$ ). Untuk reliabilitas pada aspek 1 mempunyai  $\alpha=0,876$  dan aspek 2  $\alpha=0,793$  ( $>0,7$ ) memberi indikasi bahwa instrumen reliabel.

## Pembahasan

Hasil analisis secara deskriptif terhadap tingkat keterbacaan instrumen menunjukkan bahwa pada uji coba tahap ke tiga rerata total mencapai 3,92. Hal ini memberikan indikasi bahwa tingkat keterbacaan instrumen evaluasi UKK berada pada predikat baik. Dengan demikian, validasi melalui proses pemeriksaan oleh pakar telah mengantarkan perumusan butir-butir instrumen sesuai dengan indikatornya. Mardapi (2005, pp. 15-20) mengatakan bahwa proses validasi, baik validasi teoritik maupun validasi empirik merupakan langkah yang penting. Validasi teoritik yaitu melalui pemeriksaan pakar untuk menelaah seberapa jauh dimensi merupakan jabaran yang tepat dari konstruk, seberapa jauh indikator merupakan jabaran yang tepat dari dimensi, dan seberapa jauh butir-butir instrumen yang dibuat secara tepat dapat mengukur indikator.

Sistem indikator dirancang untuk memberikan informasi akurat tentang berbagai kondisi dan memberi informasi bagaimana komponen-komponen yang akan dievaluasi

menghasilkan efek secara keseluruhan. Konsep ini sejalan dengan pendapat Shavelson (2001, p.7) bahwa merancang indikator dapat dilakukan dengan langkah-langkah: *conceptualize potential indicator, refine indicator poor, design alternative indicator system options, evaluate the options and begin developing or refining individual indicator*.

Hasil pengujian instrumen kolaborasi sekolah dengan asosiasi profesi dan DU/DI, menunjukkan bahwa berdasarkan data uji coba pada tahap implementasi, dari delapan butir instrumen seluruhnya valid karena memiliki nilai muatan faktor di atas 0,3 dan nilai  $t$  lebih besar dari 1,96. Kedua aspek instrumen memiliki koefisien Alpha di atas 0,7 sehingga instrumen pada ke dua aspek tersebut reliabel. Nilai  $p$ -value=0,0781 ( $>0,05$ ), nilai  $RMSEA=0,044$  ( $<0,80$ ) dan nilai  $GFI=0,97$  ( $>0,90$ ) memenuhi syarat sebagai fit model. Hasil pengujian memberikan indikasi bahwa semua persyaratan sebagai model yang fit terpenuhi, dengan demikian instrumen pengukuran tersebut sudah dapat digunakan sebagai model pengukuran yang cocok untuk mengumpulkan data tentang kolaborasi sekolah dengan asosiasi profesi dan DU/DI.

Kolaborasi atau kerja sama penyelenggara UKK pada satuan pendidikan dengan dunia usaha/industri/asosiasi profesi/ institusi mitra, pada hakikatnya pihak sekolah ingin memperoleh informasi dari dunia kerja mengenai standar kompetensi kerja yang dibutuhkan oleh pasar kerja. Berdasarkan standar kompetensi tersebut, maka sekolah menyusun kurikulum yang relevan. Selanjutnya, kurikulum diimplementasikan dalam proses pembelajaran. Untuk mengetahui keberhasilan pembelajaran, dilakukan uji kompetensi. Dengan demikian, materi uji kompetensi juga harus sesuai dengan cakupan SKL serta materi pembelajaran.

Hasil pengujian instrumen kinerja asesor, menunjukkan bahwa dari 25 butir instrumen kinerja asesor seluruhnya valid karena memiliki nilai muatan faktor di atas 0,3 dan nilai  $t$  lebih besar dari 1,96. Kriteria kecocokan model pengukuran hipotetik dengan data empirik, menunjukkan nilai

$RMSEA=0,032$  ( $<0,80$ ) memenuhi syarat sebagai fit model, dan nilai  $GFI=0,91$  ( $>0,90$ ) juga telah memenuhi syarat sebagai fit model, dengan demikian instrumen pengukuran kinerja asesor tersebut sudah dapat digunakan sebagai model pengukuran yang cocok untuk mengumpulkan data tentang kinerja asesor UKK.

Kualifikasi asesor sesuai dengan tujuan UKK juga perlu diperhatikan secara sungguh-sungguh. Asesor dapat berasal dari asosiasi profesi, maupun dari pengguna lulusan, yaitu DU/DI. Hal ini disebabkan, secara rasional, pihak-pihak yang paling mengetahui perkembangan kompetensi terkini yang dibutuhkan oleh dunia kerja adalah para praktisi di lapangan, yaitu asosiasi profesi dan DU/DI

Hasil pengujian instrumen sarana prasarana penunjang UKK mengindikasikan dari 21 butir instrumen seluruhnya valid karena memiliki nilai muatan faktor  $>0,3$  dan nilai  $t > 1,96$ . Hasil pengujian menunjukkan nilai  $RMSEA = 0,043$  ( $<0,80$ ) dan nilai  $GFI=0,91$  ( $>0,90$ ), artinya dua dari tiga kriteria sebagai model yang fit terpenuhi, maka instrumen tersebut sudah dapat digunakan sebagai model pengukuran yang cocok untuk mengumpulkan data tentang sarana dan prasarana penunjang UKK.

Hasil pengujian instrumen sikap siswa terhadap UKK, menunjukkan bahwa semua butir instrumen sikap siswa terhadap UKK memiliki muatan faktor di atas 0,3 dan nilai  $t$  lebih besar dari 1,96 sehingga semua butir instrumen dalam kategori valid. Nilai  $RMSEA=0,052$  ( $<0,80$ ) dan nilai  $GFI=0,91$  ( $>0,90$ ) telah memenuhi kriteria sebagai fit model.

Hasil pengujian instrumen ketersediaan informasi capaian kompetensi siswa, mengindikasikan dari 9 butir instrumen informasi capaian kompetensi siswa seluruhnya valid karena memiliki nilai muatan faktor di atas 0,3 dan nilai  $t$  lebih besar dari 1,96. Nilai  $p$ -value = 0,06516 ( $>0,05$ ), nilai  $RMSEA=0,042$  ( $<0,80$ ), dan nilai  $GFI=0,97$  ( $>0,90$ ) memenuhi tiga kriteria sebagai fit model.

Hasil pengujian instrumen pengakuan legal asosiasi profesi dan DU/DI memperlihatkan, dari 8 butir instrumen seluruhnya valid karena memiliki nilai muatan faktor di atas 0,3 dan nilai  $t$  lebih besar dari 1,96. Nilai  $RMSEA=0,054$  ( $<0,80$ ) memenuhi syarat sebagai fit model, dan nilai  $GFI = 0,97$  ( $> 0,90$ ) juga telah memenuhi syarat sebagai fit model. Oleh karena dua kriteria sebagai model yang fit terpenuhi, maka instrumen tersebut sudah dapat digunakan sebagai model pengukuran yang cocok untuk mengumpulkan data tentang pengakuan legal pihak eksternal, yaitu asosiasi profesi atau DU/DI.

Hasil analisis statistik untuk menguji validitas dan reliabilitas instrumen telah memenuhi kriteria sebagai model instrumen yang valid dan reliabel. Kriteria validitas butir instrumen sesuai dengan pendapat Hair (2010, p. 119) "*factor loadings  $\pm 0,3$  to  $0,4$  are minimally acceptable*". Pendapat senada dikemukakan Tabachnick (2007, p. 649) "*as rule of thumb, only variables with factor loadings of .32 and above are interpreted*". Hasil analisis validitas ini memberi makna penting, bahwa semua butir instrumen mampu mengukur apa yang seharusnya diukur.

Hasil-hasil statistik sebagai kriteria *goodness of fit* yang digunakan untuk menilai kecocokan model pengukuran instrumen dengan responden siswa yaitu *p-value*, *Root Mean Square Error of Approximation (RMSEA)* dan *Goodness of Fit Index (GFI)* menunjukkan kecocokan yang cukup baik antara model teoretis dengan data lapangan ketika instrumen diuji cobakan. Kecocokan model pengukuran tersebut ditentukan oleh sekurang-kurangnya terpenuhi dua dari tiga kriteria tersebut, sehingga instrumen tersebut sudah dapat digunakan sebagai model pengukuran yang cocok untuk mengumpulkan data tentang penyelenggaraan UKK Administrasi Perkantoran di SMK.

Hasil analisis fit model pada tahap ketiga (implementasi) dibandingkan dengan analisis uji coba tahap kedua, untuk nilai *p-value* mengalami peningkatan dari 0,0024 menjadi 0,078. Peningkatan tersebut memiliki makna penting, karena hasil penilaian

tersebut telah mengubah klasifikasi yang semula terdapat dua kriteria fit, kemudian menjadi terpenuhi tiga kriteria sebagai model yang cocok (*fit model*). Peningkatan tersebut terjadi karena adanya perbedaan jumlah dan karakteristik responden. Pada uji coba kedua responden berjumlah 196 siswa, sedangkan uji coba tahap ketiga responden berjumlah 256 siswa.

Instrumen dengan responden guru dianalisis menggunakan *Exploratory Factor Analysis (EFA)* dengan bantuan program *SPSS 17,00 for windows*. Analisis *EFA* ini untuk menguji apakah suatu konstruk dapat dijelaskan oleh indikator-indikatornya. Apabila indikator dapat membentuk konstruk atau variabel, maka ditunjukkan dengan muatan faktor yang tinggi ( $>0,3$ ) berarti pengukuran sudah sesuai data, dan diharapkan nilai *Kaiser-Meyer-Olkin Measure of Sampling (KMO)* lebih besar dari 0,5. Sedangkan untuk memeriksa reliabilitas sebuah instrumen adalah dengan melihat nilai koefisien *Alpha* dari formula *Cronbach's Alpha*, yaitu sekurang-kurangnya 0,7 sebagai batas terendah.

Semua butir instrumen kolaborasi sekolah dengan asosiasi profesi dan DU/DI memiliki nilai validitas  $> 0,3$  (kecuali butir A5) dan memenuhi kriteria *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* dengan muatan sebesar 0,670 ( $>0,5$ ). Untuk reliabilitas pada aspek 1 mempunyai  $\alpha=0,915$  dan aspek 2  $\alpha=0,894$  ( $>0,7$ ).

Instrumen kinerja asesor dapat diketahui bahwa semua butir instrumen memiliki nilai validitas  $> 0,3$  dan memenuhi kriteria *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* dengan muatan sebesar 0,689 ( $> 0,5$ ). Untuk reliabilitas pada aspek 1 mempunyai  $\alpha=0,885$ , aspek 2  $\alpha = 0,953$ , aspek 3  $\alpha = 0,837$ , dan aspek 4  $\alpha=0,847$  ( $>0,7$ ).

Instrumen sarana dan prasarana penunjang UKK, dapat diketahui bahwa semua butir instrumen memiliki nilai validitas  $> 0,3$  dan memenuhi kriteria *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* dengan muatan sebesar 0,810 ( $>0,5$ ). Untuk reliabilitas pada aspek 1 mempunyai  $\alpha=0,943$ , aspek 2  $\alpha=0,869$ , dan aspek 3  $\alpha=0,916$  ( $>0,7$ ).

Untuk instrumen informasi capaian kompetensi siswa, setelah dianalisis dapat diketahui bahwa semua butir instrumen memiliki nilai validitas  $> 0,3$  dan memenuhi kriteria *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* dengan muatan sebesar  $0,744 (> 0,5)$ . Untuk reliabilitas pada aspek 1 mempunyai  $\alpha=0,902$  dan aspek 2  $\alpha=0,753 (>0,7)$ .

Hasil analisis instrumen pengakuan legal asosiasi profesi atau DU/DI, diketahui bahwa semua butir instrumen memiliki nilai validitas  $> 0,3$  dan memenuhi kriteria *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* dengan muatan sebesar  $0,747 (>0,5)$ . Untuk reliabilitas pada aspek 1 mempunyai  $\alpha=0,876$  dan aspek 2  $\alpha=0,793 (> 0,7)$  memberi indikasi bahwa instrumen reliabel.

Hasil pengujian menunjukkan bahwa muatan faktor untuk setiap butir instrumen cukup tinggi ( $> 0,3$ ) dan nilai *Kaiser-Meyer-Olkin Measure of Sampling (KMO)* lebih besar dari  $0,5$ , hal ini memberi indikasi bahwa indikator dapat membentuk konstruk atau variabel, berarti pengukuran sudah sesuai data. Selanjutnya hasil pengujian reliabilitas instrumen memperlihatkan nilai koefisien *Alpha* dari formula *Cronbach's Alpha* memenuhi sekurang-kurangnya  $0,7$  sebagai batas terendah, sehingga instrumen tersebut reliabel.

Hasil uji kuantitatif, baik untuk instrumen siswa yang dianalisis dengan *second order confirmatory factor analysis (CFA)* maupun instrumen guru yang dianalisis dengan *EFA* menunjukkan bahwa butir-butir instrumen yang dikembangkan merupakan butir yang valid dan instrumen tersebut reliabel. Selain itu, instrumen evaluasi IE-UKK merupakan model instrumen evaluasi yang sesuai atau cocok untuk mengevaluasi kualitas penyelenggaraan UKK Administrasi Perkantoran di SMK, karena model tersebut secara statistik didukung oleh data lapangan, baik dari segi model strukturalnya maupun dari segi model pengukurannya.

Didukung adanya panduan evaluasi yang cukup singkat tetapi lengkap akan mempermudah penerapan instrumen evaluasi IE-UKK. Sanders & Sullins (2006, p.1) menjelaskan, "*program evaluation is the process of systematically determining the quality of a*

*program and how it can be improved*". Evaluasi program adalah proses sistematis menentukan kualitas program dan upaya meningkatkan kualitas program tersebut.

## Simpulan

Simpulan hasil penelitian ini adalah: (1) komponen penyelenggaraan UKK mencakup: (a) kolaborasi sekolah dengan asosiasi profesi dan DU/DI; (b) kinerja asesor; (c) sarana prasarana penunjang UKK; (d) sikap siswa terhadap UKK; (e) informasi capaian kompetensi siswa; dan (f) pengakuan legal asosiasi profesi dan DU/DI; (2) produk yang dikembangkan mencakup: instrumen evaluasi dan panduan penerapan instrumen. Sedangkan jenis instrumen evaluasi mencakup: (a) instrumen untuk responden siswa, dan (b) instrumen untuk responden guru; (3) hasil uji keterbacaan instrumen menunjukkan instrumen dalam kategori baik atau layak, rerata skor  $3,92$ ; (4) hasil analisis instrumen dengan responden siswa berdasarkan data uji coba pada tahap implementasi, seluruh instrumen valid (muatan faktor  $>0,3$  dan nilai  $t > 1,96$ ), reliabel ( $\alpha > 0,7$ ), dan memenuhi syarat sebagai model yang fit ( $RMSEA < 0,08$  dan  $GFI > 0,90$ ); (5) hasil analisis instrumen dengan responden guru pada tahap implementasi menunjukkan seluruh butir instrumen memiliki nilai validitas  $>0,3$  dan memenuhi kriteria  $KMO > 0,5$  serta koefisien reliabilitas  $\alpha > 0,7$ ; (6) hasil uji panduan evaluasi berdasarkan penilaian pakar dan praktisi dianalisis secara diskriptif, mencapai rerata skor  $3,99$  berarti termasuk dalam klasifikasi layak digunakan; (7) instrumen evaluasi uji kompetensi keahlian yang dikembangkan memenuhi syarat validitas dan reliabilitas sehingga layak digunakan oleh pimpinan sekolah untuk mengevaluasi kualitas penyelenggaraan UKK Administrasi Perkantoran di SMK

## Saran

Berdasarkan hasil penelitian ini dapat diajukan beberapa saran sebagai berikut: (1) instrumen yang dikembangkan dalam model ini hanya untuk responden siswa dan guru, belum melibatkan penilaian oleh pihak eks-

ternal seperti asesor dan komite sekolah. Dengan demikian, disarankan kepada peneliti lain untuk dapat mengembangkannya lebih jauh, sehingga instrumen evaluasi IE-UKK akan menjadi lebih baik; (2) cakupan objek penilaian dalam instrumen evaluasi IE-UKK, belum menjangkau komponen *outcome* yaitu kinerja para pemegang sertifikat kompetensi di dunia kerja. Oleh karena, itu disarankan kepada peneliti lain untuk mengembangkan instrumen evaluasi yang cakupannya sampai kepada komponen *outcome* UKK; (3) untuk meningkatkan efisiensi penerapan instrumen evaluasi IE-UK, pada produk pengembangan ini telah disusun *soft file* dalam *Microsoft Office Excel*, seorang evaluator disarankan memanfaatkan *soft file* tersebut sehingga menjadi lebih mudah dalam mengentri dan menganalisis data, serta menyusun rekomendasi hasil evaluasi dalam rangka perbaikan penyelenggaraan UKK ke depan.

#### Daftar Pustaka

- Cox, J. (2006). *The quality of an instructional program*. National Education Association-Alaska. Diambil tanggal 23 Februari 2009, dari <http://www.ak.nea.org/excellence/coxquality>
- Crawford, D.C. (2006). "Suggestions to Assess Nonformal Education Programs" dalam *ProQuest Education Journals*.
- Cullingford, C. (2000). *Assessment versus evaluation*. London. Cassell.
- Direktorat Pembinaan SMK. (2009). *Surat Edaran Direktorat Jenderal Manajemen Pendidikan Dasar dan Menengah Nomor: 351/C.5/MN/2007 tentang Uji Kompetensi Keahlian Kejuruan*. Diambil tanggal 22 Mei 2008 dari <http://209.85.175.104/ditpsmk.net/2008>.
- \_\_\_\_\_. (2011). *Petunjuk teknis pelaksanaan ujian kompetensi keahlian SMK tahun pelajaran 2010/2011*. Jakarta: Ditjen Pendidikan Menengah.
- Djojonegoro, W. (1998). *Pengembangan SDM melalui SMK*. Jakarta: PT Jayakarta Agung Offset.
- Fitzpatrick, J.L., Sanders, J.R., & Worthen, B.R. (2011). *Program evaluation alternative approaches and practical guidelines*. Boston: Pearson.
- Hair, J.F. et.al. (2010). *Multivariate data analysis 7<sup>th</sup> edition*. New York: Pearson Prentice Hall
- Ghozali, Imam. (2005). *Structural equation modeling, Teori, konsep dan aplikasi dengan program Lisrel 8,80*. Semarang: Badan Penerbit Universitas Diponegoro.
- Johnson, B. & Christensen, L. (2008). *Educational research quantitative, qualitative, and mixed approaches*. Los Angeles: SAGE Publications.
- Mardapi, Djemari. (2005). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendikia Offset.
- Mizikaci, F. (2007). A systems approach to program evaluation model for quality in higher education. *ProQuest Education Journals*. 130 (125-140).
- Nurosis. (1986). *SPSS/PC+for the imbbc/xt/at*. Chicago: SPSS Inc.
- Royse, D., Thyer, B. & Padgett, D., (2010). *Program evaluation in introduction*. Belmont: Wadsworth.
- Sanders, J.R. & Sullins, C.D. (2006). *Evaluation school programs an educator's guide*. California: Corwin Press.
- Shavelson, R.J. (2001). Steps in designing an indicator system (versi elektronik). *Practical assessment, research & evaluation*. Diambil tanggal 14 April 2013 dari <http://PAREonline.net/getvn.asp?v=2&n>.
- Sigalingging, Karmon. (2009). *Model evaluasi kreativitas lomba kompetensi siswa SMK*. Disertasi Doktor, tidak diterbitkan, Universitas Negeri Yogyakarta.

## ANALISIS METODE *CHEATING* PADA TES BERSKALA BESAR

<sup>1)</sup>Yance Manoppo, <sup>2)</sup>Djemari Mardapi

<sup>1)</sup>Universitas Pattimura, <sup>2)</sup>Universitas Negeri Yogyakarta

<sup>1)</sup>molucanano@yahoo.com, <sup>2)</sup>djemarimardapi@gmail.com

### Abstrak

Penelitian ini bertujuan untuk mengetahui: (1) karakteristik butir soal Kimia Ujian Nasional berdasarkan teori tes klasik dan teori respon butir; (2) besarnya kecurangan yang terjadi dengan menggunakan Metode *Angoff's B-index*, Metode *Pair1*, Metode *Pair2*, Metode *Modified Error Similarity Analysis (MESA)* dan Metode *G2*; (3) metode yang lebih banyak mendeteksi adanya kecurangan dalam pelaksanaan UN Kimia tingkat SMA/MA Negeri tahun pelajaran 2011/2012 di Provinsi Maluku. Hasil analisis dengan pendekatan teori tes klasik menunjukkan 77,5% butir memiliki tingkat kesulitan butir berfungsi baik, 55% butir daya bedanya belum memenuhi syarat, dan 70% butir memiliki pengecoh berfungsi baik dengan indeks reliabilitas tes 0,772. Analisis dengan pendekatan teori respons butir menunjukkan 14 (35%) butir cocok dengan model, fungsi informasi maksimum 11,4069 pada  $\theta = -1,6$ , dan besarnya kesalahan pengukuran 2,296. Jumlah pasangan yang diduga curang adalah: menurut Metode *Angoff's B-index* ada 13 pasangan, menurut Metode *Pair1* ada 212 pasangan, menurut Metode *Pair2* ada 444 pasangan, menurut Metode *MESA* ada 7 pasangan, dan menurut Metode *G2* ada 102 pasangan. Metode yang paling banyak mendeteksi kecurangan secara berturut-turut adalah: Metode *Pair2*, Metode *Pair1*, Metode *G2*, Metode *Angoff's B-index*, dan Metode *MESA*.

**Kata kunci:** *ujian nasional, karakteristik butir, metode kecurangan*

## AN ANALYSIS OF METHOD OF CHEATING ON LARGE TEST SCALE

<sup>1)</sup>Yance Manoppo, <sup>2)</sup>Djemari Mardapi

<sup>1)</sup>Universitas Pattimura, <sup>2)</sup>Universitas Negeri Yogyakarta

<sup>1)</sup>molucanano@yahoo.com, <sup>2)</sup>djemarimardapi@gmail.com

### Abstract

This study aimed to reveal: (1) the characteristics of items of Chemistry Test in National Examination by using the classical test theory and item response theory; (2) the amount of cheating which occurred by using Angoff's B-index Method, Pair 1 Method, Pair 2 Method, Modified Error Similarity Analysis (MESA) Method, and G2 Method; (3) the methods that detected more cheating in the implementation of the Chemistry Test in National Examination for high schools in the academic year 2011/2012 in Maluku Province. The results of the analysis with the classical test theory approach show that 77.5% items have item difficulty functioning well, 55% items have discrimination that has not met the requirement yet, and 70% items have distractor that works well with the index reliability test of 0,772. The analysis using the item response theory approach shows that 14 (35%) items fit with the model, the maximum function information is 11,4069 at  $\theta = -1,6$ , and the magnitude of the error of measurement is 2,296. The number of pairs who are suspected of cheating is as follows: 13 pairs according to Angoff's B-index Method, 212 pairs according to Pair 1 Method, 444 pairs according to Pair 2 Method, 7 pairs according to MESA Method, and 102 pairs according to G2 Method. The most widely detecting cheating in a row is a Pair 2 Method, Pair 1 Method, G2 Method, Angoff's B-index Method, and MESA Method.

**Keywords:** *national examination, items characteristics, methods of cheating*

## Pendahuluan

Pendidikan nasional bertujuan untuk mengembangkan potensi peserta didik agar menjadi manusia yang beriman dan bertakwa kepada Tuhan Yang Maha Esa, berakhlak mulia, sehat, berilmu, cakap, kreatif, mandiri, dan menjadi warga negara yang demokratis serta bertanggung jawab. Tujuan tersebut tercantum dalam Undang-Undang Nomor 20 Tahun 2003 Bab 2 pasal 3 yang ditetapkan oleh Pemerintah Indonesia melalui Direktorat Jenderal Pendidikan Dasar dan Menengah Departemen Pendidikan dan Kebudayaan yang kemudian ditegaskan kembali dalam Peraturan Pemerintah Republik Indonesia Nomor 19 Tahun 2005 Bab 2 pasal 4, mengenai tujuan standar pendidikan nasional.

Ujian Nasional (UN) merupakan sistem evaluasi standar pendidikan dasar dan menengah secara nasional. UN dilakukan oleh Depdiknas (Undang-Undang Republik Indonesia Nomor 20 Tahun 2003). Evaluasi UN dilakukan oleh lembaga yang mandiri secara berkala, menyeluruh, transparan, dan sistematis untuk menilai pencapaian standar nasional pendidikan. Proses pemantauan evaluasi UN harus dilakukan secara berkesinambungan. Pelaksanaan UN diharapkan dapat dipetakan tingkat kemampuan sekolah sehingga dapat menentukan skala prioritas penanganan proses pendidikan. Kejujuran dan prestasi menjadi kata kunci dalam pelaksanaan UN. Untuk mengukur objektivitas, validitas, dan reliabilitas mutu hasil UN, dapat ditelusuri dengan mengembangkan metode analisis pola jawaban peserta ujian. Analisis pola jawaban dapat mengidentifikasi adanya intervensi dari luar atau ketidakjujuran peserta ujian dalam menjawab soal sewaktu penyelenggaraan UN. Salah satu teknik untuk mendeteksi kejujuran dalam pelaksanaan UN yaitu dengan melihat pola jawaban peserta ujian yang bersumber dari Badan Standar Nasional Pendidikan (BSNP).

Kenyataan bahwa sistem pendidikan Indonesia yang menggunakan nilai dari tes atau evaluasi belajar terhadap materi yang diberikan sebelumnya untuk menunjukkan

kemajuan dan penguasaan ilmu anak didik, menyebabkan masyarakat memandang prestasi belajar hanya dari pencapaian nilai yang tinggi, bukan pada prosesnya. Pandangan tersebut menimbulkan tekanan pada siswa untuk mencapai nilai yang tinggi. Tekanan yang dirasakan akan membuat siswa lebih berorientasi pada perolehan nilai, bukan pada pemahaman ilmu. Siswa dapat mempersepsi ujian sebagai alat untuk menyusun peringkat dan dapat menyebabkan dirinya mengalami kegagalan, bukan sebagai instrumen yang dapat menunjukkan kemajuan dalam proses belajar.

*Cheating* (ketidakjujuran) pada tes biasanya sering terjadi di semua tingkat dan situasi tes, dari mulai tingkat kelas (Lim & Sean, 2001; Einsenberg, 2004; William, 2002; Anderman, Griesinger & Westerfield, 1998) sampai *high school* atau *academic institution* (Anderman, Cupp & Lane, 2010; Storm & Storm, 2007; Lama, 2008; Chula, Guyette & Piotrowski, 2009; Williams, Craig & Paulhus, 2010; McCabe, Trevino & Butterfield, 2001). *Cheating* berakibat buruk pada rasa keadilan dan kejujuran dari peserta tes serta dapat merusak tujuan tes tersebut di lakukan. Untuk *high-stake assessments* akibat dari ketidakjujuran ini sangat besar karena akan menurunkan tingkat kepercayaan masyarakat terhadap penyelenggaraan tes itu sendiri.

Perilaku *cheating* banyak diakibatkan oleh pengaruh kelompok yang orang cenderung berani melakukan karena melihat orang lain di kelompoknya juga melakukan. Apabila kecenderungan ini berlangsung secara terus-menerus, maka *cheating* akan menjadi kebiasaan seseorang, yang akan di-transfer tidak hanya pada kegiatan sekolah lainnya tetapi kepada kegiatan kemasyarakatan pada umumnya berdasarkan prinsip *transfer of learning*. Pada akhirnya, *cheating* akan menimbulkan kehancuran di masa mendatang. Hal ini sesuai dengan apa yang dinyatakan Cizek (2001, pp.3-4) sebagai berikut:

*There are many forms of cheating that can occur on such tests. In general, cheating can be defined as any action that violates the rules for*

*administering a test, any behavior that gives an examinee an unfair advantage over other examinees, or any action on the part of an examinee or test administrator that decreases the accuracy of the intended inferences arising from the examinee's test score or performance.*

Berdasarkan pendapat Cizek tersebut, diperoleh bahwa kecurangan secara umum dapat didefinisikan sebagai perilaku yang menyalahi aturan dalam penyelenggaraan tes, perilaku apa pun yang memberikan peserta ujian keuntungan yang tidak wajar atas peserta ujian lainnya, atau tindakan lain dari penyelenggara ujian yang mengurangi keakuratan hasil dimaksud yang berasal dari skor tes atau kinerja peserta ujian.

Sementara Lewkowicz (2007, p.147) mengatakan bahwa :

*Cheating, defined as "to full by trickery," is seen as behavior that seems to be based, at least partly, on several assumptions, including, "I don't have to work hard at school. I'll be lucky and win the lottery or become famous," and "I will never succeed so why bother to try," and "It is awful, and I must no accept myself until I measure up to everyone else.*

Pendapat Lewkowicz menjelaskan bahwa kecurangan didefinisikan sebagai sesuatu yang penuh dengan tipu daya yang mengajarkan siswa tidak perlu belajar keras di sekolah atau siswa menganggap dirinya tidak akan berhasil jadi mengapa harus bersusah payah mencoba. Kondisi ini dapat menghilangkan rasa kepercayaan diri dari siswa untuk sama seperti teman-teman lainnya.

Permasalahan *cheating* merupakan masalah yang dihadapi oleh semua negara. *cheating* pada akhirnya menjadi perhatian internasional. *Cheating* sudah tidak asing lagi bagi pelajar dan mahasiswa (Bogle, 2000; Milliron & Sandoe, 2008). Setiap orang pasti ingin mendapat nilai yang baik dalam ujian, dan sudah tentu berbagai macam cara dilakukan untuk mencapai tujuan itu. Masalah *cheating* selalu terkait dengan tes atau ujian.

Penelitian sebelumnya tentang *cheating* sering terfokus pada deskriptif isu-isu seperti gender dan perbedaan budaya. Sebagai contoh, Newstead *et al.* (Anderman, Griesinger & Westerfield, 1998, pp.84-85) melapor-

kan bahwa di antara mahasiswa, pria melaporkan kecurangan lebih daripada wanita, siswa yang lebih muda melaporkan kecurangan lebih daripada siswa yang lebih tua, dan siswa kemampuan rendah melaporkan kecurangan lebih daripada kemampuan siswa yang lebih tinggi. Dalam internasional studi tentang kecurangan, Evans, Craig & Mietzel (Anderman, Griesinger & Westerfield, 1998, p.85) menemukan bahwa siswa dari berbagai negara ditafsirkan memiliki kecurangan dalam cara yang berbeda.

Perilaku *cheating* tidak hanya terjadi pada siswa di SMP atau SMA tetapi terjadi pula di bangku kuliah atau universitas. Storm & Storm (2007, pp.105-106) menyatakan beberapa data yang memprihatinkan adalah survei nasional yang dilakukan oleh *Josephson Institute of ethics* di Amerika pada tahun 2006 dengan responden 36.000 siswa Sekolah Menengah Pertama menemukan 60% siswa menerima dan mengakui pernah melakukan kecurangan pada saat ujian dan pengerjaan tugas. Terjadi peningkatan sebesar 1% dalam kurun waktu 2 tahun, 95% di antaranya mengaku bahwa tidak pernah ketahuan ketika melakukan kecurangan. permasalahan ini dalam berbagai kajian dan penelitian perlu untuk segera mendapatkan penanganan.

Temuan hasil penelitian lain yang menarik adalah dari *The Epoch Time* 2005 yang mengambil data dari 90 mahasiswa. Dari jumlah tersebut, 83% mengaku pernah *cheating* ketika pelaksanaan tes atau ujian. (Strom & Strom, 2007, p.105). Selain itu, di China akhirnya diterapkan adanya sanksi bagi mahasiswa yang *cheating* akan dihukum dengan 7 tahun penjara. Menurut Callahan (Strom & Storm, 2007, p.105), perilaku *cheating* juga ditemukan pada siswa di Australia, Inggris, India, Jepang, Korea, Spanyol, dan Skotlandia. Anderman, Griesinger & Westerfield (1998: 84) menjelaskan bahwa menyontek merupakan hal yang biasa di kalangan remaja karena siswa sekolah lanjutan lebih berfokus pada peringkat dan performa dibandingkan dengan siswa sekolah dasar. Brades (Cizek, 1999, p.15) menemukan bahwa 65% dari siswa sekolah

dasar mengakui menyalin setidaknya sekali dari siswa lain selama dites.

Sementara itu Widiatmo (2009, p.219) menjelaskan bahwa untuk mengurangi kemungkinan peserta uji melakukan kecurangan, banyak usaha preventif yang dilakukan untuk mengurangi kesempatan peserta tes dalam melakukan *cheating*. Usaha-usaha tersebut antara lain dengan menambah *proctors* (pengawas tes) sehingga peserta uji dapat diawasi, dengan membagikan lebih dari satu paket tes, dan memperlebar jarak tempat duduk antarpeserta tes. Sebagai ilustrasi, *American College Testing (ACT)*, sebuah lembaga *testing* ternama di *United State of America*, membagikan secara random 8 sampai 12 paket tes setiap kali tes dilaksanakan. Jika ada 30 peserta uji dalam satu ruangan dan ada 10 paket tes yang dibagikan, maka kemungkinan ada 3 orang yang akan memperoleh paket tes yang sama. Lebih lanjut, kemungkinan dua peserta uji yang duduk berdekatan memperoleh paket tes yang sama sangat kecil 0,01 (satu berbanding 100). Dengan usaha-usaha preventif seperti itu kemungkinan peserta uji melakukan kecurangan sangat kecil. Kenyataan lapangan membuktikan usaha-usaha preventif tersebut tidak efektif.

Faktor kemajuan teknologi juga memberikan kontribusi terhadap *cheating* pada jalannya UN di sekolah-sekolah. Menurut Aziz (2012, p.3) yang dimuat pada *Tempo Online* soal UN 2012 untuk Sekolah Menengah Pertama (SMP) di Kota Sukabumi, Jawa Barat, diduga bocor. Hal itu ditandai dengan beredarnya kunci jawaban di kalangan siswa melalui *short message service (SMS)* padahal pihak dinas pendidikan setempat menjamin UN di Kota Sukabumi tidak ada kebocoran. Dengan kecanggihan alat komunikasi, tentunya peristiwa yang sama dapat terjadi di setiap daerah yang telah terjangkau sarana komunikasi seluler ini.

Selain permasalahan bocornya tes, pencapaian kualitas soal sesuai dengan karakteristik butir tes yang baik sangat dipengaruhi sumber daya penyusun tes. Sebagaimana yang diketahui bersama, perangkat tes UN disusun dalam bentuk tes objektif de-

ngan menggunakan *multiple choice* atau bentuk pilihan berganda. Tes seperti ini membutuhkan pemahaman yang komprehensif dan mendalam dari peserta UN sehingga mampu memberikan jawaban yang benar pula. Dengan demikian, perangkat tes harus memiliki kehandalan, kesahihan, daya pembeda, distraktor untuk mendeteksi kemampuan siswa secara objektif. Adanya perangkat tes yang kurang baik menyebabkan praktik-praktik *cheating* (kecurangan) lebih mudah dilakukan oleh karena kesulitan siswa dalam menjawab soal dengan benar disamping adanya tekanan untuk harus lulus dalam Ujian Nasional.

Berdasarkan uraian di atas maka yang dimaksud dengan *cheating* dalam tulisan ini adalah segala perbuatan atau trik-trik yang tidak jujur, perilaku tidak terpuji atau perbuatan curang yang dilakukan oleh seseorang untuk mencapai keberhasilan dalam menyelesaikan tugas-tugas akademik terutama yang terkait dengan evaluasi/ujian hasil dengan mengabaikan aturan dan kesepakatan yang ada. Perilaku *cheating* terjadi karena masyarakat memiliki pandangan bahwa prestasi belajar tercermin dari pencapaian nilai yang tinggi, sehingga membuat siswa terpaksa untuk memperoleh nilai tinggi dengan cara apa pun. Masyarakat cenderung semakin permisif sehingga menyebabkan perilaku *cheating* semakin sulit dihilangkan. Fakta-fakta tersebut menunjukkan bahwa *cheating* merupakan suatu permasalahan yang menarik untuk dianalisis lebih lanjut. Untuk itu, melalui penelitian ini penulis akan meneliti beberapa metode yang dipakai untuk mendeteksi kemungkinan peserta ujian melakukan *cheating* atau sering diidentikkan dengan *collusion* dengan melihat pola respon jawaban dari peserta tes (*examinees*). Oleh karena itu, penelitian ini bertujuan untuk mengetahui: (1) karakteristik butir soal Kimia Ujian Nasional berdasarkan teori tes klasik dan teori respon butir; (2) besarnya kecurangan yang terjadi dengan menggunakan Metode *Angoff's B-index*, Metode *Pair1*, Metode *Pair2*, Metode *Modified Error Similarity Analysis (MESA)* dan Metode *G2*; (3) metode yang lebih banyak mendeteksi

adanya kecurangan dalam pelaksanaan UN Kimia tingkat SMA/MA Negeri tahun pelajaran 2011/2012 di Provinsi Maluku.

## Metode Penelitian

### Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif menggunakan pendekatan *ex-post facto*.

### Waktu dan Tempat Penelitian

Penelitian ini dilakukan di Lab Komputer Pascasarjana Universitas Negeri Yogyakarta setelah memperoleh data respon sampel (LJK) dari Puspendik Kemdikbud Jakarta. Kegiatan penelitian ini dilaksanakan pada Maret 2013.

### Target/Subjek Penelitian

Target dalam penelitian ini adalah perangkat tes UN Kimia tingkat SMA/MA Negeri tahun pelajaran 2011/2012 di Provinsi Maluku.

### Prosedur

1. Menganalisis butir tes dengan perangkat lunak *Itean* versi 3.0 untuk melihat karakteristik butir tes yang meliputi tingkat kesukaran butir, daya pembeda butir, dan statistik sebaran jawaban. Selain itu ditentukan juga kehandalan/reliabilitas tes, kesalahan pengukuran (*SE*) dan distribusi skor.
2. Menganalisis butir dengan pendekatan teori respons butir menggunakan program *Bilog-MG* 3.0. Dari analisis ini diperoleh parameter butir, parameter peserta, fungsi informasi tes, dan grafik informasi tes. Dari analisis ini diperoleh besarnya Fungsi Informasi maksimum tiap butir, dan skala kemampuan  $\theta$  yang diambil antara -4,0 dan 4,0 dengan interval 0,25.
3. Menaksir kecocokan data dengan model dan parameter butir dengan program *Bilog-MG* 3,0. Kecocokan butir dengan model dapat dilihat dari harga  $p$  yang muncul dari hasil analisis dengan program *Bilog-MG*. Untuk butir yang memiliki harga  $p$  lebih kecil dari 0,01 dianggap tidak cocok dengan model logistik 2-parameter.

4. Mendeteksi *cheating* dengan Metode *Angoff's B-index*, Metode *Pair 1*, Metode *Pair 2*, Metode *MESA*, dan Metode *G2* dengan *software Integrity*<sup>TM</sup>.
5. Menentukan keakuratan dari kelima metode deteksi *cheating* dengan melihat metode mana yang dapat mendeteksi *cheating* lebih banyak.

### Data, Instrumen, dan Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah respon peserta UN di Propinsi Maluku khususnya mata pelajaran kimia tingkat SMA/MA sekolah Negeri Paket soal C72 sebanyak 1620 siswa. Pengumpulan data dalam penelitian ini dilakukan dengan menggunakan teknik dokumentasi, yakni mengutip respon siswa (LJK) pada perangkat tes UN Kimia Paket C72 SMA/MA Negeri di Provinsi Maluku tingkat 2011/2012. Data-data berupa LJK diperoleh dari Puspendik Jakarta.

### Teknik Analisis Data

#### *Analisis Karakteristik Butir Soal*

Analisis terhadap butir soal secara kuantitatif (empiris) menggunakan pendekatan teori tes klasik dan teori respon butir (*IRT*).

1. Analisis Empiris dengan Teori Tes Klasik
  - a. Tingkat Kesukaran Butir Soal

Tingkat kesukaran butir dianalisis dengan menghitung proporsi menjawab benar ( $p$ ) sebuah butir soal. Apabila indeks  $p < 0,30$  maka butir soal termasuk sukar, apabila  $0,30 \leq p \leq 0,70$  maka butir soal memiliki tingkat kesukaran yang sedang, jika  $p > 0,70$  maka soal termasuk soal yang mudah. Akan tetapi oleh karena Ujian Nasional dikembangkan berdasarkan acuan kriteria (tidak membedakan) bukan acuan norma (dapat membedakan), maka besarnya tingkat kesukaran memberikan makna jika minimal 75% (KKM= 75) siswanya bias menjawab butir dengan benar maka siswa tersebut dikatakan sudah "*berhasil*" atau dengan kata lain guru telah berhasil dalam mengajar.

Menurut Shapiro (2011, p.17) acuan criteria juga dapat digunakan dalam mencari sejauhmana permasalahan yang dihadapi siswa dalam belajar sehingga dapat dicari jalan keluarnya.

b. Daya Pembeda Butir Soal

Daya beda butir soal (*item*) adalah korelasi antara skor butir soal dengan skor total, dihitung dengan rumus korelasi *point biserial*. Besarnya daya beda butir soal untuk menyatakan bahwa butir soal baik adalah minimal 0,3. Sedangkan item soal dengan *point biserial* di bawah 0,3 termasuk item yang kurang baik.

c. Keberfungsian Pengecoh

Untuk mengetahui suatu pengecoh berfungsi secara baik dapat dilihat dari distribusi jawaban. Suatu pengecoh dapat dikatakan berfungsi baik jika paling sedikit pilihan jawaban dipilih oleh 5% peserta tes. Beberapa literatur pada penelitian yang lain juga menyebutkan 2%, tetapi sebenarnya jika ada yang menjawab saja sudah bisa dikategorikan bahwa pengecohnya (*distractor*) sudah berfungsi dengan baik.

d. Indeks Reliabilitas.

Nilai indeks reliabilitas tes dapat dilihat pada nilai koefisien *alpha*. Semakin besar nilai *alpha* menunjukkan semakin *reliabel* tes tersebut dan semakin kecil tingkat kesalahan pengukuran. Standar yang digunakan dalam penelitian ini adalah yang mensyaratkan instrumen yang *reliabel* haruslah memiliki koefisien reliabilitas minimal 0,7.

2. Analisis Empiris Butir dengan Teori Respon Butir

Analisis dengan program *Bilog-MG* menghasilkan *output* dalam bentuk tiga fase. Fase pertama merupakan estimasi butir berdasar teori klasik, fase kedua estimasi parameter butir berdasar Teori Respons Butir dan fase ketiga estimasi kemampuan peserta.

Pada fase pertama diperoleh informasi tentang banyaknya peserta tes yang menjawab benar, proporsi peluang menjawab benar dibagi peluang menjawab

salah serta koefisien korelasi *biserial*. butir yang memiliki nilai koefisien *biserial* negatif dapat mengganggu proses analisis selanjutnya, sehingga butir tersebut tidak diikutkan dalam analisis berikutnya. Jumlah butir yang tidak diikutkan untuk tahap analisis selanjutnya terdiri dari 2 butir tes Kimia UN SMA/MA Paket C72 tahun pelajaran 2011/2012 Provinsi Maluku.

Fase kedua, estimasi parameter teori respon butir. Pada fase ini diperoleh tentang informasi parameter butir sesuai dengan model teori tespon butir yang digunakan. Selanjutnya, untuk model 1-PL didapatkan estimasi tingkat kesukaran, model 2-PL diperoleh estimasi parameter berupa tingkat kesukaran dan daya beda, sedangkan model 3-PL didapatkan informasi tentang tingkat kesukaran, daya beda, dan tebakan semu. Selain parameter butir, pada fase kedua juga dihasilkan statistik kecocokan suatu butir dengan model atau *goodness of fit statistic*. Model yang digunakan untuk estimasi parameter adalah model logistik yang banyak menerima butir yang cocok.

Secara empiris, kualitas butir ditelaah berdasarkan kecocokan data dengan model dan nilai parameter butir. Kecocokan suatu butir dengan model dapat dilihat dari nilai *chi* kuadrat butir dibandingkan dengan harga kritik distribusi *chi* kuadrat sesuai dengan *dk* butir yang bersangkutan pada taraf signifikansi *a*. Butir dikatakan tidak cocok model jika nilai  $\chi^2$  butir lebih besar dari harga distribusi  $\chi^2$  pada nilai kritisnya, sebaliknya butir cocok dengan model jika nilai  $\chi^2$  item lebih kecil atau sama dengan nilai distribusi  $\chi^2$ . Atau dikatakan cocok model jika probabilitas  $\chi^2$  lebih besar dari 0,01.

Berdasarkan perbandingan ketiga model parameter logistik untuk uji kecocokan model, maka model 2-PL pada umumnya lebih banyak menghasilkan butir yang *fit* (cocok). Untuk keseragaman model analisis, estimasi parameter butir untuk seluruh materi uji dianalisis menggunakan model 2-PL.

Fase ketiga didapatkan estimasi parameter kemampuan peserta dan fungsi informasi tes. Pada penelitian ini informasi yang digunakan hanya fungsi informasi tes, sehingga estimasi parameter kemampuan peserta tidak dianalisis. Besarnya fungsi informasi tes dihitung menggunakan program *Excel*.

Berdasarkan kriteria di atas, untuk menentukan kualitas butir yang baik dengan pendekatan teori respon butir didasarkan pada kriteria: (1) butir cocok model, (2) mempunyai daya beda 0 sampai 2, dan (3) tingkat kesulitan -2 hingga +2 (Hambleton & Swaminathan, 1985, p.36).

#### Analisis Cheating

Selanjutnya, untuk mendeteksi adanya *cheating* maka digunakan software *Integrity*<sup>TM</sup>. Software ini adalah sebuah aplikasi *interface online* yang aman dan dirancang untuk menganalisis data ujian pilihan ganda. Program ini mengevaluasi statistik kejujuran suatu tes dan kejujuran akademik siswa yang mengikuti tes. *Integrity* menggunakan *file* data yang dikirimkan oleh pengguna (*client*) untuk mendapatkan informasi statistik uji secara rinci, serta laporan statistik deteksi penyalinan jawaban secara rinci.

### Hasil Penelitian dan Pembahasan

Karakteristik Butir Tes Klasik dengan Program *Iteman*

#### Statistik Butir Soal

Pada tampilan hasil statistik butir soal dapat diketahui hasilnya untuk soal yang dikategorikan sukar atau tidak mencapai keberhasilan 4 butir, kategori kurang berhasil 5 butir, dan yang mencapai keberhasilan 31 butir. Soal yang memiliki daya beda yang baik 22 butir, tidak baik 18 butir sedangkan untuk pengecoh 28 butir yang tergolong baik.

#### Statistik Tes/Skala

Berdasarkan statistik hasil analisis program *Microcat Iteman* versi 3.0 terhadap 40 butir soal Kimia UN Paket C72 tingkat SMA/MA tahun pelajaran 2011/2012 yang direspon oleh peserta ujian sebanyak 1620 orang pada Tabel 1 dapat bahwa peserta

rata-rata dapat menjawab dengan benar sebesar 31 butir soal ( $mean = 30,981$ ), berarti rata-rata lebih dari setengah (75%) dari jumlah butir soal dapat dijawab dengan benar oleh peserta ujian, skor tertinggi 38 dan terendah 6. Nilai rerata (30,981) yang berdekatan dengan nilai median (32,00). Rata-rata tingkat kesukaran butir soal pada umumnya sedang ( $mean p = 0,775$ ), cenderung banyak pada kisaran indeks 0,30 sampai 0,70. Butir soal cukup mampu membedakan kelompok siswa berkemampuan tinggi dengan kelompok siswa berkemampuan rendah, hal tersebut dinyatakan oleh rerata indeks daya pembeda 0,331.

Indeks reliabilitas soal 0,772 dapat dinyatakan baik, artinya 77,2% perbedaan skor yang terdapat pada peserta ujian adalah kemampuan yang sebenarnya, sedangkan 22,8% perbedaan skor yang ada merupakan kesalahan pengukuran. Menurut Linn (Mardapi, 1999, p.14), kesalahan pengukuran bisa disebabkan oleh variasi acak dalam diri manusia, faktor lingkungan, subjektivitas pengukur, dan alat ukur. Kesalahan baku pengukuran pada perangkat soal hasil analisis secara klasik adalah sebesar 2,209.

Tabel 1. Karakteristik Tes Kimia UN Paket 72 Tahun Pelajaran 2011/2012 dengan Program *Iteman*

Karakteristik	Nilai
<i>N of Item</i>	40
<i>N of Examinees</i>	1620
<i>Mean</i>	30,981
<i>Variance</i>	21,441
<i>Std. Dev.</i>	4,630
<i>Skew</i>	-1,776
<i>Kurtosis</i>	4,593
<i>Minimum</i>	6,000
<i>Maximum</i>	38,000
<i>Median</i>	32,000
<i>Alpha</i>	0,772
<i>SEM</i>	2,209
<i>Mean P</i>	0,775
<i>Mean Item-Tot.</i>	0,331
<i>Mean Biserial</i>	0,527

Berdasarkan Tabel 1, rerata tingkat kesulitan dan daya pembeda soal yaitu 0,775 dan 0,331 artinya soal tes ini rata-rata dijawab dengan benar butirnya sebesar 77,5% atau dengan kata lain apabila dengan Kriteria Ketuntasan Minimal (KKM) sebesar 75 maka dapat ditafsirkan bawa tingkat keberhasilan guru dalam melaksanakan proses pembelajaran sudah berhasil, serta soal dapat membedakan siswa yang berkemampuan tinggi dan berkemampuan rendah sebesar 33,1%. Hasil analisis soal Kimia UN Paket C72 menunjukkan bahwa tes tersebut memiliki reliabilitas yang baik yaitu di atas 0,70. Kesalahan pengukuran dapat disebabkan oleh beberapa faktor, seperti lingkungan ketika ujian yang kurang mendukung, kondisi psikologis maupun biologis peserta tes atau terjadi kecurangan ketika pelaksanaan ujian seperti saling kerja sama antara peserta atau adanya bocoran kunci jawaban.

Berdasarkan analisis program *Iteman*, soal UN Kimia Paket C72 memiliki kualitas baik, karena 31 (77,5%) butir soal memenuhi kriteria dan 9 (22,5%) butir soal yang tidak memenuhi kriteria. Hal ini didukung juga dengan 70% pengecoh bekerja dengan baik meskipun dari kategori daya beda hanya 55%. Hasil temuan dari 18 butir tersebut terdapat 2 butir soal dengan tanda "check the key" dari *print out* program *Iteman*, yaitu butir soal nomor 21 dan 39. Tanda tersebut merupakan suatu peringatan bagi pembuat soal untuk melihat kembali pada pilihan jawaban apakah sudah tepat atau belum. Selanjutnya, dari 40 butir soal Kimia Paket soal C72 yang diujikan secara nasional tahun pelajaran 2011/2012 di Provinsi Maluku ditemukan 9 butir soal yang perlu diperiksa kembali sebelum digunakan. Dengan demikian, berdasarkan kriteria yang telah ditetapkan dalam penentuan penerimaan butir soal, maka butir soal yang tidak sesuai dengan teori tes klasik adalah sebanyak 5 (12,5%) butir perlu direvisi dan 4 (10%) butir perlu diganti.

#### Karakteristik Butir Soal berdasarkan Pendekatan *IRT*

Pada analisis fase pertama, nilai *biserial* digunakan untuk menentukan apakah suatu

butir diikutkan pada analisis berikutnya atau tidak. Fase pertama hasil *output Bilog-MG* menunjukkan bahwa dari 40 butir yang dianalisis terdapat 16 butir soal dengan kategori yang tidak baik atau sukar (butir soal nomor 3,21, 22, 33, 36, 39) sedangkan sisanya adalah berada pada kategori sedang dan mudah. Berdasarkan daya beda butir tes terdapat 16 butir soal dengan kategori tidak baik (butir soal nomor 3, 9, 11, 14, 18, 19, 21, 22, 26, 27, 28, 33, 36, 38, 39 40), dan 24 butir soal berkategori baik (butir soal nomor 1, 2, 4, 5, 6, 7, 8, 10, 12, 13, 15, 16, 17, 20, 23, 24, 25, 29, 30, 31, 32, 34, 35, 37). Dalam konteks UN yang beracuan kriteria (*criterion referenced test(CRT)*) butir soal dengan tingkat kesulitan mudah dapat digunakan.

Butir soal yang tidak diikutkan dalam fase pertama disajikan pada Tabel 2.

Tabel 2. Butir Hasil Analisis *Bilog* Fase Pertama yang Tidak Dapat Dilanjutkan ke Fase Kedua

No	<i>Correlation Biserial</i>	Keterangan
3	0,062	Tidak Dilanjutkan
9	0,208	Tidak Dilanjutkan
11	0,263	Tidak Dilanjutkan
14	0,268	Tidak Dilanjutkan
18	0,234	Tidak Dilanjutkan
19	0,281	Tidak Dilanjutkan
21	0,042	Tidak Dilanjutkan
22	0,08	Tidak Dilanjutkan
26	0,246	Tidak Dilanjutkan
27	0,271	Tidak Dilanjutkan
28	0,287	Tidak Dilanjutkan
33	-0,032	Tidak Dilanjutkan
36	0,12	Tidak Dilanjutkan
38	0,249	Tidak Dilanjutkan
39	-0,044	Tidak Dilanjutkan
40	0,2	Tidak Dilanjutkan

Suatu butir dengan *biserial* kurang dari 0,3 tidak diikutkan pada analisis berikutnya, karena akan mengganggu proses analisis

yakni munculnya nilai negatif yang tidak signifikan sehingga akan menimbulkan keputusan bias. Keputusan bias yang terjadi akan cenderung menolak butir pada sampel yang besar dan menerima butir pada sampel yang kecil. Menarik jika diperhatikan butir dengan *biserial* pada *output* tersebut yang memiliki nilai negatif dimana setelah dikorelasikan dengan bentuk butir pertanyaan, ternyata opsi jawaban yang diberikan memberikan sedikit kebingungan bagi siswa dalam menjawab oleh karena terlalu sulit dibedakan butir jawabannya.

Pada fase kedua hasil analisis model teori respons-butir dapat dikemukakan bahwa dari 24 butir soal yang dianalisis terdapat 14 butir soal yang baik atau 35 % dan 10 butir soal yang tidak baik atau 25%. Butir soal yang baik adalah butir soal dengan tingkat kesukaran soal (*b<sub>i</sub>*) terletak antara (-2, 2), daya pembeda soal (*a<sub>i</sub>*) terletak antara (0, 2).

Hasil analisis estimasi parameter butir soal UN Kimia paket C72 SMA/MA menggunakan model 2-PL. Hasilnya, jumlah butir soal kimia yang layak dianalisis menggunakan IRT adalah sebanyak 24 butir. Rerata tingkat kesukaran soal kimia ditunjukkan oleh indeks -2,150 yang cenderung mengarah ke soal dengan kategori yang mudah. Rerata daya pembeda ditunjukkan oleh indeks 0,807 yang berarti memiliki daya pembeda yang kurang baik.

Pada Tabel 3 diperoleh informasi bahwa seluruh butir yang dianalisis lanjut ke fase kedua terdapat butir cocok dengan model 2-PL yang berjumlah 19 butir dengan memiliki nilai parameter daya pembeda berada pada interval 0 sampai 2. Informasi ini berarti kurang dari setengah dari keseluruhan butir soal Kimia UN paket C72 memiliki cukup kemampuan untuk membedakan tingkat kemampuan peserta ujian.

Hasil estimasi parameter tingkat kesukaran menunjukkan terdapat 15 butir yang berada pada interval -2 sampai 2, yang berarti memiliki tingkat kesukaran yang baik. Sebanyak 9 butir soal memiliki indeks kesukaran dibawah -2,0, sehingga dapat dikatakan sebagai butir yang mudah.

Tabel 3. Karakteristik Butir Soal UN Kimia Paket C72 dengan IRT

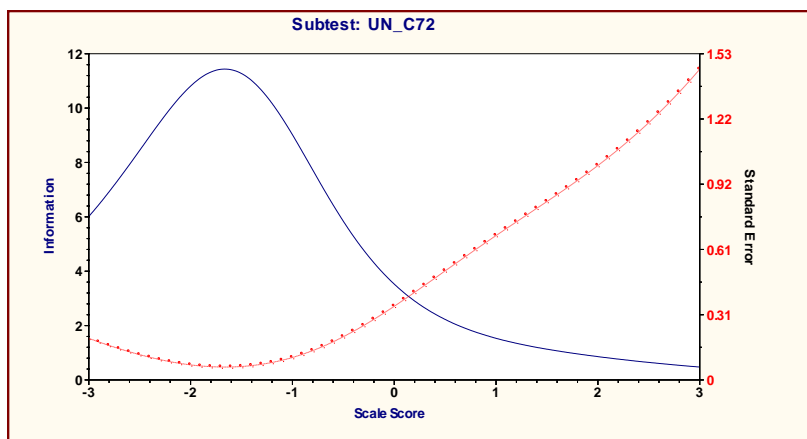
Butir	<i>a</i>	<i>b</i>	Kecocokan model	Ket
1	0,468	-1,910	Cocok	Baik
2	0,492	-4,105	TC	TB
4	0,727	-2,689	Cocok	TB
5	0,716	-2,474	Cocok	TB
6	0,572	-2,863	Cocok	TB
7	0,587	-3,432	TC	TB
8	1,174	-1,596	Cocok	Baik
10	0,808	-1,990	Cocok	Baik
12	0,805	-1,040	Cocok	Baik
13	0,792	-1,628	Cocok	Baik
15	0,409	-3,551	TC	TB
16	0,763	-1,912	Cocok	Baik
17	0,978	-1,694	Cocok	Baik
20	0,887	-1,947	Cocok	Baik
23	0,764	-1,896	Cocok	Baik
24	1,279	-1,250	TC	TB
25	0,592	-1,883	Cocok	Baik
29	1,417	-1,749	Cocok	Baik
30	1,035	-1,893	Cocok	Baik
31	1,280	-1,273	Cocok	Baik
32	0,507	-2,387	Cocok	TB
34	1,056	-1,827	Cocok	Baik
35	0,548	-2,146	Cocok	TB
37	0,703	-2,458	TC	TB

Keterangan : TC=Tidak Cocok, TB=Tidak Baik

Hasil estimasi parameter di atas dapat dijadikan informasi untuk menentukan karakteristik butir yang baik. Butir yang baik hendaknya memenuhi 3 persyaratan yaitu, butir cocok model, mempunyai daya beda 0 sampai 2, dan tingkat kesulitan -2 sampai +2. Berdasarkan kriteria tersebut maka dapat dikatakan bahwa untuk memilih soal yang baik terdapat 14 butir atau 35% butir soal yang memenuhi persyaratan sebagai butir yang baik. Butir soal tersebut adalah butir nomor 1, 8, 10, 12, 13, 16, 17, 20, 23, 25, 29, 30, 31, dan 40.

Estimasi parameter butir dengan model 2 parameter logistik pada tes Kimia UN paket C72, menghasilkan nilai informasi tes sebesar 11,4069 dengan tingkat kesalahan pengukuran 0,2961. Nilai tersebut tercapai jika parameter kemampuan siswa

( $\theta$ ) sebesar -1,60. Perhitungan fungsi informasi tes dihitung dengan bantuan program *Excel 2007 for windows*. Grafik nilai fungsi informasi tes dengan parameter kemampuan peserta yang sesuai disajikan pada Gambar 1.



Gambar 1. Grafik Nilai Fungsi Informasi, Sekor Kemampuan dan Kesalahan Baku

Kesalahan pengukuran perangkat UN Kimia Paket C72 tingkat SMA/MA Negeri tahun pelajaran 2011/2012 di Provinsi Maluku berdasarkan hasil analisis menunjukkan bahwa besarnya kesalahan pengukuran sebesar 0,296. Kesalahan pengukuran dipergunakan untuk memahami kesalahan yang bersifat acak yang mempengaruhi skor peserta yang sebenarnya dalam pelaksanaan tes. Hasil perhitungan kesalahan pengukuran, diperoleh kesalahan pengukuran sebesar 0,296 dan skor maksimum yang dapat dicapai oleh peserta adalah 38, maka skor peserta UN Kimia Paket C72 tingkat SMA/MA tahun pelajaran 2011/2012 di Provinsi Maluku sebenarnya berkisar pada  $X_i \pm 0,296$ .

#### Analisis Cheating

Deteksi *cheating* dalam penelitian ini menggunakan lima metode antara lain metode *Angoff's B-index*, metode *Pair 1*, metode *Pair 2*, metode *MESA* dan metode *G2*.

Hasil keseluruhan *cheating* untuk kelompok kota dapat dilihat pada Tabel 4 yang menyajikan frekuensi *cheating* masing-masing metode pada pelaksanaan UN Kimia SMA/MA untuk paket tes C72 berdasarkan tiga kategori tingkat *cheating* yakni tinggi, sedang dan rendah. Metode *Pair 2* yang memiliki frekuensi terbanyak dalam *cheating* ini setelah

metode *Pair 1* dan *G2* sedangkan metode *B-Index* dan *MESA* tidak ada pasangan yang didapati melakukan *cheating*. Besarnya frekuensi kasus *cheating* pada metode *Pair 2* menerangkan bahwa terdapat sebanyak 52 pasangan yang terindikasi melakukan *cheating* dengan model jumlah soal terbanyak yang memiliki jawaban yang sama pada soal-soal berurutan. Sebanyak 14 pasangan memiliki intensitas melakukan kecurangan yang tinggi diikuti 15 pasangan pada level sedang sedangkan sisanya sebanyak 23 pasangan pada level yang rendah. Adanya kemungkinan penyebaran kunci jawaban pada sekolah-sekolah di kota Ambon dibandingkan pada kota Tual juga sangat besar hal ini dapat dilihat pada tingginya frekuensi dari data *benchmark group* pada Tabel 5.

Tabel 4. Hasil *Cheating* pada Kelompok Kota (Ambon dan Tual) Berdasarkan Kedekatan Tempat Duduk

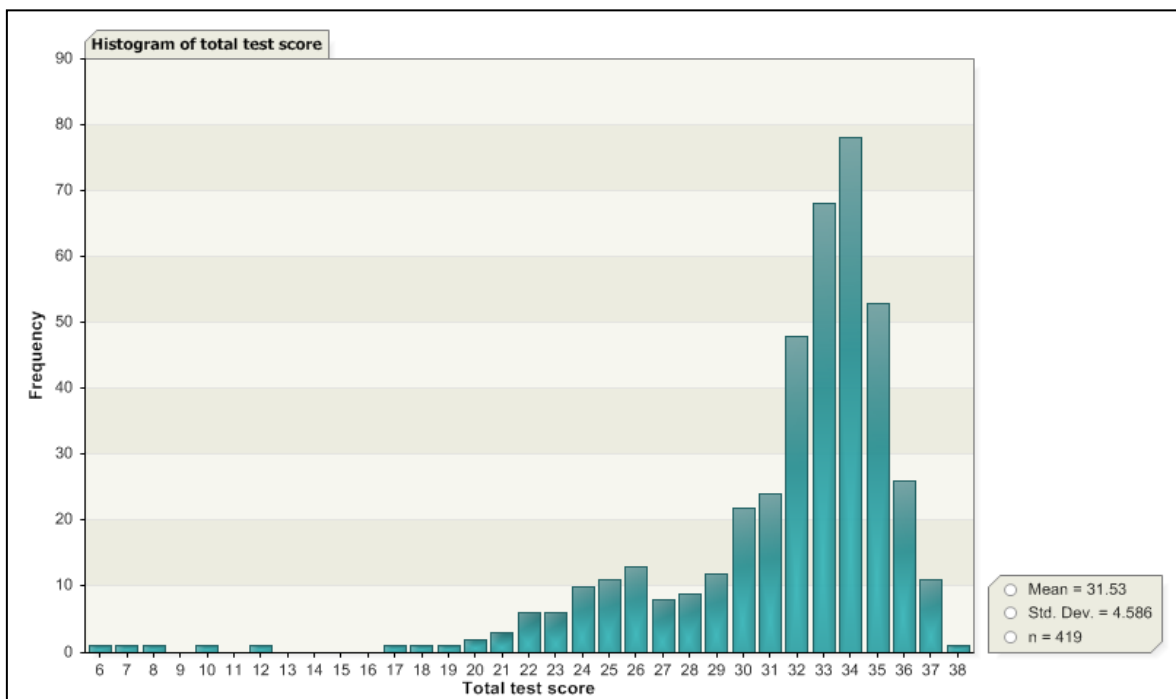
Kategori	Metode <i>Cheating</i>				
	<i>B-Index</i>	<i>Pair 1</i>	<i>Pair 2</i>	<i>MESA</i>	<i>G2</i>
Tinggi	0	0	14	0	0
Sedang	0	1	15	0	1
Rendah	0	6	23	0	4
Jumlah	0	7	52	0	5

Tabel 5. Hasil *Cheating* pada Kelompok Kota Ambon Berdasarkan *Benchmark Group*

Kategori	Metode <i>Cheating</i>				
	<i>B-Index</i>	<i>Pair 1</i>	<i>Pair 2</i>	<i>MESA</i>	<i>G2</i>
Tinggi	0	0	4	0	0
Sedang	0	0	159	0	0
Rendah	0	3	14	0	1
Jumlah	<b>0</b>	<b>3</b>	<b>177</b>	<b>0</b>	<b>1</b>

Data *benchmark group* pada Tabel 5 menjelaskan bahwa sebanyak 177 pasangan pada sekolah-sekolah yang tersebar di kota

Ambon memiliki jawaban yang seragam yang mengindikasikan kemungkinan terjadi penyebaran kunci jawaban. Hasil analisis *cheating* pada pelaksanaan UN Kimia SMA/MA Negeri Paket tes C72 di Provinsi Maluku untuk kelompok kota dengan memperhitungkan keseluruhan pasangan yang diduga melakukan *cheating* dengan pasangan terdekat atau yang termasuk dalam *benchmark group* adalah berjumlah 16 pasangan pada metode *Pair 1*, 239 pasangan pada metode *Pair 2*, 6 pasangan untuk metode *G2*, sedangkan untuk metode *Angoff's B-index* dan metode *MESA* tidak terdapat pasangan yang melakukan *cheating*.



Gambar 2. Grafik Total Sekor Tes Peserta untuk Kota Ambon

Gambar 2 memperlihatkan histogram dari total skor tes pada salah satu kelompok kota yakni kota Ambon diperoleh skor rata-rata (*mean*) sebesar 31,53 yang menunjukkan rata-rata peserta tes memiliki kemampuan yang tinggi dengan kemampuan menjawab soal cukup baik.

*Cheating* yang terjadi pada beberapa pasangan peserta tes membuktikan bahwa tidak semua peserta mampu menjawab dengan kemampuannya sendiri tetapi kemungkinan diduga terjadi kecurangan sewaktu pelaksanaan UN.

Hasil analisis *cheating* untuk daerah luar kota yang meliputi 9 Kabupaten yakni Kabupaten Maluku Tengah, Buru, Buru Selatan, Maluku Tenggara Barat, Maluku Barat Daya, Maluku Tenggara, Seram Bagian Barat, Seram Bagian Timur dan Kepulauan Aru. Tabel 6 menyajikan frekuensi *cheating* masing-masing metode pada pelaksanaan UN Kimia SMA/MA untuk Paket tes C72 pada daerah diluar kota dengan frekuensi terbesar pada metode *Pair 2* sebanyak 392 pasangan diikuti dengan metode *Pair 1* sebanyak 205 pasangan, metode *G2* sebanyak

97 pasangan, metode *B-Index* sebanyak 13 dan metode *MESA* sebanyak 7 pasangan.

Tabel 6. Hasil *Cheating* pada Kelompok Luar Kota Berdasarkan Kedekatan Tempat Duduk

Kategori	Metode <i>Cheating</i>				
	<i>B-Index</i>	<i>Pair 1</i>	<i>Pair 2</i>	<i>MESA</i>	<i>G2</i>
Tinggi	0	0	245	0	11
Sedang	0	55	53	1	60
Rendah	13	150	94	6	26
Jumlah	<b>13</b>	<b>205</b>	<b>392</b>	<b>7</b>	<b>97</b>

Tabel 6 memberikan informasi bahwa kategori *cheating* tertinggi dalam metode *Pair 2* yang mengindikasikan sebanyak 245 pasangan kemungkinan melakukan *cheating* sangat besar sekali dengan memperhatikan jumlah soal terbanyak yang memiliki jawaban yang sama pada soal-soal berurutan. Sebanyak sedangkan metode *Pair 1* yang mengindikasikan tingkat penyalinan jawaban pada pasangan terdekat tergolong sedang dan rendah. Sementara metode *G2* yang mendasarkan pada jumlah jawaban yang benar dan jumlah jawaban yang salah hasil salinan pada pasangan yang terlibat *cheating*. Hasilnya sebanyak 97 pasangan yang terlibat *cheating* untuk metode ini.

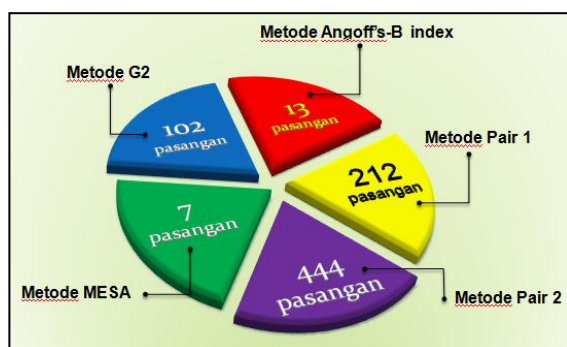
Metode *B-Index* membandingkan jumlah jawaban salah untuk semua pasangan peserta ujian yang masuk dalam interval sama dari hasil tes. interval ini didasarkan pada perkalian jumlah jawaban yang salah untuk sepasang peserta ujian kemudian diamati apakah signifikan berbeda dari nilai rata-rata jawaban salah untuk semua peserta ujian pada interval itu. Frekuensi yang terdeteksi melalui metode ini adalah sebanyak 13 pasangan dengan kategori yang rendah. Selain metode yang sudah disampaikan di atas terdapat metode *MESA* yang mendeteksi sebanyak 7 pasangan yang terlibat *cheating* dengan 1 pasangan pada kategori sedang dan 6 pasangan pada kategori rendah. Hal ini mengindikasikan bahwa probabilitas pasangan yang memiliki jawaban yang sama tetapi salah terdapat pada 7 pasangan tersebut.

Tabel 7. Hasil *Cheating* pada Kelompok Luar Kota Berdasarkan *Benchmark Group*

Kategori	Metode <i>Cheating</i>				
	<i>B-Index</i>	<i>Pair 1</i>	<i>Pair 2</i>	<i>MESA</i>	<i>G2</i>
Tinggi	0	0	352	0	0
Sedang	0	76	64	0	93
Rendah	3	199	109	3	20
Jumlah	<b>3</b>	<b>275</b>	<b>525</b>	<b>3</b>	<b>113</b>

Selanjutnya, kemungkinan terjadi penyebaran jawaban pada 9 kabupaten (luar kota), maka Tabel 7 menjelaskan dengan frekuensi yang berbeda-beda yang dilihat berdasarkan data *Benchmark Group*, metode *Pair 2*, *Pair 1*, *G2*, *B-Index* dan *MESA* menunjukkan menunjukkan frekuensi indikasi penyebaran jawaban sangat besar. Hasil analisis *cheating* pada pelaksanaan UN Kimia SMA/MA Negeri Paket C72 tahun pelajaran 2011/2012 di Provinsi Maluku untuk kelompok luar kota (kabupaten) dengan memperhitungkan keseluruhan pasangan yang diduga melakukan *cheating* dengan pasangan terdekat atau yang termasuk dalam *benchmark group* adalah terdapat 16 pasangan untuk metode *Angoff's B-index*, 480 pasangan untuk metode *Pair 1*, 917 pasangan untuk metode *Pair 2*, 10 pasangan untuk metode *MESA*, dan 210 pasangan untuk metode *G2*.

Secara keseluruhan untuk kelompok kota dan luar kota berdasarkan analisis metode *cheating* ditampilkan pada Gambar 3. Dari grafik tersebut ditampilkan terdapat jumlah pasangan yang diduga melakukan *cheating* untuk metode *Angoff's B-index* terdapat 13 pasangan, metode *Pair 1* terdapat 212 pasangan, metode *Pair 2* terdapat 444 pasangan, metode *MESA* terdapat 7 pasangan, dan metode *G2* terdapat 102 pasangan. Pasangan yang diambil adalah pasangan yang duduknya paling berdekatan sehingga adanya indikasi terjadinya *cheating*. Selanjutnya, bagi pasangan yang terindikasi *cheating* tetapi pada lokasi tempat duduk yang sudah berjauhan, pasangan ini dimasukkan ke dalam kelompok *benchmark group* untuk menduga terjadinya kebocoran atau penyebaran kunci jawaban pada sekolah maupun pada daerah tersebut.



Gambar 3. Grafik Pasangan yang Terindikasi *Cheating* pada UN SMA/ MA Kimia Paket C72 Tahun Pelajaran 2011/2012 di Provinsi Maluku

Ketidakjujuran siswa menghadapi UN memupuk ketidakjujuran siswa menghadapi kehidupan bermasyarakat, yang pada akhirnya akan melahirkan generasi-generasi muda yang tidak jujur. Ironisnya adanya kemungkinan terjadinya *cheating* pada UN 2011/2012 ini yang kemungkinan terjadi penyontekan jawaban dan penyebaran jawaban oleh pihak-pihak sekolah (bahkan pemerintah daerah) mengizinkan (bahkan membantu) *cheating* tersebut berlangsung sebagai tujuan agar kelulusannya mencapai 100%.

Akar permasalahan dari *cheating* UN menurut peneliti adalah UN sebagai alat untuk menentukan kelulusan yang hanya di administrasikan satu kali dalam setahun. Pihak-pihak tertentu (siswa, guru, kepala sekolah, orang tua, atau kepala daerah) yang merasa tidak siap menghadapi kemungkinan siswa-siswinya tidak lulus akan tergoda untuk melakukan kecurangan.

## Simpulan dan Saran

### Simpulan

Berdasarkan hasil penelitian dan pembahasan pada perangkat Ujian Nasional Kimia Paket C72 tingkat SMA/MA Negeri tahun pelajaran 2011/2012 di Provinsi Maluku dapat disimpulkan sebagai berikut: (1) berdasarkan pendekatan teori tes klasik terdapat 77,5% butir memiliki tingkat kesulitan butir berfungsi baik, 55% butir daya bedanya belum memenuhi syarat dan 70% butir pengecoh (*distractor*) berfungsi baik dengan

indeks reliabilitas tes 0,772; (2) berdasarkan pendekatan teori respons butir terdapat 14 (35%) butir soal cocok (*fit*) dengan model, fungsi informasi maksimum 11,4069 pada  $\theta = -1,6$ , dan besarnya kesalahan pengukuran (*SEM*) 2,296; (3) berdasarkan hasil analisis metode *cheating* terdapat jumlah pasangan yang diduga melakukan *cheating*: Metode *Angoff's B-index* terdapat 13 pasangan, metode *Pair 1* terdapat 212 pasangan, metode *Pair 2* terdapat 444 pasangan, metode *MESA* terdapat 7 pasangan, dan metode *G2* terdapat 102 pasangan yang terdeteksi *cheating*; (4) berdasarkan hasil analisis, metode yang lebih banyak mendeteksi adanya *cheating* dalam Ujian Nasional Kimia secara berturut-turut adalah metode *Pair 2*, metode *Pair 1*, metode *G2*, metode *Angoff's B-index* dan metode *MESA*.

### Saran

Berdasarkan hasil penelitian, disarankan beberapa hal yang berkaitan dengan perangkat Ujian Nasional Kimia Paket C72 SMA/MA Negeri di Provinsi Maluku sebagai berikut: (1) agar dapat memenuhi standar perangkat tes yang baik dalam upaya mengestimasi kemampuan peserta tes, maka pada perakitan perangkat tes dan ujicoba di waktu mendatang Dinas Provinsi perlu mengusulkan ke Pemerintah Pusat agar pelaksanaan ujicoba soal Ujian Nasional perlu dilaksanakan juga pada daerah-daerah yang sulit/jauh dari pusat; (2) sekolah dan guru perlu menanamkan sikap kejujuran kepada setiap peserta didik dalam menghadapi Ujian Nasional sebab praktik-praktik *cheating* bukanlah alternatif yang baik dalam mendongkrak nilai Ujian Nasional dari sekolah melainkan dengan melakukan dan melaksanakan sistem pembelajaran yang baik, metode pengajaran yang tepat dan efektif sehingga semuanya itu akan menjadi solusi yang positif untuk peserta didik; (3) dalam mendeteksi adanya perilaku *cheating* yang berdampak negatif pada dunia pendidikan, maka perlu dikembangkan metode *cheating* yang tepat berdasarkan kriteria yang baik sehingga data yang diperoleh lebih akurat.

## Daftar Pustaka

- Anderman, E. M., Griesinger, T., & Westfield, G. (1998). Motivation and cheating during early adolescence. *Journal of Educational Psychology*, 90, 84-93.
- Anderman, E. M., Pamela, K. C., & Derek, L. (2010). Impulsivity and academic cheating. *Journal of Educational Psychology*, 90, 84-93.
- Aziz, Deden Abdul. (September 2012). Soal Ujian Nasional SMP Diduga Bocor. *Tempo Online* diakses pada tanggal 30 September 2012, dari: <http://www.tempo.co/read/news>
- Bogle, K. D. (2000). Effect of perspective, type of student, and gender on the attribution of cheating. *Proceedings of Oclahoma Academic Science*. Oclahoma City, 80, 91-97.
- Chula, G. K., Roger, W.G., & Chris, P. (2009). Online exams and cheating: an empirical analysis of business students' views. *The Journal of Educators Online*, 6, 1.
- Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*. Mahwah, New Jersey: Lawrence Erlbaum.
- \_\_\_\_\_. (April 2001). An overview of issues concerning cheating on large-scale tests. *Paper presented at the annual meeting of the National Council on Measurement in Education*, in Seattle, WA.
- Depdiknas (2003). Undang-Undang RI Nomor 20, Tahun 2003, tentang Sistem Pendidikan Nasional.
- \_\_\_\_\_.(2005). Peraturan Pemerintah RI Nomor 19, Tahun 2005, tentang Standar Nasional Pendidikan.
- Eisenberg, J. (2004). To cheat or not cheat: effects of moral perspective and situational variables on students' attitudes. *Journal of Moral Education*, 33, 2, 163-178.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston, MA: Kluwer Academic Publishers.
- Lama, M. Al. (2008). Student's attitudes toward cheat and relation to demographic factors. *European Journal of Social Science*, 7, 1, 140-146.
- Lewkowicz, A. B. (2007). *Teaching emotional intelligence*, California: Corwin Press.
- Lim, V. K. G., Sean, K. B. S. (2001). Attitudes toward, and intentions to report, academic cheating among students in Singapore. *Ethnic & Behavior*, Lawrence Erlbaum Associates, 11(3), 261-274
- Mardapi, Djemari. (1999). Estimasi kesalahan pengukuran dalam bidang pendidikan dan implikasinya pada ujian nasional. Yogyakarta: UNY.
- McCabe, D. L., Linda K. T., & Kenneth, D.B. (2001). Cheating in academic institutions: a decade of research. *Ethnic & Behavior*, 11(3), 219-232.
- Shapiro, E. S. (2011). *Academic skill problems* (4th Ed.). New York: The Guildford Press.
- Strom, P. S., & Strom R., D. (2007) Cheating middle school and high school. The Education Forum, 71, 104-116*
- Widiatmo, H. (2009). Metode untuk mendeteksi penyontekan jawaban pada tes pilihan ganda: studi kasus SMP di Kabupaten Garut. *Pusat Penelitian Pendidikan, Balitbang Diknas*. 219-226.
- Williams, J. B. (2002). The plagiarism problem: are students entirely to blame. *Proceedings of ASCILITE*. Australia
- Williams, K., M., Craig N., Delroy L.P. (2010). Identifying and profiling scholastic cheaters: their personality, cognitive ability, and motivation. *Journal of Experimental Psychology*. 16 (3), 293-307.

## MODEL ASESMEN KEPEMIMPINAN PEMBELAJARAN KEPALA SEKOLAH PENDIDIKAN DASAR

<sup>1)</sup>Zaenal Fanani, <sup>2)</sup>Djemari Mardapi, <sup>3)</sup>Wuradji

<sup>1)</sup>LPMP Provinsi Kalimantan Selatan, <sup>2,3)</sup>Universitas Negeri Yogyakarta

<sup>1)</sup>zaenalfanani2005@yahoo.com, <sup>2)</sup>djemarimardapi@gmail.com

### Abstrak

Penelitian ini bertujuan untuk: (1) mengembangkan model asesmen kepemimpinan pembelajaran kepala sekolah, dan (2) mengetahui keefektifan kepemimpinan pembelajaran kepala sekolah. Model asesmen kepemimpinan pembelajaran ini menggunakan pendekatan *360-degree assesement*, yakni asesmen dari guru, kepala sekolah dan pengawas. Jumlah responden sebanyak 560 yang terdiri dari 466 guru, 47 kepala sekolah, dan 47 pengawas pada jenjang SD dan SMP. Dua daerah dipilih untuk mewakili Provinsi Kalimantan Selatan, yakni masing-masing satu dari Kabupaten Banjar dan Kota Banjarbaru. Ukuran sampel sekolah sebanyak 47 diambil dari wilayah berbeda, yakni wilayah kota, pinggiran, dan desa. Pengumpulan data menggunakan kuesioner skala Likert (1-4) yang terdiri atas 36 item. Responden dari setiap sekolah terdiri atas 10 guru, satu kepala sekolah, dan satu pengawas dari sekolah yang sama untuk mengisi instrumen. Teknik analisis data untuk uji kecocokan model dengan data dan invariansi parameter model antar *group rater* menggunakan *Confirmatory Factor Analysis*. Hasil penelitian menunjukkan bahwa: (1) model asesmen kepemimpinan pembelajaran kepala sekolah dengan 4 dimensi dan 36 items menunjukkan kecocokan dengan data empiris ( $\chi^2$ ,  $p=0,347>0,050$ ;  $RMSEA=0,006<0,080$ ) dan semua item memiliki *standardized loading factor* yang signifikan ( $t>1,96$ ;  $\lambda>0,50$ ), dan (2) persentase kepala sekolah yang memiliki keefektifan kepemimpinan pembelajaran pada kategori tinggi dan sedang berturut-turut adalah sebesar 19% dan 79%, dan dengan skala 100, rerata skor dimensi visi, supervisi, penilaian kinerja guru, dan pengembangan keprofesian guru berturut-turut adalah sebesar 74, 65, 65 dan 63.

**Kata kunci:** *model asesmen, kepemimpinan pembelajaran*

## A MODEL FOR ASSESSMENT OF PRINCIPAL INSTRUCTIONAL LEADERSHIP OF BASIC EDUCATION

<sup>1)</sup>Zaenal Fanani, <sup>2)</sup>Djemari Mardapi, <sup>3)</sup>Wuradji

<sup>1)</sup>LPMP Provinsi Kalimantan Selatan, <sup>2,3)</sup>Universitas Negeri Yogyakarta

<sup>1)</sup>zaenalfanani2005@yahoo.com, <sup>2)</sup>djemarimardapi@gmail.com

### Abstract

This study aims to: (1) develop a model for assessment of principal instructional leadership, and (2) describe the effectiveness of the principal instructional leadership. This assessment model was developed by using 360 degree assessment approach, i.e. the assessment from teachers, principals, and supervisors. The respondents were 560 people that consisted of 466 teachers, 47 principals, and 47 supervisors from elementary and secondary school. Banjar and Banjarbaru districts were selected to represent the South Kalimantan Province. A sample of 47 schools was established from the different areas: urban, suburban, and rural areas. The data were collected using a Likert scale questionnaire (1-4) with 36 items. The respondents who were selected from each school consisted of 10 teachers, one principal, and one supervisor to complete the questionnaire. The Confirmatory Factor Analysis was used to test the fitness between model and data and invariance of measurement across group raters. The findings of this study show that: (1) the model for assessment of instructional leadership with four dimensions and 36 items shows a good fitness to data ( $\chi^2$ ,  $p=0.347>0.050$ ;  $RMSEA= 0.006<0.080$ ) and all items have the significant standardized loading factors ( $t>1.96$ ;  $\lambda>0.50$ ), and (2) the percentage of principals having instructional leadership effectiveness at high and moderate category respectively is 19% and 79%, and by 100 scale, the mean score of vision for learning, supervision, assessment of teacher performance, teacher professional development, respectively is 74, 65, 65, and 63.

**Keywords:** *assessment model, instructional leadership*

## Pendahuluan

Asemen kepemimpinan pembelajaran kepala sekolah sebagai fungsi pembinaan keprofesian kepala sekolah menjadi domain penting dalam pendidikan. Pembinaan ini penting karena kepemimpinan kepala sekolah secara luas telah dipandang sebagai salah satu faktor yang menentukan terhadap keberhasilan sekolah dalam menciptakan prestasi belajar siswa (Leithwood et.al, 2004, p.17). Masa kini, kepala sekolah sebagai pemimpin pembelajaran dipandang sebagai kepala sekolah efektif. *Asia Society Partnership for Global Learning* (2012) melakukan pertemuan tingkat tinggi dengan 24 Negara membahas tentang *Teaching and Leadership for 21<sup>st</sup> Century* dan menyepakati tentang kepemimpinan kepala sekolah yang efektif, yaitu fungsi kepala sekolah yang lebih fokus pada kepemimpinan pembelajaran dari pada managerial. *Trend* peran kepala sekolah secara progresif telah bergeser dari peran sebagai manajer ke peran sebagai pemimpin pembelajaran (Shelton (2010, p.5). Masa kini, ada dua model kepemimpinan kepala sekolah yang menjadi perhatian para peneliti, yaitu kepemimpinan transformasional dan kepemimpinan instruksional (pembelajaran). Pada era kebutuhan akuntabilitas yang semakin besar, terutama berkaitan dengan pengukuran berbasis outcome, peran kepala sekolah yang berorientasi pada pembelajaran semakin diperlukan (Lunenburg, 2010, p.1). Asesmen kepemimpinan pembelajaran kepala sekolah yang berkualitas menjadi kebutuhan global untuk menghasilkan kepala sekolah yang efektif.

Robinson et al. (2009, p.90) melakukan analisis meta terhadap 12 dari 13 penelitian tentang kepemimpinan kepala sekolah dan hasilnya menunjukkan bahwa *effect size* model kepemimpinan pembelajaran adalah tertinggi di antara model-model kepemimpinan yang lain. *Effect size* kepemimpinan pembelajaran adalah tiga kali lebih besar dibandingkan dengan *effect size* kepemimpinan transformasional. Temuan ini menunjukkan bahwa model kepemimpinan pembelajaran kepala sekolah adalah

model kepemimpinan yang paling efektif sehingga sangat dibutuhkan saat ini untuk memperbaiki prestasi siswa. Kosekuensinya, model asesmen kepemimpinan pembelajaran kepala sekolah menjadi urgen.

Hallinger & Heck (1998) melakukan *review* terhadap 40 hasil penelitian tentang pengaruh kepemimpinan kepala sekolah terhadap prestasi siswa dan menyimpulkan bahwa pengaruh tidak langsung lebih konsisten dari pada pengaruh langsung. Kepala sekolah perlu konsentrasi pada pembentukan sikap dan perilaku guru untuk mencapai prestasi siswa yang tinggi. Upaya kepala sekolah untuk meningkatkan prestasi siswa tidak bisa secara langsung akan tetapi harus melalui perbaikan kinerja guru. McEwan (2003, p.6) berpendapat bahwa pemimpin pembelajaran harus memiliki pengetahuan tentang teori belajar, instruksional dan kurikulum. Kepala sekolah sebagai pemimpin pembelajaran secara langsung berhubungan dengan proses pembelajaran dimana guru, peserta didik dan kurikulum berinteraksi. Kepemimpinan pembelajaran bertujuan untuk perbaikan pembelajaran sebagai *core bussiness* pendidikan. Hal yang sama disampaikan oleh Bush (2008, p.17) bahwa kepemimpinan pembelajaran adalah dimensi yang sangat penting karena fokus pada kegiatan inti sekolah, yakni kegiatan belajar-mengajar (pembelajaran) di kelas. Perbaikan pembelajaran di satuan pendidikan dapat berlangsung secara terus menerus apabila kompetensi guru sebagai aktor utama ditingkatkan secara terus menerus.

Berdasarkan hasil pembahasan tersebut, konsep kepemimpinan pembelajaran kepala sekolah dapat didefinisikan sebagai perilaku kepala sekolah yang berpengaruh tidak langsung terhadap prestasi siswa melalui guru. Dengan kata lain, kepemimpinan pembelajaran adalah perilaku kepala sekolah yang berpengaruh langsung terhadap kinerja guru yang pada gilirannya berdampak pada prestasi siswa. Holifield & Cline (1997, p.109) menyatakan bahwa salah satu tugas utama kepala sekolah adalah untuk meningkatkan performa guru. Definisi ini telah disepakati oleh semua ahli pendidikan dan menjadi

kebijakan pemerintah. Hal serupa juga dinyatakan oleh DuFour & Berkey (1995, p.1) bahwa kesuksesan dari upaya perbaikan sekolah bergantung pada upaya pengembangan keprofesian di dalam sekolah, terutama pengembangan keprofesian guru. Dengan kata lain, hampir seluruh energi kepala sekolah digunakan untuk memperbaiki mutu pembelajaran melalui mengembangkan keprofesian guru secara terus menerus.

Menurut Gaziel (2007), sasaran utama dari perbaikan pembelajaran secara terus menerus adalah untuk peningkatan prestasi siswa melalui guru dan kultur sekolah. Dengan demikian, pembinaan keprofesian kepala sekolah melalui asesmen kepemimpinan pembelajaran berkaitan erat dengan program peningkatan kinerja guru dan program peningkatan prestasi siswa sehingga merupakan kerangka konseptual penjaminan mutu satuan pendidikan.

Kepemimpinan pembelajaran kepala sekolah sebagai salah satu dari enam kompetensi kepala sekolah telah diakomodasi oleh pemerintah dengan telah ditetapkannya Permendiknas No. 35 Tahun 2010 tentang Petunjuk Teknis Pelaksanaan Jabatan Fungsional Guru dan Angka Kreditnya. Dalam Permendiknas nomor 28 tahun 2010 tentang Penugasan Guru sebagai Kepala Sekolah, dimana pada pasal 12 ayat 1 disebutkan bahwa penilaian kinerja kepala sekolah/madrasah dilakukan secara berkala setiap tahun dan secara kumulatif setiap 4 (empat) tahun. Pada satu sisi, informasi hasil penilaian tahunan adalah sebagai dasar untuk pengembangan program perbaikan kinerja kepala sekolah. Pada sisi lain, informasi hasil penilaian empat tahunan adalah sebagai dasar untuk pengambilan keputusan berkaitan dengan promosi kepala sekolah.

Namun, masalah yang terjadi adalah bahwa asesmen kepemimpinan pembelajaran kepala sekolah tersebut belum sesuai dengan harapan, yakni belum memiliki penjelasan tentang bukti validitas dan reliabilitas yang menjadi kriteria utama sebagai alat ukur yang berkualitas. Asesmen yang tidak memenuhi syarat validitas dan reliabilitas

memberikan data dan informasi yang tidak akurat sehingga berdampak pada ketidakakuratan pada pengambilan keputusan dan program pembinaan. Program pembinaan kepala sekolah selama ini salah sasaran. Faktor-faktor yang tidak efektif menjadi sasaran pembinaan, namun faktor-faktor yang efektif tidak menjadi sasaran pembinaan. Meskipun kepala sekolah diberi pembinaan terus-menerus, namun profesionalitas kepala sekolah tidak mengalami perkembangan yang berarti karena program pembinaannya kurang relevan dengan perilaku kepemimpinan yang efektif.

Reeve (2005) menyatakan bahwa asesmen kepemimpinan memberikan kontribusi yang signifikan terhadap penyediaan pendidik dengan informasi yang dibutuhkan yang dapat digunakan untuk tujuan perbaikan kepemimpinan dan akuntabilitas. Reeve menemukan bahwa para kepala sekolah setuju bahwa asesmen secara umum berdampak positif, akurat dan konsisten dengan tugas. Namun, Reeve jarang menemukan proses asesmen relevan terhadap perbaikan motivasi dan perbaikan kinerja. Asesmen pada umumnya fokus pada tugas atau daftar tanggung jawab sebagai kepala sekolah dan karakteristik yang meliputi sifat, disposisi dan atribusi.

Goldring (2007) melakukan kajian secara komprehensif tentang asesmen kepala sekolah dan menyimpulkan bahwa isi dari asesmen kepala sekolah dalam praktek di lapangan kurang berkaitan dengan perilaku kepemimpinan yang berdampak pada belajar siswa. Karakteristik instrumen pengukuran (psikometrik) hampir tidak pernah dilaporkan. Goldring menemukan hanya 4 dari 66 alat ukur kepemimpinan kepala sekolah yang mendiskripsikan karakteristik psikometrik. Sebagian besar instrumen pengukuran tidak menyediakan informasi tentang standar yang digunakan dan tidak menyediakan norma sebagai pembanding.

#### Pendekatan Asesmen

Pendekatan asesmen kepemimpinan dalam penelitian ini meliputi dua pendekatan, yaitu pendekatan *upward feedback* dan

360° *feedback*. Heslin dan Latham (2004, p.24) menyatakan bahwa *upward feedback* atau *subordinate appraisal* merupakan cara yang efektif dalam mengubah perilaku manajer. Instrumen appraisal fokus pada perilaku dan bukan pada sifat pribadi pemimpin. Skala appraisal fokus pada frekuensi perilaku yang diobservasi. Salah satu perusahaan besar yang menggunakan asesmen bawahan (*upward feedback*) adalah perusahaan IBM. IBM melaporkan bahwa pegawainya menjadi lebih terlibat dan lebih menunjukkan kepuasan kerja dengan adanya asesmen dari bawahan. Heslin dan Latham dengan penelitian *quasi-experimental study* menemukan bahwa *upward feedback* berdampak pada kinerja manager lebih tinggi setelah 6 bulan kemudian dibandingkan dengan manager tanpa *upward feedback* sebagai grup kontrol.

Pendekatan 360° *feedback* dipandang sebagai pendekatan ideal, yakni penilaian dari pengawas, teman sesama kepala sekolah, kepala sekolah bersangkutan dan guru. Menurut Olson (2008, p.2), Universitas Vanderbilt dan Universitas Pennsylvania telah mengembangkan instrumen untuk mengukur perilaku kepemimpinan kepala sekolah yang berkaitan dengan prestasi siswa. Instrumen penilaian ini menggunakan “360-degree” *feedback* yakni penilaian dilakukan oleh seluruh guru di sekolah, kepala sekolah dan pengawas. Hasil uji coba menunjukkan bahwa instrumen memiliki reliabilitas yang tinggi, total 0.95 dan tiap subscale 0,90.

Hallinger (2008) melakukan kajian terhadap 25 tahun hasil penelitian yang menggunakan PIMRS dan menemukan bahwa penilaian dari guru lebih konsisten dibandingkan dengan kepala sekolah dan pengawas. Hallinger & Murphy (1985) menyatakan bahwa konsistensi guru ditunjukkan dengan konsistensi hasil isian instrumen, hasil wawancara dan dokumen. Penilaian dari diri kepala sekolah selain menggambarkan realita yang ada, kemungkinan juga dipengaruhi oleh keinginan citra yang baik tentang dirinya. Penilaian dari pengawas kurang menggambarkan realita yang

ada karena pengawas tidak selalu ada di sekolah setiap harinya sehingga tidak banyak tahu tentang pelaksanaan kepemimpinan pembelajaran kepala sekolah. Oleh karenanya, personil pendidikan yang tahu banyak dan mengalami langsung serta lebih objektif menilai pelaksanaan kepemimpinan pembelajaran kepala sekolah adalah guru-guru di sekolah. Namun, belum pernah dicoba seberapa lebar perbedaan ketiga kelompok tersebut dalam memberikan penilaian

Pengembangan model asesmen kepemimpinan pembelajaran kepala sekolah yang memenuhi syarat alat ukur yang baik sangat dibutuhkan saat ini untuk mengatasi permasalahan sebagaimana diuraikan di atas. Pengembangan model asesmen adalah penting karena beberapa alasan: asesmen merupakan bagian penting dari pembinaan, pembelajaran adalah *Core business* pendidikan, asesmen kepemimpinan pembelajaran kepala sekolah belum dikembangkan, sebagian besar kinerja kepala sekolah lemah pada aspek perbaikan pembelajaran, prestasi siswa rendah pada laporan PISA dan TIMMS, dan kompetensi guru rendah pada hasil uji kompetensi guru dan hasil temuan JICA.

Dalam penelitian ini, model asesmen kepemimpinan pembelajaran kepala sekolah dikembangkan dari hasil sintesa model Hallinger & Murphy (1985), Blasé & Blasé (2000), SEQIP (2004), dan sistem pendidikan Indonesia. Model asesmen kepemimpinan pembelajaran kepala sekolah memiliki empat dimensi: (1) visi belajar, (2) supervisi pembelajaran, (3) mengembangkan keprofesian guru secara berkelanjutan, dan (4) penilaian kinerja guru. Keempat dimensi ini adalah tindakan yang harus kepala sekolah lakukan sebagai pemimpin efektif untuk perbaikan pembelajaran yang berdampak langsung terhadap peningkatan prestasi siswa.

#### Visi Belajar

Hallinger & Heck (1998) mengatakan bahwa visi adalah faktor kunci dalam membangun organisasi sekolah yang efektif. Tanpa visi, personil di dalam sekolah hanya

bekerja dan bekerja secara rutin tanpa sebuah arah. Menurut Smith & Piele (2006, p.43), visi berfungsi memberikan arah bagi setiap orang yang terlibat dalam suatu organisasi. Dalam konteks pendidikan, visi harus berkaitan dengan belajar. Visi belajar yang dibangun bersama di sekolah memberikan arah kepada setiap personil sekolah (siswa, guru, kepala sekolah, pengawas dan orangtua) untuk mencapai kompetensi lulusan yang ditetapkan bersama. Murphy et al. (2006, p.7) menyatakan bahwa pemimpin yang berpusat pada belajar menggunakan sebagian besar energinya untuk kegiatan pengembangan, artikulasi, penerapan, dan pengamanan visi belajar. Visi belajar dibangun bersama antar *stakeholders* sekolah yang berfungsi sebagai arah dan sumber inspirasi bagi semua perilaku personil sekolah dalam perbaikan pembelajaran.

Visi belajar dibangun berdasarkan standar kompetensi lulusan yang ingin dicapai bersama. Penetapan standar belajar yang tinggi dapat memberikan tantangan bagi personil sekolah untuk mencapainya. Menurut Goldring et al, (2007, p.5), tugas kepala sekolah adalah untuk menetapkan standar hasil belajar siswa yang tinggi dan kurikulum yang ketat. Dengan standar dan kurikulum yang jelas, personil sekolah terbantu untuk menetapkan tujuan sekolah. Personil sekolah termotivasi dengan tujuan yang dapat dicapai dan tertantang untuk mencapainya (Leithwood, 2004, p.8). Oleh karena itu, penetapan tujuan yang jelas dan komitmen untuk mencapai tujuan tersebut harus ditempatkan pada posisi terdepan dan menjadi perhatian utama bagi kepala sekolah. Hasil analisis meta oleh Witziers, Bosker, & Kruger (2003) antara tahun 1986 dan 1996 ditemukan bahwa “menetapkan dan mengkomunikasikan misi sekolah” memberikan pengaruh langsung terhadap prestasi siswa.

Kepala sekolah sebagai pemimpin pembelajaran memiliki peran dan tanggung jawab untuk menjamin bahwa proses pembelajaran dan penilaian yang dikembangkan guru sesuai dengan kompetensi siswa yang ingin dicapai. Oleh karenanya, kepala

sekolah diharapkan mampu memberikan arah dan fasilitas bagi guru untuk mampu menjabarkan standar kompetensi menjadi tahapan pembelajaran dan alat penilaian. Kesesuaian proses pembelajaran dengan standar kompetensi akan menjamin atau memungkinkan (*enabling*) tercapainya standar kompetensi siswa yang diinginkan. Kesesuaian penilaian dengan standar kompetensi dapat meningkatkan motivasi siswa belajar lebih giat untuk mencapai kompetensi yang ditetapkan.

Secara operasional, dimensi standar dalam kepemimpinan pembelajaran adalah kemampuan kepala sekolah untuk mengkomunikasikan yang direfleksikan ke dalam empat indikator perilaku: (1) data ketercapaian standar kompetensi siswa, (2) mengharapkan ketercapaian standar kompetensi siswa yang lebih tinggi, (3) mengharapkan kesesuaian pembelajaran (standar proses) dengan standar kompetensi siswa yang tinggi, dan (4) mengharapkan kesesuaian penilaian (standar penilaian) dengan standar kompetensi siswa dan model pembelajaran yang digunakan.

### Supervisi Pembelajaran

Sullivan & Glanz (2005, p.27) mendefinisikan supervisi sebagai proses untuk melibatkan para guru dalam dialog pembelajaran dengan tujuan untuk memperbaiki pengajaran dan prestasi siswa. Tujuan utama supervisi adalah perbaikan pembelajaran melalui pengembangan profesional guru. Kegiatan supervisi pembelajaran ini bukan untuk penilaian kinerja guru akan tetapi sebagai bentuk bantuan teknis dalam mendiagnosis dan saran perbaikan.

Glickman, Gordon & Gordon (2007, p.8) menyatakan bahwa *Effective supervision requires knowledge, interpersonal skills, and technical skills*. Supervisi yang efektif memerlukan pengetahuan, keterampilan personal dan keterampilan teknis. Kepala sekolah harus memiliki pengetahuan tentang teori belajar dan pembelajaran sehingga ia mampu membedakan antara pembelajaran guru yang sesuai dan pembelajaran yang tidak sesuai dengan hasil belajar siswa yang ingin

dicapai. Kepala sekolah harus memiliki keterampilan personal sehingga komunikasi antara kepala sekolah dan guru lebih bersifat kolegial dari pada *top down* sehingga menyadarkan guru memperbaiki sendiri. Kepala sekolah harus memiliki keterampilan teknis seperti mendemonstrasikan contoh mengajar yang efektif di depan para siswa sehingga guru paham cara dan waktu menerapkan teknik-teknik pembelajaran yang efektif. Blasé & Blasé (2004, p.162), menemukan tiga elemen utama dalam kepemimpinan pembelajaran yang sukses: (1) melakukan konferensi pembelajaran seperti saran, *feedback*, *modeling*, menggunakan *inquiry*, (2) mendorong guru untuk menyampaikan saran dan pendapat. (3) mengembangkan refleksi guru seperti modeling, observasi kelas, dialog, saran, dan pujian. Indikator-indikator perilaku kepala sekolah yang merefleksikan dimensi supervisi antara lain: melakukan observasi proses pembelajaran, mendiskusikan data hasil observasi, mendorong guru melakukan refleksi, dan mendemonstrasikan contoh mengajar.

#### Pengembangan Keprofesional Berkelanjutan bagi Guru

DuFour & Berkey (1999) mengatakan bahwa pengembangan keprofesional guru di sekolah adalah sebagai prioritas perbaikan pembelajaran. DuFour & Berkey menyatakan bahwa kesuksesan dari upaya perbaikan sekolah bergantung pada kegiatan keprofesional di dalam sekolah. Kepala sekolah dapat menciptakan kondisi yang menjamin bahwa pertumbuhan profesionalisme adalah bagian kultur sekolah. Pengembangan keprofesional di sekolah dapat dibangun jika para guru memiliki visi bersama, bekerjasama untuk merancang kurikulum, pembelajaran dan penilaian, serta saling melakukan observasi kelas. Hal senada dinyatakan oleh Smith & Piele (2006, p.43) bahwa sekolah efektif berkaitan erat dengan program sekolah untuk mengembangkan kapasitas sumberdaya manusia secara berkelanjutan di sekolah, khususnya guru.

Blasé & Blasé (2004, p.162) menyatakan bahwa dalam program pengembangan

keprofesional guru, kepala sekolah melakukan studi tentang pembelajaran, mendukung kolaborasi, pengembangan hubungan pelatihan, menggunakan penelitian tindakan, penyediaan sumber-sumber pendukung, menerapkan prinsip-prinsip pertumbuhan orang dewasa, dan pengembangan seluruh tahapan program pengembangan staf. Definisi operasional pengembangan keprofesional guru dalam konsep kepemimpinan pembelajaran direfleksikan oleh perilaku kepala sekolah untuk memastikan para guru secara kolaboratif dan rutin: (1) memperdalam materi kurikulum yang sehari-hari diajarkan, (2) saling belajar dengan guru lain tentang teknik pembelajaran melalui banyak melakukan pengamatan, (3) banyak mencoba teknik-teknik pembelajaran bersama teman guru, (4) dan membiasakan melakukan refleksi terhadap hasil pengamatan pengajaran dan hasil mencoba teknik-teknik pembelajaran.

#### Penilaian Kinerja Guru

Menurut Marzano (2012, p.9), pemimpin sekolah memastikan bahwa para guru dievaluasi dengan jelas dan secara terus menerus tentang kekuatan dan kelemahan pedagogi berdasarkan data dari berbagai sumber dan konsisten dengan data prestasi siswa. Informasi tentang kekuatan dan kelemahan pedagogi guru sangat bermanfaat bagi guru sendiri sebagai *feedback* untuk perbaikan. Menurut Sweeney (2003: 11), *feedback* sebagai bagian integral kultur sekolah. Penilaian dan *feedback* terhadap guru berfungsi sebagai saran komunikasi akademik antara kepala sekolah dan guru sehingga dapat memebtuk pemahaman yang sama tentang permasalahan pembelajaran. Padangan senada disampaikan oleh Stronge (2006: 241) bahwa penilaian kinerja guru berfungsi untuk mengidentifikasi kekuatan dan kelemahan guru. Penilaian kinerja guru yang dilakukan oleh kepala sekolah memberikan informasi tentang kekuatan dan keberhasilan guru yang berfungsi meningkatkan kepercayaan diri dan motivasi guru. Di samping sebagai *feedback*, hasil penilaian kinerja guru dapat dimanfaatkan untuk me-

nyusun profil kinerja guru sebagai input dalam penyusunan program pengembangan keprofesian berkelanjutan.

Dalam konteks sistem pendidikan di Indonesia, tugas kepala sekolah untuk menilai kinerja guru dijelaskan dalam Pemandiknas dan KaBKN Nomor 03/V/PB/2010 dan Nomor 14 Tahun 2010 tentang Petunjuk Pelaksanaan Jabatan Fungsional Guru dan Angka Kreditnya. Pada pasal 22 ayat 1 disebutkan bahwa penilaian kinerja guru dilakukan oleh kepala sekolah. Hasil penilaian kinerja guru menentukan karir seorang guru. Penilaian kinerja guru oleh kepala sekolah yang dilakukan dengan objektif akan memotivasi guru untuk selalu meningkatkan kapasitas profesionalnya. Dimensi akuntabilitas kepemimpinan pembelajaran kepala sekolah salah satu faktor pendorong perbaikan pembelajaran yang berdampak langsung terhadap peningkatan prestasi siswa.

Definisi konseptual akuntabilitas kepemimpinan pembelajaran adalah kegiatan kepala sekolah untuk menilai kinerja guru, yakni dengan membandingkan antara kinerja sebenarnya dengan standar kinerja. Hasil penilaian ini dalam bentuk keputusan tercapai standar, melampaui standar atau di bawah standar. Definisi operasional akuntabilitas direfleksikan dalam bentuk indikator perilaku kepala sekolah: (1) menyampaikan jadwal penilaian, (2) menyampaikan criteria penilaian, (3) melaksanakan penilaian, (4) membuat keputusan, (5) menyampaikan laporan.

Maksud dari pengembangan asesmen kepemimpinan pembelajaran kepala sekolah yang efektif adalah untuk menjadikan program pembinaan kepala sekolah lebih efektif. Dengan program pembinaan yang efektif, peran kepemimpinan pembelajaran kepala sekolah dapat berfungsi optimal sehingga dapat memberikan kontribusi besar terhadap peningkatan profesional guru yang berdampak pada peningkatan prestasi siswa. Pengembangan model asesmen kepemimpinan pembelajaran kepala sekolah memiliki efek signifikan terhadap peningkatan prestasi siswa.

Sasaran pengukuran dibatasi pada asesmen perilaku kepemimpinan pembelajaran kepala sekolah, tidak melibatkan tes pengetahuan, tidak pula keterampilan atau kepribadian serta tidak menilai kinerja yang berkaitan dengan daftar tugas. Lebih khusus sasaran pengukuran dibatasi pada perilaku kepemimpinan yang memberikan efek langsung pada kinerja guru dan tidak langsung pada hasil belajar siswa. Penelitian adalah untuk mencoba memadukan beberapa model kepemimpinan pembelajaran kepala sekolah pendidikan dasar yang dikembangkan di negara-negara maju dan model yang telah dikembangkan di Indonesia. Hasil perpaduan ini diharapkan mengasikkan model asesmen yang memenuhi syarat validitas dan reliabilitas, lebih efektif, lebih sederhana, dan mudah digunakan.

Tujuan penelitian ini adalah untuk mengembangkan model asesmen kepemimpinan pembelajaran kepala sekolah pendidikan dasar dan untuk mengetahui keefektifan kepemimpinan pembelajaran kepala sekolah pendidikan dasar.

## Metode Penelitian

Metode penelitian meliputi jenis penelitian, waktu dan tempat penelitian, subyek penelitian, prosedur, instrumen dan teknik analisis data.

### Jenis Penelitian

Peneliti ini menggunakan pendekatan *Research and Development* (R&D). Produk yang dikembangkan dalam penelitian ini adalah model asesmen kepemimpinan pembelajaran kepala sekolah pendidikan dasar, jenjang SD dan SMP. Model asesmen kepemimpinan pembelajaran kepala sekolah pendidikan dasar ini dikembangkan dengan mengadaptasi model Cennamo dan Kalk (2005) yang meliputi 6 (enam) langkah, yakni (1) identifikasi kebutuhan produk, (2) desain produk, (3) evaluasi produk, (4) ujicoba produk, (5) analisis dan penyempurnaan produk, (6) penggunaan produk.

## Waktu dan Tempat Penelitian

Waktu penelitian dilakukan selama tiga bulan, yaitu pada bulan Januari s.d Maret Tahun 2013. Tempat penelitian dilakukan di dua daerah Provinsi Kalimantan Selatan, yakni kota Banjarbaru dan kabupaten Banjar.

## Subjek Penelitian

Ukuran sampel dalam penelitian ini sebanyak 560 responden yang meliputi guru 466 orang, dari kepala sekolah 47 orang dan dari pengawas 47 orang pada jenjang SD dan SMP. Dua daerah dipilih untuk mewakili Provinsi Kalimantan Selatan, yakni masing-masing satu dari Kabupaten Banjar dan Kota Banjarbaru. Ukuran sampel sekolah sebanyak 47 diambil dari wilayah berbeda, yakni wilayah kota, pinggiran, dan desa. Ukuran sampel responden ditentukan oleh jumlah variabel teramati (*observed variables*) atau item. Menurut Hair, et. al. (1998), ukuran sampel yang disarankan untuk menggunakan estimasi Maximum Likelihood adalah sebesar 100-200.

## Prosedur

Desain uji coba produk asesmen kepemimpinan pembelajaran dalam penelitian ini menggunakan multi-rater, yakni rater dari grup pengawas (*supervisor*), grup kepala sekolah (*self*) dan grup guru (*subordinate*). Peneliti mendatangi setiap sekolah sampel untuk menemui kepala sekolah dan menjelaskan tujuan dan manfaat penelitian bagi kepala sekolah sendiri dan bagi guru. Disamping itu, peneliti menjelaskan tugas kepala sekolah untuk menjelaskan kepada guru-guru tentang tujuan dan manfaat penelitian bagi guru serta perlunya mengisi instrumen apa adanya. Kepala sekolah dan guru diminta untuk mengisi instrumen selama 10 menit dan hasilnya langsung dikumpulkan kepada peneliti. Untuk responden pengawas SD, peneliti mendatangi di kantor pengawas di Unit Pelaksana Teknis Pendidikan dan untuk responden pengawas SMP, peneliti mendatangi kantor dinas kab/kota untuk meminta pengawas mengisi

instrumen selama 10 menit dan hasilnya langsung dikumpulkan kepada peneliti.

## Data, Instrumen, dan Teknik Pengumpulan Data

Jenis data dalam penelitian ini adalah data dengan skala ordinal. Instrumen yang digunakan adalah instrumen kuesioner dengan skala Likert (1-4) hasil pengembangan dalam penelitian ini sebagai bagian dari model asesmen kepemimpinan pembelajaran kepala sekolah yang terdiri dari 4 dimensi dan 36 item.

Teknik pengumpulan data dengan pengisian instrumen oleh 10 guru, satu kepala sekolah dan satu pengawas untuk setiap sekolah dari sekolah yang sama. Tempat pengisian instrumen bagi guru dan kepala sekolah di sekolah masing-masing, bagi pengawas SMP di kantor Dinas Pendidikan Kabupaten dan bagi pengawas SD di Unit Pelaksana Teknis Pendidikan di masing-masing kecamatan. Waktu pengisian hanya memerlukan  $\pm 10$  menit dan hasil isian instrumen dikumpulkan kepada peneliti pada hari yang sama.

## Teknik Analisis Data

Teknik analisis data kualitatif dengan menggunakan *expert judgment* untuk validasi isi model asesmen kepemimpinan pembelajaran kepala sekolah. Teknik analisis data kuantitatif untuk uji kecocokan model dengan data empiris, uji konvergensi variabel laten antar grup rater, dan analisis invariansi parameter model pengukuran antar grup rater menggunakan *Confirmatory Factor Analysis*. Uji signifikansi korelasi antara model asesmen antar grup rater dengan prestasi siswa digunakan *Pearson Correlation*

## Hasil Penelitian dan Pembahasan

Hasil penelitian ini dirangkum dalam bentuk informasi rerata skor dan standar deviasi keempat dimensi dan total untuk masing-masing grup rater seperti dipaparkan pada Tabel 1 berikut ini.

Tabel 1. Retata Skor dan Standar Deviasi Dimensi setiap Grup Rater

Dimensi	Grup Rater							
	Self		Supervisor		Subordinate		Total	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Visi	3,16	0,45	3,18	0,39	2,93	0,37	3,09	0,40
PKB	2,69	0,52	2,61	0,49	2,52	0,45	2,61	0,49
Supervisi	2,64	0,63	2,80	0,52	2,55	0,44	2,67	0,53
Penilaian	2,60	0,66	2,75	0,60	2,58	0,47	2,64	0,58
Composit	2,77	0,57	2,84	0,50	2,65	0,43	-	-

Tabel 1 menunjukkan bahwa grup *self* dan *supervisor* dalam menilai dimensi Visi relatif sama, 3,16 dan 3,18 tetapi kedua grup lebih tinggi dibandingkan dengan pada grup *subordinate*, 2,93. Perbedaan kedua grup rater dengan grup *subordinate* adalah signifikan. Dalam menilai dimensi PKB, grup *self* cenderung paling tinggi, 2,69, disusul grup *supervisor*, 2,61, dan grup *subordinate* paling rendah, 2,52. Pada dimensi Supervisi dan Penilaian, grup *self* cenderung menilai lebih rendah, 2,64 dan 2,60, dari pada grup *supervisor*, 2,80 dan 2,75, tetapi lebih tinggi dari grup *subordinate*, 2,55 dan 2,58. Namun demikian, perbedaan ketiga grup rater pada dimensi PKB, Supervisi dan Penilaian tidak signifikan.

Pada rerata composit, rerata skor grup *self*, 2,77, lebih rendah dari grup *supervisor*, 2,84, tetapi lebih tinggi dari grup *subordinate*, 2,65. Standar deviasi pada grup *self* paling besar, 0,57, SD di bawahnya grup *supervisor*, 0,50, dan SD paling kecil grup *subordinate*, 0,43. Pada total grup rater, rerata dimensi Visi paling tinggi, 3,09 disusul dimensi Supervisi, 2,67, dimensi Penilaian, 2,64 dan paling rendah dimensi PKB, 2,64. Standar deviasi pada dimensi Visi paling kecil, 0,40, SD di atasnya dimensi PKB, 0,49, SD di atasnya lagi dimensi Supervisi, 0,53, dan SD dimensi Penilaian paling besar, 0,58.

Berdasarkan analisis tersebut dapat disimpulkan bahwa ketiga grup rater memiliki kecenderungan berbeda dalam menilai dimensi yang sama tentang kepemimpinan pembelajaran kepala sekolah. Grup *supervisor* (pengawas) cenderung paling *lenient* (murah) di antara ketiga grup rater dalam memberikan penilaian terhadap semua dimensi yang

sama, kecuali terhadap dimensi PKB. Sebaliknya grup *subordinate* (guru) cenderung paling *strict* (mahal) dalam menilai semua dimensi dan grup *self* berada di antara keduanya. Variabilitas semua dimensi dalam grup *self* adalah paling besar diantara ketiga grup rater, sebaliknya variabilitas semua dimensi dalam grup *subordinate* adalah paling kecil dan variabilitas semua dimensi dalam grup *supervisor* di antara keduanya.

Perbedaan kecenderungan antara ketiga grup rater untuk menilai dimensi yang sama belum memberikan penjelasan yang lengkap sebelum analisis karakteristik model pengukuran berdasarkan respon dari grup yang berbeda. Analisis karakteristik model pengukuran ini adalah analisis validitas konstruk dengan menggunakan *Confirmatory Factor Analysis* (CFA). Metode yang digunakan untuk estimasi parameter model dengan CFA bergantung pada keadaan normalitas distribusi data. Hasil uji multivariat normalitas distribusi data menunjukkan data tidak memenuhi normalitas distribusi data. Oleh karena itu, metode estimasi parameter model pengukuran menggunakan *Robust Maximum Likelihood*. Input data bukan data mentah, akan tetapi data mentah dikonversi menjadi data (Schumacker & Lomac, 2004: 34) dalam format *asymptotic covariance matrix* dan *covariance matrix data* untuk koreksi terhadap bias dalam *standard error* dan *fit statistics*. Format data *asymptotic covariance matrix* ini menjadi input untuk analisis kecocokan model dengan data empiris dengan CFA.

Hasil analisis kecocokan model pengukuran dengan data empiris pada level *first order* CFA seperti disajikan pada Tabel 2.

Tabel 2. Hasil *First Order* CFA

Dimensi	Model	$\chi^2$	df	p-value	RMSEA	Fit
Visi Belajar	A priori	60,52	27	0,000	0,047	Tidak
	Revisi	29,58	26	0,286	0,016	Fit
PKB Guru	A priori	25,98	27	0,519	0,000	Fit
	Supervisi	30,72	27	0,283	0,016	Fit
Penilaian	A priori	46,66	27	0,011	0,036	Tidak
Kinerja Guru	Revisi	20,66	26	0,759	0,000	Fit

Tabel 2 menunjukkan bahwa hasil *first order* CFA model a priori (model awal *Congeneric*) menunjukkan dua model fit dengan data, yaitu dimensi PKB Guru,  $\chi^2(p=0,519 < 0,05)$  dan Supervisi,  $\chi^2(p=0,283 < 0,05)$  dan dua model tidak fit dengan data, yaitu dimensi Bisi Belajar,  $\chi^2(p=0,000 < 0,05)$  dan Penilaian Kinerja Guru,  $\chi^2(p=0,011 < 0,05)$ . Namun setelah dilakukan revisi dengan membebaskan *error variance* berkorelasi antara pasangan item7 dan item8, dimensi visi fit dengan data,  $\chi^2(p=0,286 < 0,05)$  serta membebaskan *error variance* berkorelasi antara pasangan item31 dan item32, dimensi penilaian kinerja guru fit dengan data,  $\chi^2(p=0,759 < 0,05)$ . Semua item yang merefleksikan masing-masing dimensi memiliki nilai t muatan faktor lebih besar dari nilai kritis (nilai  $t > 1,96$ ) atau nilai muatan faktor standar  $\geq 0,70$ . Dengan demikian, keempat model pengukuran masing-masing dimensi telah memenuhi syarat unidimensi (validitas konstruk) dan semua (36) item layak menjadi alat ukur.

Model selanjutnya diuji pada level lebih tinggi, yaitu pengujian *second order* CFA. Dalam analisis *second order* CFA, keempat dimensi hasil revisi *first order* CFA sebagai indikator dan konstruk umum kepemimpinan pembelajaran sebagai domain target ukur. Analisis awal, model menunjukkan bahwa *error variance* dimensi Visi mengandung nilai negatif (*Haywoodcases*) dan variabel laten belum memiliki skala, *unit of measurement*. *Haywoodcases* ini dapat diatasi dengan mengubah nilai negatif menjadi positif pada *error variance* pada dimensi Visi. Sedangkan pemberian skala variable laten dapat dilakukan dengan menetapkan koefisien hubungan antara salah satu indikator

variable laten dengan angka 1 dan model dianalisis ulang.

Hasil *second order* CFA satu faktor kepemimpinan pembelajaran menghasilkan model fit dengan data empiris, yaitu  $\chi^2$  dengan  $p\text{-value}=0,3474 > 0,05$  dan nilai RMSEA=  $0,006 < 0,08$ . Dimensi Supervisi memiliki nilai muatan faktor *second order* yang paling baik ( $\gamma=0,87$ ) disusul oleh muatan faktor dimensi Penilaian ( $\gamma=0,77$ ), muatan faktor dimensi Visi ( $\gamma=0,75$ ) dan muatan faktor dimensi PKB ( $\gamma=0,72$ ). Keempat dimensi ini memiliki nilai t muatan faktor lebih besar dari nilai kritis (nilai  $t > 1,9$ ). Keempat dimensi *first order* hanya mengukur konstruk umum satu faktor kepemimpinan pembelajaran kepala sekolah. Dengan demikian dapat disimpulkan bahwa model asesmen kepemimpinan pembelajaran dengan 4 dimensi dan 36 indikator telah terbukti memenuhi syarat validitas konstruk.

Konstruk model asesmen kepemimpinan pembelajaran kepala sekolah pendidikan dasar telah teruji memenuhi validitas dan reliabilitas. Analisis selanjutnya adalah untuk menguji signifikansi hubungan antara model asesmen masing-masing grup rater dengan ujian nasional (UN). Hasil analisis ini disajikan pada Tabel 3.

Tabel 3. Hubungan antara Model Grup Rater dengan Ujian Nasional (UN)

Model Grup Rater-UN	r	p	Kesimpulan
<i>Self</i> -UN	0,373	0,010 < 0,05	Signifikan
<i>Supervisor</i> -UN	0,214	0,148 > 0,05	Tidak Signifikan
<i>Subordinate</i> -UN	0,425	0,003 < 0,05	Signifikan

Tabel 3 menunjukkan bahwa hubungan model grup *supervisor* dengan UN adalah tidak signifikan ( $p=0,149>0,05$ ), sebaliknya hubungan model grup *self* dengan UN ( $p=0,010<0,05$ ) dan *subordinate* dengan UN ( $p=0,003<0,05$ ) adalah signifikan. Dengan demikian, dapat disimpulkan bahwa model asesmen grup *self* (kepala sekolah) dan *subordinate* (guru) memiliki peluang sebagai prediktor yang baik terhadap prestasi siswa. Sementara model asesmen grup *supervisor* (pengawas) memiliki peluang sebagai prediktor yang buruk terhadap prestasi siswa. Berdasarkan besarnya kekuatan hubungan, nilai koefisien korelasi model *subordinate*-UN ( $r=0,425$ ) lebih besar dari pada model *self*-UN ( $r=0,373$ ). Dengan demikian, model asesmen grup *subordinate* (guru) memiliki peluang sebagai prediktor paling baik terhadap prestasi siswa. Artinya, model pengukuran kepemimpinan pembelajaran berdasarkan respon grup guru lebih akurat sebagai prediktor keberhasilan prestasi siswa dibandingkan dengan grup kepala sekolah dan grup pengawas.

Analisis validitas konstruk yang dibahas sebelumnya adalah analisis konstruk model asesmen dengan pendekatan *360-degree assessment*, dari keseluruhan grup rater yakni asesmen dari *self* (kepala sekolah), *supervisor* (pengawas) dan *subordinate* (guru). Analisis berikut ini adalah analisis validitas konstruk model asesmen kepemimpinan pembelajaran secara terpisah yakni model asesmen dari masing-masing grup rater.

Tabel 4. Hasil CFA setiap Grup Rater

Grup Rater	$\chi^2$	p-value	RMSEA	Fit
<i>Self</i>	59,26	0,174>0,05	0,063<0,08	Fit
<i>Supervisor</i>	73,96	0,012<0,05	0,105>0,08	Tidak Fit
<i>Subordinate</i>	57,72	0,184>0,05	0,062<0,08	Fit

Tabel 4 menunjukkan bahwa model pengukuran grup *subordinate* dan grup *self* adalah fit dengan data empiris, secara berturut-turut nilai  $\chi^2$ ,  $p=0,184>0,05$  dan  $0,17>0,05$  serta RMSEA=  $0,062<0,08$  dan  $0,63$

$<0,08$ . Model pengukuran grup *supervisor* tidak fit dengan data karena nilai  $\chi^2$ ,  $p=0,012<0,05$  dan nilai RMSEA=  $0,105>0,08$ . Dengan demikian dapat disimpulkan bahwa kedua model pengukuran grup *self* dan grup *subordinate* memiliki validitas konstruk, sebaliknya model pengukuran grup *supervisor* tidak memiliki validitas konstruk.

Analisis konstruk model pengukuran ketiga grup rater dilanjutkan pada level konvergensi variabel laten terhadap setiap dimensi.

Tabel 5. Hasil Analisis Variansi Variabel Laten antar Grup Rater

Variabel Laten	<i>Self</i>		<i>Supervisor</i>		<i>Subordinate</i>		
	$\lambda_1$	$\lambda_2$	$t_2$	$\delta_2$	$\lambda_3$	$t_3$	$\delta_3$
Visi	1,00	0,44	1,52	0,91	0,64	2,09*	0,81
PKB	1,00	0,17	0,61	0,99	0,82	3,57*	0,60
Supervisi	1,00	0,38	1,37	0,97	0,69	2,35*	0,76
Penilaian	1,00	0,28	1,12	0,96	1,01	3,23*	0,49

\*t-value >1,96, estimasi *factor loading* adalah signifikan

Tabel 5 menunjukkan bahwa semua keempat variabel laten grup *subordinate* memiliki nilai t signifikan karena nilai  $t>1,96$ , yakni nilai t variabel laten visi=2,09, PKB=3,57, supervisi=2,35 dan penilaian=3,23. Sebaliknya semua keempat variabel laten grup *supervisor* memiliki nilai t tidak signifikan karena nilai  $t<1,96$ , yakni nilai t variabel laten visi=1,52, PKB=0,61, supervisi=1,37 dan penilaian=1,12. Dengan demikian, dapat disimpulkan bahwa keempat variabel laten model asesmen berdasarkan respon grup guru adalah konvergen terhadap konstruk kepemimpinan pembelajaran yang menjadi tujuan ukur. Sementara, keempat variabel laten model asesmen berdasarkan respon grup pengawas adalah tidak konvergen terhadap konstruk yang menjadi tujuan ukur.

Hasil analisis dapat dirangkum bahwa model pengukuran dengan tiga grup rater secara keseluruhan memenuhi validitas dan reliabilitas konstruk, akan tetapi secara terpisah model pengukuran grup *supervisor* tidak memenuhi syarat validitas konstruk sementara model pengukuran grup *self* dan *subordinate* memenuhi syarat validitas konstruk.

Kedua model pengukuran grup *self* dan grup *subordinate* telah teruji memenuhi syarat validitas, namun demikian masih perlu diuji apakah kedua grup memiliki parameter model yang invariant (ekuivalen). Dalam analisis invariansi parameter kedua grup, ada tiga model untuk dibandingkan: model A, model B, dan model C. Model A sebagai *baseline* atau acuan dimana di dalamnya 12 *factor loadings*, 12 *error variances* dan 4 *factor correlations* ditetapkan *equivalent* antar kedua grup. Model B adalah model dengan semua parameter *factor loading* kedua grup dibebaskan berbeda ( $\lambda_1 \neq \lambda_2$ ), dan model C adalah model dengan semua parameter *error variance* kedua grup dibebaskan berbeda ( $\delta_1 \neq \delta_2$ ). Analisis invariansi parameter model kedua grup adalah untuk membandingkan ( $\chi^2$  dan *df*) model B dengan model A dan model C dengan model A. Hasil analisis invariansi parameter model kedua grup disajikan seperti Tabel 6.

Tabel 6. Hasil Analisis Invariansi Parameter Model antara Grup *Self* dan *Subordinate*

Model	$\chi^2$	df	$\Delta\chi^2$	$\frac{\Delta d}{f}$	p-value
A ( <i>Baseline</i> )	283,51	127	-	-	0,0000
B ( $\lambda_1 \neq \lambda_2$ )	256,32	115	27,19	12	0,0073
C ( $\delta_1 \neq \delta_2$ )	132,67	115	150,84	12	0,0000

Tabel 6 menjelaskan bahwa model B dan model A (*baseline*) yang memiliki perbedaan nilai  $\Delta\chi^2 = 27,19$  dan  $\Delta df = 12$  menghasilkan p-value = 0,0073 < 0,05 berbeda secara signifikan. Sedangkan model C dan model A yang memiliki perbedaan nilai  $\Delta\chi^2 = 150,84$  dan  $\Delta df = 12$  menghasilkan p-value = 0,0000 < 0,05 juga berbeda secara signifikan. Dengan demikian, dapat disimpulkan bahwa semua parameter *factor loading* ( $\lambda$ ) dan *error variance* ( $\delta$ ) antara grup *self* dan grup *subordinate* adalah non invariant. Perbedaan parameter model antar kedua grup *self* dan grup *subordinate* dipaparkan pada Tabel 7.

Tabel 7. Perbedaan Parameter antara Model Grup *Self* dan Grup *Subordinate*

Variabel	<i>Self</i> (Kepsek)		<i>Subordinate</i> (Guru)	
	$\lambda_1$	$\delta_1$	$\lambda_2$	$\delta_2$
Visi				
• Indikator1	0,83	0,50	0,69	0,32
• Indikator2	0,82	0,65	0,73	0,14
• Indikator3	0,44	0,98	0,64	0,42
PKB				
• Indikator4	0,77	0,63	0,77	0,19
• Indikator5	0,84	0,42	0,93	0,00
• Indikator6	0,87	0,58	0,66	0,23
Supervisi				
• Indikator7	1,05	0,38	0,66	0,07
• Indikator8	0,96	0,46	0,76	0,05
• Indikator9	1,03	0,21	0,82	0,07
Penilaian				
• Indikator10	0,94	0,58	0,69	0,06
• Indikator11	1,09	0,14	0,78	0,06
• Indikator12	0,95	0,55	0,70	0,07

Tabel 7 menjelaskan bahwa 12 indikator sebagai alat ukur konstruk umum kepemimpinan pembelajaran kepala sekolah pendidikan dasar memiliki parameter yang berbeda secara signifikan antara grup *self* dan grup *subordinate*. Semua *factor loadings* keduabelas indikator berbeda antara grup kepala sekolah ( $\lambda_1$ ) dan grup guru ( $\lambda_2$ ). Di samping itu, semua *error variances* keduabelas indikator pada grup guru ( $\delta_2$ ) jauh lebih kecil dibandingkan dengan *error variances* pada grup kepala sekolah ( $\delta_1$ ). Rentang nilai *error variances* keduabelas indikator pada grup guru dari 0,00 s.d 0,32 (rerata  $\delta_2 = 0,14$ ), sementara rentang nilai *error variances* pada grup kepala sekolah dari 0,14 s.d 0,98 (rerata  $\delta_1 = 0,51$ ). Hal ini menunjukkan bahwa model asesmen berdasarkan respon guru (*upward assessment*) lebih reliabel dibandingkan dengan model asesmen berdasarkan respon diri sendiri kepala sekolah. Model asesmen kepemimpinan pembelajaran kepala sekolah oleh grup guru disebut juga model *upward assessment*.

Berdasarkan semua analisis di atas dapat ringkas bahwa model asesmen grup *self*, *supervisor* dan *subordinate* secara terpadu memenuhi syarat validitas dan reliabilitas. Namun, secara terpisah model asesmen

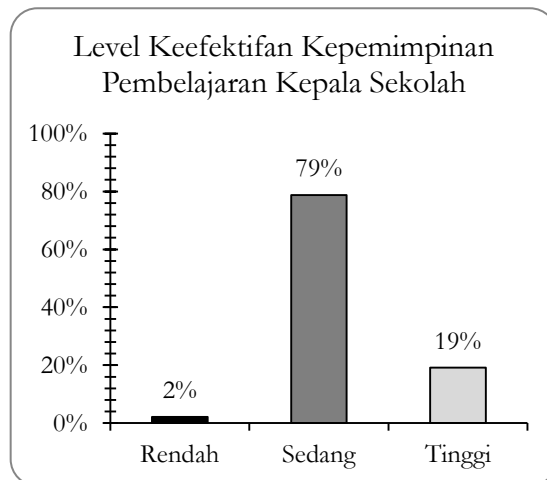
grup *supervisor* tidak memenuhi syarat validitas konstruk, sementara model asesmen grup *self* dan grup *subordinate* memenuhi syarat validitas konstruk. Meskipun model asesmen grup *self* dan grup *subordinate* telah memenuhi syarat validitas konstruk, akan tetapi kedua grup tersebut memiliki reliabilitas berbeda, yaitu reliabilitas model asesmen grup *subordinate* lebih tinggi pada semua dimensi dibandingkan dengan model asesmen grup *self*.

Model asesmen berdasarkan respon guru adalah model asesmen yang memiliki kehandalan tinggi sehingga dapat dipercaya untuk menyediakan data yang paling tepat dan cermat tentang perilaku kepemimpinan pembelajaran kepala sekolah pendidikan dasar. Model asesmen berdasarkan respon guru dapat dijadikan dasar yang paling kuat untuk pengambilan keputusan dan untuk program pengembangan keprofesian kepala sekolah secara berkelanjutan.

Model asesmen yang telah memenuhi syarat validitas dan reliabilitas dapat digunakan untuk menilai level keefektifan kepemimpinan pembelajaran kepala sekolah. Skor komposit dimensi dengan menjumlahkan atau merata-ratakan secara langsung semua skor indikator belum dapat dilakukan sebelum model paralel telah teruji. Apabila model paralel terbukti fit dengan data empiris, maka skor komposit masing-masing dimensi dengan penjumlahan semua skor indikator dalam dimensi yang sama dapat dilakukan. Hasil uji model paralel telah terbukti fit dengan data empiris sehingga dapat dilakukan penjumlahan semua skor indikator menjadi skor dimensi dan kemudian skor masing-masing dimensi tersebut dikonversi menjadi skala 100.

Penafsiran skor komposit masing-masing dimensi belum dapat dilakukan tanpa kriteria. Menurut Mardapi (2008, p.122), interpretasi skor hasil pengukuran dengan cara membandingkan skor dengan kriteria. Asesmen level keefektifan kepemimpinan pembelajaran dilakukan dengan membandingkan skor dimensi setiap individu kepala sekolah dengan kriteria ideal, yaitu skor level tinggi  $\geq 75$ , level sedang  $\geq 50$

dan level rendah  $< 50$ . Persentasi kepala sekolah yang berhasil mencapai masing-masing level tersebut disajikan pada Gambar 1.

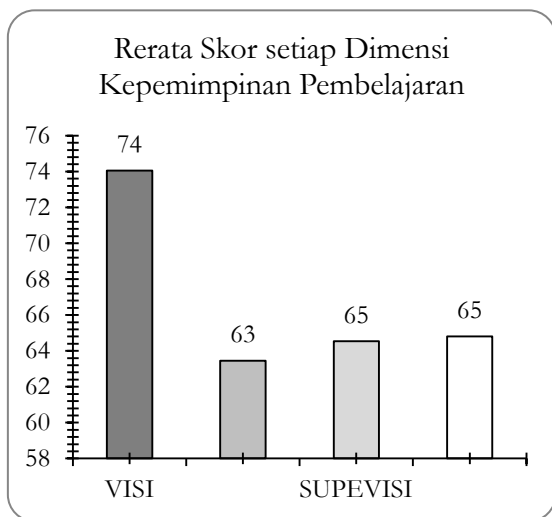


Gambar 1. Level Keefektifan Kepemimpinan Pembelajaran Kepala Sekolah

Gambar 1 menjelaskan bahwa persentasi kepala sekolah yang memiliki level keefektifan kepemimpinan pembelajaran pada kategori tinggi sebesar 19%, pada kategori sedang sebesar 79% dan pada kategori rendah sebesar 2%. Apabila harapan ideal bahwa semua kepala sekolah mencapai kategori tinggi, berarti sebanyak 81% kepala sekolah belum sesuai dengan harapan. Sebagai tindak lanjut perbaikan, sebanyak 81% kepala sekolah harus meningkatkan keefektifan kepemimpinan pembelajaran ke level tinggi.

Analisis berikutnya diperlukan untuk mendeskripsikan kekuatan dan kelemahan setiap dimensi kepemimpinan pembelajaran. Hasil analisis kekuatan dan kelemahan kepala sekolah disajikan rerata skor setiap dimensi seperti pada Gambar 2. Gambar 2 menjelaskan bahwa level keefektifan perilaku kepemimpinan pembelajaran kepala sekolah relatif tinggi pada dimensi visi belajar sebesar 74 dan relatif rendah pada dimensi PKB guru, supervisi pembelajaran dan penilaian kinerja guru secara berturut-turut sebesar 63, 65 dan 65. Dengan kata lain, mayoritas kepala sekolah memiliki kekuatan pada dimensi pengembangan visi, akan tetapi memiliki kelemahan pada dimensi PKB guru,

supervisi dan penilaian kinerja guru. Berdasarkan informasi kekuatan dan kelemahan ini, program tindak lanjut peningkatan level keefektifan kepemimpinan pembelajaran kepala sekolah perlu difokuskan pada dimensi-dimensi yang relatif rendah, yaitu PKB guru, supervisi dan penilaian kinerja guru. Ketiga dimensi ini merupakan faktor-faktor kunci dalam perbaikan mutu pembelajaran.



Gambar 2. Rerata Skor setiap Dimensi Kepemimpinan Pembelajaran Kepala Sekolah

Pembahasan tentang semua temuan hasil penelitian ini meliputi: konstruk model asesmen dengan pendekatan *360-degree feedback*, konstruk model asesmen masing-masing grup rater, hubungan model masing-masing grup rater dengan prestasi siswa, dan keefektifan kepemimpinan pembelajaran kepala sekolah. Konstruk model asesmen dengan pendekatan *360-degree feedback* cocok dengan data empiris atau memenuhi validitas konstruk setelah *error variance* dua pasangan item dibebaskan berkorelasi. Korelasi *error* pasangan item ini dapat dijelaskan bahwa sebagian varian bersama pasangan item tersebut disebabkan oleh faktor yang menjadi tujuan ukur dan sebagian varian bersama pasangan item tersebut disebabkan faktor luar. Dua pasangan item memiliki konten yang sama, yaitu tentang “strategi pembelajaran” dan “evaluasi belajar”. Sebagian besar responden memandang kedua

item tersebut memiliki konten yang sama sehingga merespon dengan cara yang sama.

Konstruk model asesmen seluruh grup rater (*360-degree feedback*) memiliki validitas bukti konstruk dan reliabilitas, akan tetapi setelah diuji secara terpisah masing-masing grup rater memiliki validitas dan reliabilitas berbeda. Model asesmen grup pengawas tidak akurat untuk menilai perilaku kepemimpinan pembelajaran kepala sekolah karena tiga alasan: (1) model grup pengawas tidak memenuhi validitas konstruk, (2) keempat variabel laten model grup pengawas tidak konvergen terhadap konstruk yang menjadi tujuan ukur, dan (3) model grup pengawas tidak memiliki korelasi signifikan dengan prestasi siswa. Penjelasan tentang ketidakakuratan model asesmen grup pengawas antara lain, pengawas cenderung paling *lenient* (murah) di antara ketiga grup rater. Pengawas cenderung memberikan skor tinggi yang tidak sesuai dengan kenyataan, tidak sesuai dengan yang dialami oleh kepala sekolah dan guru. Kemungkinan pengawas tidak banyak mengetahui tentang indikator-indikator perilaku kepemimpinan pembelajaran kepala sekolah dibandingkan dengan kepala sekolah dan guru.

Berbeda dengan model pengawas, model grup kepala sekolah dan grup guru layak menjadi model asesmen kepemimpinan pembelajaran kepala sekolah karena model kedua grup: (1) memenuhi validitas konstruk, (2) keempat variabel laten model kedua grup konvergen terhadap konstruk tujuan ukur, dan (3) model kedua grup memiliki korelasi signifikan dengan prestasi siswa. Kedua grup cenderung memberikan skor yang sesuai dengan kenyataan, sesuai dengan yang dialaminya.

Namun demikian, model asesmen kedua grup kepala sekolah dan guru memiliki kualitas berbeda. Model asesmen grup guru lebih layak dibandingkan dengan model asesmen grup kepala sekolah karena model asesmen grup guru memiliki korelasi lebih tinggi dengan prestasi siswa dibandingkan dengan model grup kepala sekolah dan model asesmen grup guru lebih reliabel dibandingkan dengan model grup kepala sekolah. Model

asesmen grup guru memiliki eror jauh lebih kecil dibandingkan dengan model asesmen grup kepala sekolah terhadap semua indikator perilaku kepemimpinan pembelajaran kepala sekolah. Besarnya eror pada model grup kepala sekolah kemungkinan disebabkan sebagian kepala sekolah dipengaruhi oleh *social desirability*, ingin menunjukkan citra baik sehingga merespon yang tidak semestinya.

Berbeda dengan model grup kepala sekolah, kecilnya eror pada model grup guru karena kecil kemungkinan dipengaruhi oleh *social desirability*, mayoritas guru lebih merespon sesuai kenyataan dari pada ingin menunjukkan citra baik. Di samping itu, guru menginginkan kepemimpinan pembelajaran kepala sekolah semakin baik karena berdampak langsung terhadap keprofesian guru.

### Simpulan dan Saran

Berdasarkan hasil analisis data dan pembahasan, hasil penelitian ini dapat disimpulkan sebagai berikut.

Pertama, model asesmen kepemimpinan pembelajaran dengan 4 dimensi dan 12 indikator telah teruji secara teori dan empiris. Keempat dimensi asesmen kepemimpinan pembelajaran meliputi dimensi Visi Belajar, Pengembangan Keprofesian Berkelanjutan guru, Supervisi Pembelajaran dan Penilaian Kinerja Guru. Dimensi Visi Belajar memiliki 3 indikator: visi dan tujuan bersama, komitmen pada mutu pembelajaran, dan dukungan sumber belajar. Dimensi Pengembangan Keprofesian Berkelanjutan (PKB) guru memiliki tiga indikator: program pengembangan, *peer training*, dan *peer evaluation*. Dimensi supervisi memiliki tiga indikator: praobservasi, observasi, paska observasi. Dimensi Penilaian Kinerja Guru memiliki tiga indikator: prapenilaian, penilaian, dan paska penilaian.

Kedua, model asesmen berdasarkan respon guru dan kepala sekolah memiliki bukti validitas konstruk, sementara model asesmen berdasarkan respon grup pengawas tidak memiliki bukti validitas konstruk. Ketiga, model asesmen berdasarkan respon guru dan kepala sekolah memiliki peluang sebagai prediktor yang baik terhadap pres-

tasi siswa, sementara model asesmen berdasarkan respon pengawas memiliki peluang sebagai prediktor yang kurang baik terhadap prestasi siswa. Keempat, model asesmen berdasarkan respon guru lebih reliabel dibandingkan dengan model asesmen berdasarkan respon kepala sekolah. Kelima, persentase kepala sekolah yang mencapai level keefektifan kepemimpinan pembelajaran pada kategori sedang dan rendah sebesar 81%, dan dengan skala 100, rerata skor dimensi visi, supervisi, penilaian kinerja guru, dan pengembangan keprofesian guru berturut-turut adalah sebesar 74, 65, 65 dan 63.

Saran tindak lanjut berdasarkan kesimpulan tersebut adalah sebagai berikut. Pertama, pemerintah daerah dan pusat disarankan untuk meningkatkan keefektifan kepemimpinan pembelajaran kepala sekolah sebanyak 81% pada kategori rendah dan sedang ke level kategori tinggi dengan memfokuskan pada dimensi yang rendah, yakni PKB, supervisi dan penilaian. Kedua, pemerintah daerah dan pusat disarankan untuk menggunakan model asesmen kepemimpinan pembelajaran kepala sekolah karena validitas dan reliabilitas telah teruji. Ketiga, kepala sekolah disarankan untuk memanfaatkan hasil asesmen sebagai dasar pengembangan keprofesian kepemimpinan pembelajaran. Keempat, pengawas sekolah disarankan menggunakan model asesmen berdasarkan respon guru untuk pengambilan keputusan.

### Daftar Pustaka

- Asia Society Partnership for Global Learning. (2012). *Teaching and leadership for the twenty-first century: The International Summit on The Teaching Profession*.
- Blasé, J. & Blasé, J. (2004). *Handbook of instructional leadership*. California: Corwin Press.
- Bush, T. & Glover, D. (2003). *School leadership: Concept and evidence*. Nottingham: National College for School Leadership.

- Cennamo, K., & Kalk, D. (2005). *Real world instructional design*. Thomson Learning.
- DuFour, R. & Berkey, T. (1995). The principal as staff developer. *Journal of Staff Development, Fall 1995 (Vol. 16, No. 4)*
- Gaziel, H. (2007). *Re-examining the relationship between principals' instructional leadership and student achievement*. ©Kamla-Raj 2007 J. Soc. Sci.15(1): 17-24 (2007)
- Glickman, C.D., Gordon, S.P., & Gordon, J.S. (2007). *Supervision and instructional leader*. Boston: Pearson Education.
- Goldring, E. et al. (2007). *Assessing learning-center leadership: Connection to research, professional standard, and current practice*. The Wallace Foundation.
- Hair, J. F., et.al. (1998). *Multivariate data analysis*. Prentice Hall International: UK.
- Hallinger, P., & Heck, R.H. (1998). *Exploring the principal contribution to school effectiveness: 1980-1995*. School Effectiveness and School Improvement 9,2: 157-191.
- Hallinger, P. (2008). *Methodologies for studying school leadership: A review of 25 years of research using the principal instructional management rating scale*. Paper prepared for presentation at the annual meeting of American Educational Research Association. New York.
- Hallinger, P., & Murphy, J. (1985). Assessing the instructional management behavior of principals. *Elementary School Journal*.
- Heslin, P.A., & Latham, G.P. (2004). The effect of upward feedback on managerial behavior. *Applied Psychology: An International Review*,53(1),23-37
- Holifield, M. & Cline, D. (1997). Clinical supervision and its outcome: Teachers and principals report. *National Association of Secondary School Principals, NASSP Bulletin*; Sep 1997; 81, 590; Research Library pg. 109
- Kemendiknas. (2010). *Peraturan Menteri Pendidikan Nasional Nomor 28 Tahun 2010 tentang Penugasan Guru sebagai Kepala Sekolah/Madrasah*
- Kemdiknas. (2010). Permendiknas No. 35 tahun 2010 tentang Petunjuk Teknis Pelak. JabFung Guru dan Angka Kreditnya
- Kemdiknas. (2010). Permendiknas & KaBKN No 03/V/PB/2010 dan No 14 tahun 2010 tentang Juklak Jabatan Fungsional & Angka Kreditnya
- Leithwood, K. et.al. (2004). *How Leadership Influences Student Learning: Review of Research*. New York: The Wallace Foundation.
- Lunerburg, Fred C. (2010). *The Principal as Instructional Leader*. National Forum of Educational and Supervision Journal, Volume 27, Number 4
- Mardapi, Djemari. (2008). *Teknik penyusunan instrumen tes dan nontes*. Jogjakarta. Mitra Cendikian Press.
- Marzano, R.J. (2012). *Marzano school leadership evaluation model*. Learning Sciences Marzano Center for Teacher and Leadership Evaluation.
- McEwan, A.K. (2003). *Seven steps to effective instructional leadership*. California: A Sage Publication Company, Crowin Press, Inc.
- Murphy, J., et al. (2006). *Learning-Center Leadership: A Conceptual Foundation*. The Wallace Foundation.
- Olson, L. (2008). *Assessment to Rate Principal Leadership to Be Field-Tested*. Education Week, Focus on Leadership & Management. Publish Online: January 11, 2008. Publish in Print: January 16, 2008.
- Reeve, D.B. (2005). *Assessing educational leaders: Evaluating performance for improved individual and organizational result*. Thousand Oaks, CA: Crowin Press.

- Robinson, V. J. M., Lloyd, C.A., & Rowe, K. J. (2009). *School Leadership and Student Outcomes: Identifying What Works and Why: Best Evidence Synthesis Iteration*. New Zealand: Ministry of Education.
- Science Education Quality Improvement Project (SEQIP). (2002). *Proyek Kerjasama Bilateral antara pemerintah Indonesia dan Jerman*.
- Shelton, S. (2010, May). *Strong Leaders Strong Schools: 2009 State Laws*. Colorado and Washington, D.C.: National Conference of State Legislatures. The Forum for America's Ideas
- Schumacker, Randall E. & Lomax, Richard G. (2004). *A Beginner's Guide to Structural Equation Modeling*. Lawrence Erlbaum Associate. Second Edition
- Smith, Stuart C., & Piele, Philip K. (2006). *School Leadership: Handbook for Excellence in Student Learning*. California, A Sage Publication Company: Corwin Press.
- Strong, J.H. (2006). *Evaluating teaching*. California: Corwin Press.
- Sullivan, S. & Glanz, J. (2005). *Supervision that improve teaching*. California: Corwin Press.
- Sweeney, D. (2003). *Professional Development by and for Teacher: Learning Along the Way*. Maine, Portland: Stenhouse Publisher
- Witziers, B., Bosker, R., & Kruger, M. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, 39(3), 398–425.