

PENGEMBANGAN MODEL PENILAIAN KETERAMPILAN INTRAPRIBADI DAN ANTARPRIBADI DALAM PENDIDIKAN KARAKTER DI SEKOLAH DASAR

¹⁾Akif Khilmiyah, ²⁾Sumarno, ³⁾Darmiyati Zuchdi

¹⁾Universitas Muhammadiyah Yogyakarta, ^{2,3)}Universitas Negeri Yogyakarta

¹⁾khilmiyahakif@yahoo.com, ²⁾sumarno_unj@yahoo.co.uk, ³⁾darmiyati@uny.ac.id

Abstrak

Penelitian ini bertujuan untuk mengembangkan model penilaian keterampilan intrapribadi dan antarpribadi dalam pendidikan karakter di SD yang valid, reliabel, dan cocok secara statistik. Penelitian ini menggunakan model *research and development* dari Plomp dengan lima fase, yaitu: investigasi, desain, realisasi, tes, dan implementasi. Subjek penelitian yaitu siswa kelas V SD di DI Yogyakarta. Pengumpulan data menggunakan kuesioner, dokumentasi, wawancara, FGD, dan observasi. Analisis data menggunakan analisis deskriptif, formula *Aiken's V*, *EFA* dan *CFA*. Simpulan penelitian ini yaitu konstruk penilaian keterampilan intrapribadi dan antarpribadi dalam pendidikan karakter telah memenuhi syarat validitas, reliabilitas, dan *goodness of fit statistic*; dan panduan penilaian keterampilan intrapribadi dan antarpribadi dalam pendidikan karakter di SD dinilai komprehensif, mudah, dan praktis oleh para guru selaku pengguna.

Kata kunci: *keterampilan intrapribadi, keterampilan antarpribadi, pendidikan karakter.*

DEVELOPING OF ASSESSMENT MODEL OF INTRAPERSONAL AND INTERPERSONAL SKILLS BASED ON CHARACTER EDUCATION IN ELEMENTARY SCHOOLS

¹⁾Akif Khilmiyah, ²⁾Sumarno, ³⁾Darmiyati Zuchdi

¹⁾Universitas Muhammadiyah Yogyakarta, ^{2,3)}Universitas Negeri Yogyakarta

¹⁾khilmiyahakif@yahoo.com, ²⁾sumarno_unj@yahoo.co.uk, ³⁾darmiyati@uny.ac.id

Abstract

The objectives of this study is to develop the assessment model of intrapersonal and interpersonal skills in character education of elementary school students that fit into the criteria of validity, reliability and goodness of fit statistic. This study used plomp's research and development model which consisted of five phases, namely: investigation, design, realization, testing, and implementation. The subjects of this study were the fifth grade students of the elementary schools in Yogyakarta Special Region. The data were collected using a questionnaire, documentation, interview, focus group discussion, and observation, and analyzed using the descriptive analysis, formula aiken's V, Exploratory FactorAnalysis (EFA), and Confirmatory Factor Analysis (CFA). The results of this study are as follows, (1) the assessment construct of intrapersonal and interpersonal skills based on character education for elementary school students has fulfilled the criteria of validity, reliability, and the goodness of fit statistic.

Keywords: *intrapersonal skills, interpersonal skills, character education*

Pendahuluan

Tujuan sekolah dalam mendidik karakter adalah agar siswa menjadi pribadi yang unggul secara menyeluruh, yakni pribadi yang cerdas secara intelektual, emosional, sosial, dan spiritual. Penanaman pendidikan karakter ini harus dimulai dari jenjang pendidikan yang paling rendah yakni Sekolah Dasar (SD). Pada jenjang SD, anak-anak masih bisa dibentuk dengan mudah dan diperkenalkan dengan berbagai karakter positif, agar tumbuh menjadi pribadi yang memiliki pola pikir yang baik, kemauan baik, dan perilaku yang baik sebagai bekal kelak dalam mengarungi kehidupan. Pembentukan pribadi yang unggul secara menyeluruh ini, telah tercantum dalam pasal 3, UU nomor 20/2003 tentang Sistem Pendidikan Nasional (Kemendiknas, 2011, p.243). Tujuan untuk membentuk pribadi yang memiliki kecerdasan secara menyeluruh ini, selaras dengan Permendiknas nomor 22 dan 23/2006 tentang Standar Isi dan SKL, Inpres nomor 1/2010, tentang Percepatan Pelaksanaan Prioritas Pembangunan Nasional Tahun 2010 yang memerintahkan perlunya pengembangan karakter peserta didik melalui pendidikan di sekolah.

Namun, berdasarkan survei peneliti pada beberapa SD di Yogyakarta, ditemukan fakta bahwa 70% kecerdasan dan penilaian yang banyak dikembangkan di SD adalah kecerdasan intelektual melalui penilaian kognitif. Perhatian dan pemahaman para guru akan pentingnya pengembangan kecerdasan emosional, sosial, dan spiritual masih sangat rendah dalam proses pembelajaran dan penilaian. Penilaian tentang karakter siswa hanya didasarkan pada pengamatan guru selama di sekolah saja. Penilaian khusus tentang keterampilan intrapribadi dan antarpribadi di SD belum ada. Model penilaian yang digunakan juga bervariasi dan biasanya dikembangkan sendiri oleh guru agama. Kebanyakan guru beranggapan bahwa pembentukan karakter dan akhlak siswa lebih dekat dengan muatan materi pelajaran agama.

Hal ini didukung pula oleh hasil penelitian Zuchdi (2010, p.68), bahwa praktik

penilaian hasil belajar di sekolah, khususnya di SD, sarat dengan penilaian kognitif semata. Penghargaan pada siswa ditunjukkan guru melalui pemberian ranking dan nilai ujian. Meskipun semua pendidik tahu bahwa ranah pendidikan itu ada kognitif, afektif, dan psikomotor (perilaku), namun pada praktiknya penilaian ranah afektif dan psikomotor (perilaku) belum mendapat porsi perhatian yang memadai. Kondisi ini terjadi karena kurangnya kemampuan pendidik untuk mendeskripsikan indikator capaian ranah afektif, sehingga penilaian yang dilakukan kepada siswa tidak menggambarkan kemampuan siswa secara menyeluruh. Akibatnya pendidik tidak mampu melakukan pembinaan aspek afektif yang meliputi kecerdasan emosional, sosial, dan spiritual, yang termanifestasikan dalam keterampilan intrapribadi (*intrapersonal*) dan antarpribadi (*interpersonal*). Padahal kedua keterampilan ini merupakan inti dari pendidikan karakter yang harus dimiliki siswa, sebagai hasil dari proses pembelajaran di sekolah.

Pentingnya pendidikan karakter yang termanifestasikan dalam keterampilan intrapribadi dan antarpribadi memberikan pengaruh yang besar pada kesuksesan seseorang dalam kehidupan. Hal itu, sebagaimana hasil penelitian Goleman (1995, p.15), Gardner (1996, p.26), dan Bar-On (1988, p.24; 1997, p.123; 2000, p.46; 2005, p.94), Berkowitz, (2002, p.23) yang menemukan bahwa 80% penentu kesuksesan seseorang adalah kecerdasan emosional. Demikian pula hal itu sesuai dengan hasil penelitian Zamroni, yang menemukan bahwa 70% orang sukses dalam kehidupannya bukan karena kecerdasan intelektual dan vokasional, tetapi justru karena kecerdasan generik (2005, p.49). Kemampuan generik ini ditunjukkan melalui: (1) kemampuan personal (kemampuan mengambil keputusan secara rasional, kemampuan merancang masa depan, dll), (2) kemampuan sosial (kemampuan bekerja sama, kemampuan memahami orang lain, dll).

Hal ini diperkuat pula dari hasil penelitian pendahuluan yang dilakukan oleh peneliti pada SD yang menjadi *piloting* pen-

didikan karakter di Bantul, tahun 2011 dan 2012 diperoleh kesimpulan bahwa 80% siswa mengalami permasalahan rendahnya keterampilan intrapribadi yang terlihat dari rendahnya sikap kejujuran, rasa percaya diri, dan kemandirian. Sebanyak 78% siswa mengalami rendahnya keterampilan antarpribadi yang ditunjukkan dengan rendahnya sikap kerja sama, tanggung jawab, kemampuan mengatasi konflik, dan rendahnya sikap toleransi antarsiswa (Khilmiyah, 2011, p.65; 2012, p.62). Kondisi ini menunjukkan masih rendahnya keterampilan intrapribadi dan antarpribadi siswa SD dalam mengelola emosi diri sendiri dan kemampuan bersosialisasi dengan orang lain, yang berdampak pada munculnya sikap apatis, egois, dan inferior. Padahal menurut Zuchdi, (2010, p.69) keterampilan intrapribadi dan antarpribadi dapat mempengaruhi perkembangan kepribadian ke arah positif. Pengaruh positif tersebut antara lain berwujud dalam bentuk dapat menghargai orang lain, mampu menemukan alternatif pemecahan masalah, kreatif, sabar, dan mandiri.

Untuk mewujudkan pribadi siswa yang memiliki karakter terpuji, maka guru perlu melakukan penilaian keterampilan intrapribadi dan antarpribadi secara periodik agar diperoleh informasi lebih awal tingkat kekurangan capaian kemampuan karakter siswa yang harus segera dibenahi. Melalui penilaian karakter yang komprehensif, diharapkan guru mampu mengembangkan potensi siswa secara utuh, yang meliputi kemampuan intelektual ataupun kemampuan emosional, sosial, dan spiritual.

Berdasarkan uraian tersebut, maka penelitian ini dipandang penting dan strategis untuk dilakukan guna mengembangkan model penilaian keterampilan intrapribadi dan antarpribadi. Hasil penelitian ini dapat dimanfaatkan oleh guru untuk meningkatkan karakter siswa yang termanifestasikan dalam kecerdasan emosional, sosial, dan spiritual, karena selama ini belum tersedia acuan kriteria model penilaiannya.

Permasalahan penelitian ini terfokus pada tiga hal yakni: (1) mengkaji sejauhmana kelemahan model penilaian yang se-

lama ini digunakan untuk mendeteksi capaian pendidikan karakter di SD; (2) bagaimana mengembangkan model penilaian keterampilan intrapribadi dan antarpribadi dalam pendidikan karakter di SD yang komprehensif dan memenuhi kriteria valid, reliabel, dan *goodness of fit statistic*; (3) bagaimana mengembangkan panduan model Penilaian Keterampilan Intrapribadi dan Antarpribadi dalam pendidikan karakter di SD (PKIA-SD) yang komprehensif.

Tujuan penelitian ini adalah (1) untuk mengidentifikasi kelemahan model penilaian yang selama ini digunakan untuk mendeteksi capaian pendidikan karakter di SD; (2) untuk mengembangkan model penilaian keterampilan intrapribadi dan antarpribadi dalam pendidikan karakter di SD yang komprehensif dan memenuhi kriteria valid, reliabel, dan *goodness of fit statistic*; dan (3) untuk mengembangkan panduan model Penilaian Keterampilan Intrapribadi dan Antarpribadi dalam pendidikan karakter di SD (PKIA-SD) yang komprehensif.

Secara teoretis, temuan pengembangan model PKIA-SD ini dapat dijadikan rujukan untuk mengembangkan model penilaian kepribadian anak, pada aspek nilai-nilai pendidikan karakter yang lain secara lebih mendalam. Secara metodologis, hasil penelitian ini dapat dijadikan rujukan untuk pengembangan model penilaian keterampilan intrapribadi dan antarpribadi pada siswa SD tingkat rendah (kelas 1-3). Secara praktis, temuan penelitian ini dapat dipakai oleh guru kelas di SD tingkat tinggi (kelas 4-6), untuk melakukan penilaian capaian keterampilan intrapribadi dan antarpribadi dalam rangka melaksanakan pendidikan karakter, sehingga guru dapat melakukan pembinaan karakter lebih awal. Bagi pembuat kebijakan, khususnya penentu kebijakan terkait dengan pelaksanaan karakter dan akhlak mulia di SD/ MI, temuan penelitian ini dapat dijadikan model pembinaan kepada para guru SD/MI untuk melakukan penilaian capaian keterampilan intrapribadi dan antarpribadi dalam pelaksanaan pendidikan karakter secara komprehensif, yang mencakup ranah kognitif, afektif, dan psikomotorik.

Metode Penelitian

Jenis penelitian ini yaitu *research and development* dengan menggunakan model Plomp. Model ini terdiri dari lima fase, yaitu: investigasi, desain, realisasi, tes, dan implementasi (1997, p.5).

Penelitian ini dilaksanakan selama satu tahun mulai bulan Juni 2013. Tempat penelitiannya yaitu di Sekolah Dasar (SD) yang ada di Daerah Istimewa Yogyakarta (DIY).

Subjek penelitian ini adalah siswa kelas V SD di satu kota, dan tiga kabupaten di DIY. Jumlah subjek uji coba secara akumulatif sebanyak 518 siswa, yang terdiri dari subjek uji coba terbatas sebanyak 54 siswa dari 2 SD, dan subjek uji coba diperluas sebanyak 464 siswa dari 15 SD.

Pemilihan subjek uji coba penelitian ini dilakukan secara *purposive sampling* yang didasarkan pada pertimbangan usia anak (10-12 tahun) dan kemampuan melakukan penalaran moral secara otonom. Selain itu, dipertimbangkan berdasarkan variasi SD pada aspek: (1) perbedaan penerapan budaya pendidikan karakter di SD, (2) perbedaan sebagai SD piloting pendidikan karakter, (3) perbedaan agama, (4) perbedaan lama waktu belajar di sekolah, (5) perbedaan program pendidikan karakter di sekolah. Berdasarkan kriteria tersebut, dipilih subjek uji coba yang meliputi: 4 SD Negeri piloting pendidikan karakter dari Dinas Dikpora DIY, 2 SD Negeri bukan piloting pendidikan karakter, 1 SD Swasta bercorak kebangsaan, 4 SD Swasta berbasis Islam, 2 SD Swasta berbasis Katholik, 2 SD IT yang menerapkan *full day school*, 2 MI Negeri dan Swasta yang berada dalam pembinaan Kemenag RI.

Penelitian dan pengembangan ini dilakukan dalam dua tahap kegiatan, yaitu tahap prapengembangan dan tahap pengembangan. Kegiatan prapengembangan model PKIA-SD terdiri dari tiga tahap yaitu tahap investigasi, tahap desain, dan tahap realisasi. Tahap investigasi yaitu mengidentifikasi kelemahan model penilaian pendidikan karakter di SD dan mengkaji kebutuhan model PKIA-SD bagi guru dan siswa. Tahap desain yaitu kegiatan-kegiatan yang dilakukan

oleh peneliti dalam merancang perangkat model PKIA-SD, meliputi: merumuskan konstruk teori, mengembangkan indikator, mendeskripsikan indikator, menyusun item instrumen. Tahap realisasi terdiri dari kegiatan validasi melalui pakar dan praktisi, membuktikan validitas dan reliabilitas empirik, evaluasi hasil uji coba, revisi, membuktikan validitas dan estimasi reliabilitas konstruk model PKIA-SD.

Data kuantitatif diperoleh dari angket semi terbuka, sedangkan data kualitatif diperoleh dari hasil FGD, wawancara, dan observasi. Pengumpulan data melalui teknik dokumentasi, wawancara, observasi, FGD (*Focus Group Discussion*), dan kuesioner. Instrumennya berupa dokumen *software* dan *hardware*, pedoman wawancara, lembar pengamatan, dan angket semi terbuka.

Penelitian ini menggunakan analisis deskriptif kuantitatif dan kualitatif untuk mendeskripsikan data capaian keterampilan intrapribadi dan antarpribadi. Formula *Aiken's V*, dipakai untuk membuktikan validitas isi. Analisis metode *Cronbach's Alpha*, digunakan untuk mengetahui tingkat reliabilitas instrumen. Teknik *Exploratory Factor Analysis* (EFA), digunakan untuk mengetahui nilai *eigenvalue* yang terbentuk menjadi faktor atau dimensi dari variabel laten. Teknik analisis faktor konfirmatori orde dua (*Second Order Confirmatory Factor Analysis*) digunakan untuk menguji kecocokan model PKIA-SD (Joreskog & Sorbon, 1999, p.115).

Hasil Penelitian dan Pembahasan

Hasil Penelitian

Hasil investigasi awal yang dilakukan selama tiga bulan pada 17 SD di DIY, ditemukan adanya perbedaan variasi penerapan budaya pendidikan karakter pada masing-masing sekolah. Ada 7 macam variasi penerapan budaya pendidikan karakter di SD, sehingga ada 7 jenis SD, yaitu: (1) SD Negeri yang menjadi *piloting* pendidikan karakter, (2) SD Negeri yang bukan *piloting* pendidikan karakter, (3) SD Swasta berbasis Islam, (4) SD Swasta berbasis Katholik, (5)

SD swasta berbasis kebangsaan, (6) SDIT yang menerapkan *full day school*, dan (7) MI di bawah Kementerian Agama.

Hasil yang diperoleh dari investigasi awal ini, antara lain: model penerapan pendidikan karakter di SD dilaksanakan melalui empat model yakni: (1) pembelajaran di kelas, (2) pembiasaan rutin dalam kegiatan sekolah, (3) keteladanan dalam menjaga ketertiban, dan (4) kegiatan keagamaan.

Evaluasi praktek penilaian karakter di SD dilaksanakan dengan dua model penilaian yakni: (1) model evaluasi perilaku, dan (2) model monitoring akhlak yang dilakukan oleh guru PAI. Kedua model penilaian mengandung kelemahan sebagai berikut: penilaiannya bersifat mekanik, penilaiannya tidak menyeluruh, hanya aspek perilaku saja. Hasil penilaian belum menyentuh keterampilan yang dibutuhkan untuk memperkuat kepribadian siswa, dan siswa tidak dilibatkan untuk menilai sendiri penalaran dan perasaan moral yang dimilikinya.

Hasil analisis kebutuhan menunjukkan adanya kebutuhan yang dirasakan guru tentang pentingnya buku panduan penilaian keterampilan intrapribadi dan antarpribadi dalam pendidikan karakter untuk siswa SD. Panduan ini dapat dijadikan sebagai pedoman baku bagi para guru untuk menilai keterampilan intrapribadi dan antarpribadi siswa sebagai keterampilan *softskills* yang perlu dikuasai oleh siswa sebagai bekal untuk memperoleh kesuksesan hidup.

Hasil Pembahasan

Desain kisi-kisi instrumen model PKIA-SD

Untuk mendesain kisi-kisi instrumen model PKIA-SD diawali dengan perumusan definisi konseptual keterampilan intrapribadi dan antarpribadi. Keterampilan intrapribadi didefinisikan sebagai kemampuan individu dalam mengenali, menghargai, mengelola, mengendalikan perasaan diri, dan merasakan realita spiritualitas diri untuk memahami keberadaan diri dan berani bertanggung jawab terhadap kehidupan pribadi, sehingga terbentuk kepribadian berakhlak mulia. Definisi ini dirumuskan sebagai hasil

synthesis lima konsep teori keterampilan intrapribadi dari pendapat Gardner (1993, p.24-25), Lazear (2000, p.24), Bar-On (1997b, p.120), Lwin. (2005, p.233), dan Amin (1974, p.61).

Keterampilan antarpribadi didefinisikan sebagai kemampuan untuk menghargai pandangan orang lain, menyadari tanggung jawab sosial, bekerja sama, tenggang rasa, dan mampu berkomunikasi dengan orang lain. Definisi ini sebagai hasil synthesis lima konsep teori keterampilan antarpribadi yang berasal dari pendapat Gardner (1993, p.240), Bar-On (1997b, p.161), Lazear (2000, p.23), Mant (1997, p.161), dan Cunningham & Corderio (2003, p.205).

Berdasarkan definisi konseptual tersebut dapat dirumuskan indikator-indikator dari keterampilan intrapribadi dan keterampilan antarpribadi. Keterampilan intrapribadi mempunyai lima indikator, yaitu: (1) mengenali perasaan diri sendiri, (2) menghargai perasaan diri sendiri, (3) mengelola perasaan diri sendiri, (4) mengendalikan perasaan diri sendiri, dan (5) merasakan realita spiritualitas diri sendiri. Sedangkan keterampilan antarpribadi mempunyai lima indikator, yaitu: (1) menghargai pandangan orang lain, (2) menyadari tanggung jawab sosial, (3) bekerja sama dengan orang lain, (4) tenggang rasa terhadap orang lain, dan (5) berkomunikasi secara efektif dengan orang lain.

Berdasarkan indikator tersebut disusunlah kisi-kisi instrumen model PKIA-SD yang dibagi dalam tiga ranah penilaian keterampilan intrapribadi dan antarpribadi dalam pendidikan karakter, yaitu kognitif, afektif, dan psikomotor.

Desain model PKIA-SD

Model PKIA-SD yang dikembangkan dalam penelitian ini berupa seperangkat pengukuran nontes yang terdiri atas 3 jenis model penilaian untuk menilai keterampilan intrapribadi dan antarpribadi. Adapun ketiga jenis penilaian dalam model PKIA-SD sebagai berikut.

Model penilaian keterampilan intrapribadi dan antarpribadi pada aspek kognitif disingkat IN-TRAKOG dan ANTARKOG

Model penilaian ini digunakan untuk mengetahui tahapan penalaran moral siswa tentang keterampilan intrapribadi dan antarpribadi. Model penilaian INTRAKOG dan ANTARKOG disusun dalam bentuk *projective test*. Pernyataan dalam *projective test* dibuat dalam bentuk *dilema moral* artinya kasus tersebut merupakan kasus tentang moral yang dilematis. Penilaian dan pemberian skornya dilakukan dengan cara melihat alasan yang dikemukakan oleh siswa. Kemudian jawaban tersebut dinilai berdasarkan klasifikasi tahapan moral sesuai pendapat Kohlberg seperti ditunjukkan pada Tabel 1.

Tabel 1. Tingkatan Moral Kognitif

Tingkatan Moral	Skor	Tahapan Moral
Pra-konvensional	1	Egosentrik
	2	Individualis
Konvensional	3	Persetujuan mayoritas
	4	Pemeliharaan aturan sosial
Pasca-konvensional	5	Komitmen kontrak sosial
	6	Etis universal

Model penilaian keterampilan intrapribadi dan antarpribadi pada aspek afektif disingkat IN-TRAFEK dan ANTARFEK

Model penilaian ini digunakan untuk mengetahui perasaan diri dan rasa spiritualitas diri. Model penilaian ini dibuat dalam bentuk *grading skills* yakni berupa sejumlah tingkatan keterampilan yang disusun secara tidak berurutan, dan siswa diminta untuk memilih norma, perasaan, dan perilaku yang paling sesuai dengan kebiasaan yang dilakukan siswa.

Model penilaian keterampilan intrapribadi dan antarpribadi pada aspek psikomotor disingkat INTRATOR dan ANTARTOR

Model penilaian ini digunakan untuk mengetahui performan perilaku siswa sehari-hari. Model penilaian ini dibuat dalam bentuk lembar observasi dengan skala peng-

amatan yang bergerak dari belum muncul sampai sudah muncul secara konsisten. Penilaian ini dilakukan oleh guru kelas untuk mengetahui tingkat intensitas perilaku dan sikap siswa sehari-hari.

Validasi Model PKIA-SD

Validasi isi model PKIA-SD diberikan oleh tiga unsur yakni pakar, praktisi, dan responden. Dari hasil validasi pakar yang meliputi berbagai bidang: psikometri, karakter, psikologi, bahasa, dan metodologi diperoleh saran penyempurnaan draf model PKIA-SD. Kelima pakar juga melakukan penilaian secara kuantitatif terhadap draf model PKIA-SD yang meliputi lima aspek penilaian yakni: (1) petunjuk penilaian, (2) cakupan penilaian, (3) penggunaan bahasa, (4) tata tulis penilaian, dan (5) penilaian secara menyeluruh. Hasil penilaian pakar dianalisis dengan formula Aiken's V dan diperoleh nilai koefisien kesepakatan rater untuk validitas isi sebesar 0,88. Nilai koefisien ini termasuk tinggi bagi item tersebut. Hal ini berarti item tersebut memiliki validitas isi yang baik dan mendukung isi tes secara keseluruhan.

Berdasarkan masukan dari praktisi, diperoleh saran berupa pengurangan jumlah item model penilaian INTRAKOG dan ANTARKOG dari 30 butir soal menjadi 10 butir soal, dan skala penilaiannya dari 1-4 menjadi 1-6. Alasan pengurangan butir yaitu: (1) untuk mempermudah pembuatan kasus moral dilematis secara tepat, (2) dapat mengukur kemampuan penalaran moral siswa SD pada setiap indikator keterampilan intrapribadi dan antarpribadi, (3) memudahkan dalam melakukan penskoran setiap item, (4) tidak terjadi duplikasi item yang dipergunakan untuk mengukur tahapan penalaran moral anak SD, dan (5) kata-kata yang digunakan lebih operasional untuk mengukur penalaran moral siswa SD, dan (6) memudahkan dalam analisis.

Dari uji keterbacaan model PKIA-SD oleh siswa diperoleh saran untuk perbaikan penggunaan istilah-istilah dan kalimat yang sulit dimengerti oleh siswa. Hal itu perlu disederhanakan agar sesuai dengan tingkat penalaran siswa SD.

Uji Coba Model PKIA-SD Tahap I

Uji coba model PKIA-SD dilakukan dalam dua tahap. Uji coba tahap pertama dilakukan pada sampel kecil, yaitu 2 SD yang berjumlah 54 siswa. Penentuan sekolah didasarkan perbedaan budaya pendidikan karakter yang diterapkan pada sekolah tersebut.

Hasil uji coba tahap pertama menunjukkan bahwa keterampilan intrapribadi dan antarpribadi aspek psikomotor siswa lebih baik dibandingkan keterampilan intrapribadi dan antarpribadi aspek kognitif dan afektif. Hal ini membuktikan bahwa pendidikan karakter tidak harus dilakukan secara linier yakni berurutan dari kognitif, afektif, kemudian psikomotor. Oleh karena itu, metode yang tepat digunakan untuk mendidik karakter adalah dengan pendekatan konstruktivistik. Artinya siswa dapat secara bersamaan dilatih untuk berkarakter positif, bahkan dapat dimulai dari metode pembiasaan kemudian baru dilanjutkan dengan metode penalaran terhadap perilaku.

Hasil uji coba ini juga menunjukkan bahwa siswa SD memiliki keterampilan intrapribadi yang rendah pada kemampuan mengelola emosi, mengenali emosi, dan menghargai emosi diri sendiri, sedangkan kemampuan yang paling tinggi adalah kemampuan merasakan rasa spiritualitas diri. Keterampilan antarpribadi siswa SD yang paling rendah adalah kemampuan menghargai pikiran dan perasaan orang lain. Kemampuan yang paling tinggi adalah kemampuan tanggung jawab sosial dan kemampuan untuk bersikap tenggang rasa kepada orang lain.

Hasil analisis validitas konstruk dengan *Exploratory Factor Analysis* (EFA), dan hasil analisis reliabilitas menggunakan formula *Cronbach's Alpha* terhadap model PKIA-SD disajikan pada Tabel 2.

Tabel 2. menunjukkan bahwa keenam instrumen model PKIA-SD termasuk kategori memuaskan, dan sebagian besar faktor yang terbentuk adalah *unidimensi*. Pada model PKIA-SD aspek intrator terdapat dua faktor yang memiliki nilai *eigenvalue* yang lebih besar dari satu. Namun, faktor yang

dipilih hanyalah satu, yakni faktor yang memiliki kedekatan makna dengan variabel yang diukur, dan memiliki nilai varian yang lebih besar dari kedua faktor tersebut.

Tabel 2. Hasil EFA dan Uji Reliabilitas

Model PKIA	KMO MSA	Nilai <i>eigenvalue</i>	Reliabilitas
Intrakog	0,869	1 faktor	0,823
Antarkog	0,872	1 faktor	0,818
Intrafek	0,938	1 faktor	0,779
Antarfek	0,937	1 faktor	0,776
Intrator	0,924	2 faktor	0,775
Antartor	0,902	1 faktor	0,774

Kemampuan siswa yang sudah baik dalam aspek keterampilan intrapribadi, yaitu kemampuan menghargai dan mengendalikan emosi diri, sedangkan kemampuan siswa yang masih perlu pembinaan adalah kemampuan mengelola dan mengenali emosi diri. Dalam aspek keterampilan antarpribadi, kemampuan siswa yang sudah baik yaitu kemampuan dalam berkomunikasi dengan orang lain, kemampuan bekerja sama, dan kemampuan berempati pada orang lain, sedangkan kemampuan siswa yang masih perlu pembinaan adalah kemampuan dalam bertanggung jawab dan kemampuan menghargai orang lain.

Uji Coba Model PKIA-SD Tahap II

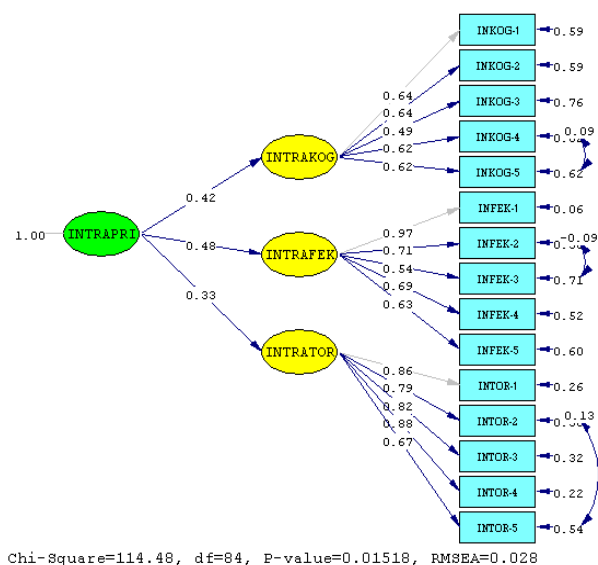
Uji coba tahap kedua dilakukan pada sampel diperluas, berjumlah 464 siswa berasal dari 15 SD. Penentuan sekolah didasarkan perbedaan budaya sekolah yang beragam dalam pelaksanaan pendidikan karakter.

Uji coba tahap kedua bertujuan untuk mengujicobakan model penilaian yang telah direvisi, baik bahasa, konten, maupun prosedur pelaksanaan. Hal itu sesuai saran yang diperoleh dari hasil uji coba tahap pertama.

Pengujian kecocokan model PKIA-SD dilakukan dengan menggunakan model pengukuran analisis faktor konfirmatori orde dua (*second order confirmatory factor analysis*). Ada dua model pengukuran, yaitu: model

keterampilan intrapribadi (*intrapersonal skill*) dan model keterampilan antarpribadi (*interpersonal skill*) yang masing masing keterampilan tersebut memiliki tiga aspek yaitu aspek kognitif, afektif, dan psikomotor.

Berdasarkan hasil *second order* CFA untuk konstruk keterampilan intrapribadi diperoleh *path diagram* yang ditampilkan pada Gambar 1. Hasil analisis kecocokan model (*goodness of fit statistics* (GOF)) untuk konstruk keterampilan intrapribadi (Wijanto, 2008, p.51-58) ditampilkan pada Tabel 3.



Gambar 1. Hasil Pengujian konstruk Keterampilan Intrapribadi

Tabel 3. Hasil Uji Kecocokan Model Keterampilan Intrapribadi

Kriteria <i>Goodness of Fit</i>	Tingkat Kecocokan	Hasil Estimasi	Hasil
P-value	> 0,05	0,01518	Kurang
RMSEA	≤ 0,08	0,028	Baik
RMR	< 0,05	0,041	Baik
NFI	≥ 0,90	0,97	Baik
NNFI	≥ 0,90	0,99	Baik
CFI	≥ 0,90	0,99	Baik
IFI	≥ 0,90	0,99	Baik
RFI	≥ 0,90	0,97	Baik
GFI	≥ 0,90	0,97	Baik
AGFI	≥ 0,90	0,95	Baik

Sumber: Hasil Analisis LISREL

Dari hasil pengujian kecocokan model tersebut diperoleh 9 ukuran GOF yang menunjukkan kecocokan model yang baik dan 1 ukuran GOF yang menunjukkan kecocokan model yang baik kecuali untuk *p-value*-nya. Dengan demikian, dapat disimpulkan bahwa model pengukuran konstruk keterampilan intrapribadi telah memenuhi ketentuan *goodness of fit statistic*.

Dari hasil pengujian *second order* CFA konstruk keterampilan intrapribadi diperoleh 14 nilai variabel teramati (indikator) yang memiliki nilai t_{hitung} lebih besar dari 1,96 dan nilai koefisien muatan faktor yang distandarkan lebih besar dari 0,50. Dengan demikian dapat disimpulkan bahwa variabel teramati (indikator) tersebut mempunyai validitas yang baik terhadap variabel latennya. Sedangkan 1 nilai variabel teramati (INKOG-3) memiliki nilai t_{hitung} lebih besar dari 1,96 tetapi nilai koefisien muatan faktor yang distandarkan (λ) lebih kecil dari 0,50, sehingga dapat dinyatakan bahwa variabel teramati tersebut (INKOG-3) mempunyai validitas yang cukup terhadap variabel laten kognitif (INTRAKOG). Hal ini dapat disimpulkan bahwa keempatbelas indikator yang digunakan untuk mengukur aspek kognitif, afektif, dan psikomotor pada keterampilan intrapribadi mempunyai validitas yang baik, sedangkan satu indikator mempunyai validitas yang cukup. Hasil Pengujian *Second Order CFA* konstruk keterampilan intrapribadi ditampilkan pada Tabel 4.

Dari hasil pengujian itu juga diperoleh nilai koefisien reliabilitas komposit (*construct reliability*, CR) lebih besar dari 0,7 dan nilai koefisien ekstrak varian (*variance extracted*: VE) lebih besar dari 0,5 untuk ketiga variabel laten dari konstruk keterampilan intrapribadi. Hal ini berarti reliabilitas semua variabel laten adalah baik. Dengan demikian, dapat disimpulkan bahwa reliabilitas model pengukuran semua variabel laten dari konstruk keterampilan intrapribadi adalah baik.

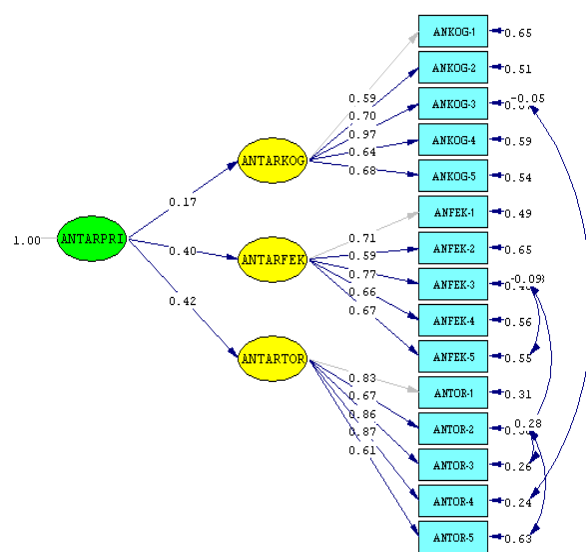
Berdasarkan pengujian *second order* CFA untuk konstruk keterampilan antarpribadi diperoleh *path diagram* yang ditampilkan pada Gambar 2. Hasil pengujian ke-

cocokan model (*goodness of fit statistics*) untuk konstruk keterampilan antarpribadi ditampilkan pada Tabel 5.

Tabel 4. Hasil Pengujian *Second Order CFA* Konstruk Keterampilan Intrapribadi

Variabel	Stand. Loading Factors	Nilai t	Reliabilitas	
			CR	VE
INTRAKOG				
INKOG-1	0,64	---	0,74	0,52
INKOG-2	0,64	9,74		
INKOG-3	0,49	8,15		
INKOG-4	0,62	9,06		
INKOG-5	0,62	9,04		
INTRAFEK				
INFEK-1	0,97	---	0,84	0,59
INFEK-2	0,71	17,96		
INFEK-3	0,54	12,48		
INFEK-4	0,69	17,61		
INFEK-5	0,63	15,46		
INTRATOR				
INTOR-1	0,86	---	0,90	0,65
INTOR-2	0,79	20,29		
INTOR-3	0,82	21,67		
INTOR-4	0,88	24,11		
INTOR-5	0,67	16,10		

Sumber: Hasil Analisis LISREL



Gambar 2. Hasil Pengujian konstruk Keterampilan Antarpribadi

Dari hasil pengujian kecocokan model tersebut diperoleh 9 ukuran GOF yang menunjukkan kecocokan model yang baik dan 1 ukuran GOF yang menunjukkan kecocokan model yang cukup baik kecuali pada *p-value*-nya. Dengan demikian, dapat disimpulkan bahwa model pengukuran konstruk keterampilan antarpribadi telah memenuhi ketentuan *goodness of fit statistic*.

Tabel 5. Hasil Uji Kecocokan Model Keterampilan Antarpribadi

Kriteria <i>Goodness of Fit</i>	Tingkat Kecocokan	Hasil Estimasi	Hasil
P-value	> 0,05	0,00000	Kurang
RMSEA	≤ 0,08	0,048	Baik
RMR	< 0,05	0,040	Baik
NFI	≥ 0,90	0,96	Baik
NNFI	≥ 0,90	0,97	Baik
CFI	≥ 0,90	0,98	Baik
IFI	≥ 0,90	0,98	Baik
RFI	≥ 0,90	0,95	Baik
GFI	≥ 0,90	0,95	Baik
AGFI	≥ 0,90	0,93	Baik

Sumber: Hasil Analisis LISREL

Dari hasil pengujian *second order CFA* konstruk keterampilan antarpribadi diperoleh 15 nilai variabel teramati (indikator) yang memiliki nilai t_{hitung} lebih besar dari 1,96 dan nilai koefisien muatan faktor yang distandarkan lebih besar dari 0,50. Dengan demikian dapat disimpulkan bahwa kelima belas variabel teramati (indikator) tersebut mempunyai validitas yang baik terhadap variabel latennya. Hal ini dapat disimpulkan bahwa semua indikator yang digunakan untuk mengukur aspek kognitif, afektif, dan psikomotor pada keterampilan intrapribadi mempunyai validitas yang baik. Hasil Pengujian *Second Order CFA* konstruk keterampilan antarpribadi ditampilkan pada Tabel 6.

Dari hasil pengujian itu juga diperoleh nilai koefisien reliabilitas komposit (*construct reliability*: CR) lebih besar dari 0,7 dan nilai koefisien ekstrak varians (*variance extracted*: VE) lebih besar dari 0,5 untuk ketiga variabel laten dari konstruk keterampilan antar-

pribadi. Hal ini berarti reliabilitas semua variabel laten adalah baik. Dengan demikian, dapat disimpulkan bahwa reliabilitas model pengukuran semua variabel laten dari konstruk keterampilan antarpribadi adalah baik.

Tabel 6. Hasil Pengujian *Second Order CFA* Konstruk Keterampilan Antarpribadi

Variabel	Stand. Loading Factors	Nilai t	Reliabilitas	
			CR	VE
ANTARKOG				
ANKOG-1	0,59	---	0,84	0,53
ANKOG-2	0,70	12,30		
ANKOG-3	0,97	14,11		
ANKOG-4	0,64	11,48		
ANKOG-5	0,68	12,01		
ANTARFEK				
ANFEK-1	0,71	---	0,81	0,50
ANFEK-2	0,59	11,49		
ANFEK-3	0,77	13,35		
ANFEK-4	0,66	12,86		
ANFEK-5	0,67	11,57		
ANTARTOR				
ANTOR-1	0,83	---	0,88	0,60
ANTOR-2	0,67	15,14		
ANTOR-3	0,86	21,36		
ANTOR-4	0,87	21,69		
ANTARKOG	0,61	13,66		

Sumber: Hasil Analisis LISREL

Berdasarkan hasil analisis *second order CFA* tersebut, maka dapat diketahui bahwa dua model yaitu: (1) model instrumen penilaian keterampilan intrapribadi (INTRAPRI) yang terdiri dari aspek keterampilan intrapribadi kognitif (INTRAKOG), aspek keterampilan intrapribadi afektif (INTRAFEK), aspek keterampilan intrapribadi psikomotor (INTRATOR) dan model (2) instrumen penilaian keterampilan antarpribadi (ANTARPRI) yang terdiri dari aspek keterampilan antarpribadi kognitif (ANTARKOG), aspek keterampilan antarpribadi afektif (ANTARFEK), dan aspek keterampilan

pribadi antarpribadi psikomotor (ANTARTOR) telah memenuhi persyaratan statistik dan dikategorikan sebagai model pengukuran yang cocok (*fit*). Hal ini berarti model PKIA-SD telah memenuhi *goodness of fit statistic*, sehingga PKIA-SD dapat dinyatakan sebagai model penilaian yang dapat untuk mendeteksi capaian keterampilan intrapribadi dan antarpribadi dalam pendidikan karakter di Sekolah Dasar.

Dengan demikian, dapat disimpulkan bahwa hasil analisis uji coba tahap II pada keterampilan intrapribadi, kemampuan siswa yang masih rendah, yaitu bersikap inovatif, sehingga diperlukan pembinaan lebih lanjut dari guru kelas melalui pembiasaan dan pembelajaran di sekolah. Pada keterampilan antarpribadi, semua kemampuan siswa sudah baik sehingga guru kelas perlu menjaga pembinaan tersebut melalui pembiasaan dan pembelajaran di sekolah.

Simpulan dan Saran

Simpulan

Simpulan penelitian ini dapat dirumuskan sebagai berikut: (1) Model penilaian untuk mendeteksi capaian pendidikan karakter di SD selama ini adalah model evaluasi perilaku dan model monitoring akhlak siswa berbasis pelajaran Agama; (2) Kelemahan kedua model penilaian tersebut adalah penilaiannya bersifat mekanik, tidak komprehensif, belum menilai aspek kepribadian, dan tidak melibatkan siswa dalam penilaian; (3) Konstruk model PKIA-SD terdiri dari aspek kognitif, afektif, dan psikomotorik. Penilaian dilakukan komprehensif dengan instrumen berbentuk *projective test*, *rating scale*, dan *cek list*, skala penilaiannya diukur melalui kasus *dilema moral*, *grading skills*, dan *performance assessment*; (4) Instrumen model PKIA-SD telah memenuhi validitas isi dan reliabilitas; (5) Model PKIA-SD dinyatakan valid dan reliabel, dan memenuhi kriteria dalam pengujian kecocokan model (*goodness of fit statistics*); (6) Prosedur model PKIA-SD diawali dengan menjumlahkan semua hasil penilaian *self report* siswa pada aspek penalaran dan aspek

perasaan moral dengan hasil penilaian guru melalui pengamatan aspek perilaku moral selama siswa di sekolah. Hasil penjumlahan ketiga aspek tersebut disusun dalam kategorisasi, sehingga dapat diketahui siswa yang perlu pembinaan dan siswa yang perlu pengayaan karakter; (7) Panduan model PKIA-SD ini dinilai komprehensif, mudah, praktis, dan lengkap oleh para guru selaku pengguna, karena dilengkapi dengan kriteria dan rubrik yang jelas, serta diberikan contoh teknik analisis dan teknik pelaporannya sesuai dengan aturan penilaian dalam Kurikulum 2013.

Saran

Beberapa saran pemanfaatan model PKIA-SD sebagai berikut: (1) Model PKIA-SD ini dapat dijadikan sebagai pegangan bagi guru SD untuk mendeteksi capaian keterampilan intrapribadi dan antarpribadi siswa pada aspek kognitif, afektif, dan psikomotor; (2) Penggunaan model PKIA-SD ini dapat mendorong guru lebih objektif dalam melaporkan tingkat capaian keterampilan intrapribadi dan antarpribadi siswa SD kepada pihak-pihak terkait, terutama kepada orang tua yang membutuhkan laporan perkembangan capaian keterampilan anaknya di SD; (3) Model PKIA-SD ini dapat mendorong para guru SD untuk lebih kreatif dan inovatif dalam melakukan kegiatan pembelajaran di kelas, sehingga mampu mendukung peningkatan keterampilan intrapribadi dan antarpribadi siswanya. Penggunaan panduan model PKIA-SD ini telah dilengkapi dengan kriteria dan rubrik yang jelas sehingga dapat membantu guru dalam mengembangkan karakter positif siswa-siswanya.

Sebagai tindak lanjut dari penelitian ini, diharapkan para peneliti atau pihak-pihak yang memiliki minat terhadap pengembangan model penilaian keterampilan intrapribadi dan antarpribadi siswa SD, hendaknya penelitian ini dapat dijadikan acuan untuk mengembangkan model penilaian lebih luas lagi, baik dari subjek maupun objeknya, artinya tidak terbatas pada SD kelas tinggi saja, namun dapat diguna-

kan untuk acuan pengembangan model penilaian keterampilan intrapribadi dan antarpribadi pada siswa-siswa SD kelas rendah, atau jenjang sekolah di atasnya.

Daftar Pustaka

- Amin, A. (1974). *Kitab al-akhlak*. Beirut-Lubnan: Dar al-Kitab al-'Araby.
- Bar-On, R. (1988). *The development of a concept of psychological well-being*. Doctoral dissertation: Rhodes University, South Africa.
- Bar-On, R.. (1997a). *The emotional quotient inventory (EQ-i): a test of emotional intelligence*. Toronto, Canada: Multi-Helth Systems, Inc.
- Bar-On, R.. (1997b). *The emotional quotient inventory (EQ-i): technical manual*. Toronto. Canada: Multi-Health Systems, Inc.
- Bar-On, R.. (2000). *Emotional dan social intelligence: Insights from the emotional quotient inventory (EQ-i)*. In R.Bar-On and J.D.A.Parker (Eds), *Handbook of emotional intelligence*. San Francisco: Jossey-Bass.
- Bar-On, R.. (2005). *Emotional intelligence and subjective wellbeing*. Manuscript submitted for publication.
- Bar-On, R., & Fund, S. (2004). *The impact of emotional and social intelligence on self-perceived physical helth*. Manuscript submitted for publication.
- Berkowitz, M.W., Battistich, V.C., & Bier, M.C. (2002). *What works in character education: What is known and what needs to be known*. In Nucci, Larry P.& Narvaez, Darcia. (Eds.). *Handbook of moral and character education* pp.414-431. New York and London: Routledge Taylon & Francis Group.
- Cunningham, W.G., & Paula, A.C. (2003). *Educational leadership: a problem based approach*. Boston: Pearson Education.

- Gardner, H. (1993). *Frames of mind: the theory of multiple intelligence*. New York: Basic Books.
- Gardner, H. (1996). *Intelligence: multiple perspectives*. Harcourt Brace College. Fort Worth. Tex.
- Goleman, D. (1995). *Emotional intelligence, why it can matter more than IQ*. New York: Bantam Books.
- Joreskog, K.G., & Sorbom, D. (1999). *Lisrel 8: User's reference guide*. Chicago: Scientific Software International.
- Kemendiknas.(2011). *Panduan pelaksanaan pendidikan karakter*. Jakarta: Balitbang Diknas.
- Khilmiyah, A. (2011). Evaluasi implementasi pendidikan karakter pada Sekolah Dasar percontohan pendidikan karakter di kabupaten Bantul. *Laporan Penelitian*. Yogyakarta: LP3M UMY.
- Khilmiyah, A. (2012). Perbandingan keterampilan intrapersonal dan interpersonal siswa Sekolah Dasar Negeri Kasihan Bantul. *Jurnal Afkaruna*, Vol. 8 No.2 Juli-Desember 2012. ISSN 1693-4040. Yogyakarta: FAI UMY.
- Lazear, D. (2000). *Pathways of learning: teaching students and parents about multiple intelligences*. Arizona: Zephyr.
- Lwin, M. (2005). *How to multiply your child's intelligence*. Terjemahan Christine S. Jakarta: Indeks.
- Mant, A. (1997). *Intelligent leadership*. New South Wales: Allen & Unwin.
- Plomp, T.J. (1997). *Educational design: introduction. from Plomp (eds.). Educational & training system design: introduction. Design of education and training (in Dutch)*. Utrecht (the Netherlands): Lemma. Netherland. Faculty of Educational Science and Technology, University of Twente.
- Wijanto, S. H. (2008). *Structural equation modeling dengan Lisrel 8.8 konsep dan tutorial*. Jogjakarta: Graha Ilmu.
- Zamroni. (2005). Pengembangan sistem penilaian pendidikan yang menerapkan KBK dalam kerangka otonomi daerah. *Jurnal Penelitian dan Evaluasi Pendidikan*. Vol 2. No.12, Mei 2005, Tahun 11, ISSN 1410-4725. Yogyakarta: HEPI. PPs UNY.
- Zuchdi, D. (2010). *Humanisasi pendidikan, menemukan kembali pendidikan yang manusiawi*. Jakarta: Bumi Aksara.
- Zuchdi, D. et.al. (2010). Pengembangan model pendidikan karakter dengan pendekatan komprehensif di Sekolah Dasar terpadu dengan pembelajaran Bahasa Indonesia, IPA dan IPS. *Laporan Penelitian Hibah Pascasarjana*. Yogyakarta: UNY.

PENYETARAAN VERTIKAL DENGAN PENDEKATAN KLASIK DAN ITEM RESPONSE THEORY PADA SISWA SEKOLAH DASAR

¹⁾Anak Agung Purwa Antara, ²⁾Bastari

¹⁾IKIP Saraswati Tabanan, ²⁾Universitas Indonesia

¹⁾purwa.antara@gmail.com, ²⁾bastari@kemdikbud.go.id

Abstrak

Penelitian ini bertujuan untuk mengetahui (1) akurasi metode linear dan ekipersentil pada pendekatan Klasik dan *mean & mean* dan *mean & sigma* pada pendekatan teori respons butir; (2) karakteristik tes prestasi belajar matematika model campuran kelas III, IV, V, dan VI Sekolah Dasar yang digunakan dalam penelitian ini; dan (3) perkembangan kemampuan siswa (θ) kelas III, IV, V, dan VI pada pelajaran matematika berdasarkan tes yang digunakan. Data penelitian adalah skor siswa pada ujian akhir semester tahun 2013 yang dikumpulkan dengan desain tes dengan butir bersama. Hasil penelitian menunjukkan bahwa, (1) *Standart Error of Equating (SEE)* metode ekipersentil 0,505116 dan metode linear 0,608, sedangkan *Root Mean Square Different (RMSD)* metode *mean & sigma* 0,344 dan metode *mean & mean* 0,394, perbedaan nilai *SEE* pada pendekatan Klasik dan *RMSD* pada pendekatan *IRT* sangat kecil sehingga kebermaknaannya masih perlu diteliti; (2) tes prestasi belajar matematika kelas III, IV, V, dan VI berturut-turut memiliki rerata *location* katagori sedang -0,070, -0,228, -0,324, -0,430, *slope* yang baik 0,537, 0,558, 0,639, 0,644, dan *items fit statistics* minimum 0,077, 0,052, 0,055, dan 0,198; (3) keempat metode penyetaraan memberikan informasi bahwa nilai θ meningkat seiring dengan meningkatnya peringkat kelas, dengan rerata θ kelas III, IV, V, dan VI berturut-turut sebesar -0,104, -0,076, 0,004, dan 0,173. Namun, kenaikan nilai tersebut sangat kecil sehingga kebermaknaannya masih perlu diteliti.

Kata kunci: *Metode penyetaraan, Tes model campuran*

VERTICAL EQUATING USING CLASICAL METHOD AND ITEM RESPONSE THEORY IN ELEMENTARY SCHOOL STUDENTS

¹⁾Anak Agung Purwa Antara, ²⁾Bastari

¹⁾IKIP Saraswati Tabanan, ²⁾Universitas Indonesia

¹⁾purwa.antara@gmail.com, ²⁾bastari@kemdikbud.go.id

Abstract

The study was aimed at finding: (1) the accurate linear method and equipercntile method in Classic approach and *mean&mean* and *mean & sigma* method in *IRT* approach, (2) the characteristics of the mathematic mixed-model achievement test for elementary students grade III, IV, V, and VI used in this study, (3) the progress of elementary students gade III, IV, V, and VI ability (θ) in Mathematic based on the developed test. The data of this study were students' scorer on final exams Elementary school 2013 collected with anchor test design. The result showed that: (1) Standard Error of Equating (SEE) within the linear method was 0.607 and ekipercntile method was 0.505116, while Root Mean Square Different (RMSD) within the *mean&mean* was 0.394 and *mean&sigma* method was 0.344. The differences between the value of SEE and RMSD was very little, therefore the meaningfulness needed further study; (2) the Mathematic achievement test for Grade III, IV, V, and VI had location in the middle category within the values respectively -0.070, -0.228, -0.324, -0.430, the good mean of slope 0.537, 0.558, 0.639, 0.644, and items fit statistics minimum 0.077, 0.052, 0.055, 0.198; (3) the four equation method provided information that θ increased in accordance with the increased of the class rank, with mean of the θ students Grade III, IV, V, and VI as 0.104, -0.076, 0.004, dan 0.173. But the increased of the θ value was relatively small, in that its meaningfulness also required further study.

Keywords: *equating method, mixed-model test*

Pendahuluan

Isu penting yang selalu hangat dibicarakan dalam dunia pendidikan adalah peningkatan kualitas hasil belajar atau kompetensi lulusan dari suatu lembaga pendidikan agar lulusan yang dihasilkan menjadi sumber daya yang bermanfaat, mampu bertahan dan bersaing dalam kehidupan sekarang ataupun dalam kehidupan yang akan datang. Usaha-usaha pencapaian kualitas hasil belajar tersebut telah dilakukan pemerintah misalnya melalui penyempurnaan kurikulum yang mencakup empat hal yaitu Standar Kompetensi Lulusan (SKL), materi yang harus dipelajari atau diajarkan, pelaksanaan atau proses pembelajaran, dan penilaian ketercapaian tujuan pembelajaran atau hasil belajar (Kemendikbud, 2013).

Berkaitan dengan hal tersebut, Mardapi (2008, p.5) menyatakan bahwa, peningkatan kualitas hasil belajar dari suatu lembaga pendidikan dapat dimulai dari peningkatan kualitas program pembelajaran dan peningkatan kualitas penilaian yang dilakukan di dalam kelas. Hal tersebut dapat dimaknai bahwa guru sebagai komponen yang bersentuhan langsung dengan siswa hendaknya mampu mengembangkan potensi-potensi yang dimiliki siswa, baik melalui pelaksanaan program pembelajaran, maupun melalui model penilaian yang digunakan. Model penilaian yang digunakan guru, hendaknya mampu memantau atau memberikan informasi tentang proses pembelajaran yang telah dilakukan, perkembangan kemampuan siswa setelah mengikuti proses pembelajaran, dan hasil belajar yang telah dicapai siswa, sehingga dapat dilakukan evaluasi untuk perbaikan menuju pencapaian tujuan pendidikan yang diharapkan.

Penilaian yang baik memerlukan data yang benar sehingga perlu didukung oleh proses pengukuran yang baik (Mardapi, 2008, p.5). Penerapan program pengukuran yang bermutu akan mendorong semua komponen yang terlibat dalam proses pembelajaran, terutama guru dan siswa untuk melakukan upaya-upaya dalam pencapaian kualitas hasil belajar sesuai dengan tujuan yang diharapkan. Guru dituntut meningkat-

kan pengetahuan dan memperluas penguasaan materi pelajaran sehingga mampu menyusun tes sebagai alat ukur yang berkualitas untuk mengukur ketecapaian program pembelajaran yang telah dilakukan.

Alat ukur pada pelajaran matematika memiliki struktur tersendiri karena materi matematika bersifat hierarkis. Hal ini berarti penguasaan siswa terhadap materi pada kelas tertentu, berkaitan dengan penguasaan materi pada kelas sebelumnya. Dengan demikian, untuk mengungkap perkembangan kemampuan siswa pada pelajaran matematika diperlukan kemampuan menyusun tes yang dapat menghubungkan kemampuan siswa pada seluruh peringkat kelas. Guru hendaknya mampu memilih model tes yang sesuai, mampu menyusun tes yang baik, dan menggunakan tes tersebut secara tepat untuk memperoleh informasi perkembangan kemampuan siswa.

Rasiman (2008, p.258), menyatakan bahwa informasi perkembangan kemampuan siswa dapat diperoleh secara optimal apabila guru memperhatikan perkembangan kognitif siswa yang diukur. Penggunaan tes yang berada di luar kecakapan siswa akan menyebabkan siswa tersebut tidak mampu menjawab butir soal yang diberikan sehingga guru tidak memperoleh informasi yang cukup tentang perkembangan kemampuan siswanya. Perbedaan kemampuan siswa yang ekstrim dalam satu kelas akan menyebabkan guru menemui masalah dalam melakukan pengukuran. Siswa yang pintar kemungkinan akan mampu menjawab secara tuntas semua butir soal yang berada pada tingkatnya, dan ada kemungkinan mampu menjawab soal dengan materi yang berada di atas tingkatnya, demikian sebaliknya.

Menurut Naga (1992, p.347), cara untuk mengatasi masalah tersebut adalah dengan memberikan siswa yang lambat tes yang berada pada tingkat di bawahnya, sedangkan siswa yang cepat diberikan tes pada tingkat di atasnya, kemudian melalui penyamaan skala, skor yang mereka peroleh dikonversi kembali pada kedudukan semula.

Menurut Kolen & Brennan (2004, p. 5), skor dua tes berbeda dari dua kelompok

atau lebih dapat diperbandingkan jika butir-butir soalnya setara dan memiliki skala yang sama. Kesetaraan antarskor dapat dilakukan secara statistik. Proses statistik yang digunakan untuk menghasilkan skala tunggal dari skor dua tes yang berbeda tersebut dinamakan penyetaraan atau *equating* (Kolen & Brennan, 1995, p.2).

Penyetaraan dapat dilakukan secara horizontal ataupun secara vertikal (Crocker & Algina, 1986, p.456; Hambleton & Swaminathan, 1985, p.197). Penyetaraan yang dilakukan terhadap tes yang memiliki tingkat kesulitan berbeda pada tingkat kelas yang berbeda tapi mengukur *trait* yang sama disebut penyetaraan vertikal. Penyetaraan vertikal dapat digunakan guru untuk mengungkapkan perkembangan kemampuan siswa, walaupun siswa tersebut berada pada tingkat kelas yang berbeda dan memiliki tingkat kemampuan berbeda, asalkan tes yang digunakan mengukur *trait* yang sama.

Penggunaan penyetaraan vertikal pada tingkat kelas yang berbeda pada pelajaran matematika akan membantu guru memperoleh informasi tentang kemampuan siswa pada mata pelajaran tersebut sehingga perkembangan kemampuan siswa tersebut dapat dipantau. Dengan demikian, sangat jelas bahwa betapa pentingnya melakukan penyetaraan vertikal bagi lembaga pendidikan karena melalui penyetaraan vertikal dapat diperoleh informasi tentang kemampuan siswa dan berbagai kecenderungannya. Informasi tersebut akan menjadi acuan dalam usaha perbaikan menuju peningkatan kualitas hasil belajar atau lulusan pada lembaga tersebut.

Penelitian tentang penyetaraan telah banyak dilakukan oleh para peneliti, di antaranya penyetaraan dengan pendekatan *IRT* untuk butir dikotomis dengan metode *mean & mean*, *mean & sigma*, dan *Stocking-Lord* oleh Kim et al. (2009); penyetaraan dengan pendekatan *IRT* menggunakan metode *Stocking-Lord*, *mean & sigma* dan *concurrent calibration* untuk format model campuran oleh Pang et al. (2010); penyetaraan dengan pendekatan Klasik menggunakan metode linear untuk tes model campuran *multiple-*

choice dan *constructed-response* oleh Kim, Walker, & Frederick (2010); penyetaraan dengan pendekatan *IRT* menggunakan metode *characteristic curve* dan *mean & sigma* untuk butir *GPCM* oleh Swediati (1997); penyetaraan dengan pendekatan *IRT* menggunakan metode *stocking-lord*, *mean & sigma* pada respon berjenjang oleh Syaifuddin (2005); penyetaraan dengan pendekatan *IRT* menggunakan metode *mean & mean*, *mean & sigma*, *haebara*, *stocking-lord* pada butir *PCM* oleh Sugeng (2010); dan masih banyak penelitian-penelitian lainnya. Penelitian-penelitian tersebut kebanyakan menggunakan pendekatan *IRT* dan masih sedikit menggunakan pendekatan Klasik.

Hasil penelitian yang telah ada menunjukkan, tidak ada suatu metode yang unggul pada semua kondisi. Suatu metode penyetaraan mungkin akurat pada kondisi tertentu tapi tidak akurat pada kondisi yang lain. Oleh karena itu, masih perlu diteliti lebih lanjut keakuratan berbagai metode penyetaraan yang ada baik dengan pendekatan Klasik maupun *IRT*. Guru perlu memperoleh informasi, metode mana yang memberikan hasil penyetaraan yang akurat pada kondisi yang ada, sehingga informasi yang diperoleh dalam memantau perkembangan kemampuan siswanya mendekati kebenaran.

Dari uraian yang telah disampaikan sebelumnya, penelitian ini bertujuan untuk mengetahui: (1) akurasi metode linear dan ekipersentil pada pendekatan Klasik dan *mean & mean* dan *mean & sigma* pada pendekatan *IRT*; (2) karakteristik tes prestasi belajar matematika model campuran kelas III, IV, V, dan VI Sekolah Dasar yang digunakan dalam penelitian ini; dan (3) perkembangan kemampuan siswa (θ) kelas III, IV, V, dan VI pada pelajaran matematika berdasarkan tes yang digunakan.

Metode Penelitian

Penelitian ini adalah penelitian tentang penyetaraan vertikal yang secara umum menggunakan pendekatan kuantitatif, dan dalam pelaksanaannya terdiri dari beberapa bagian penelitian yang saling terkait yaitu

penyusunan instrumen penelitian, analisis karakteristik instrumen yang disusun, perbandingan keakuratan metode penyetaraan pada pendekatan klasik dan IRT, dan penelitian tentang perkembangan kemampuan siswa.

Penelitian diawali dengan pengembangan instrumen (tes) prestasi belajar matematika model campuran untuk kelas III, IV, V, dan VI Sekolah Dasar yang diujikan pada semester 2 (tes sumatif) yang disusun berdasarkan pokok bahasan bilangan, geometri dan pengukuran sesuai dengan Kurikulum Tingkat Satuan Pendidikan (KTSP). Empat paket tes memuat butir-butir *anchor* sebanyak 20 % baik untuk pilihan ganda maupun uraian. Penentuan Standar Kompetensi, Kompetensi Dasar dan indikator yang diujikan dilakukan melalui *FGD (forum group discussion)*. Peserta *FGD* adalah guru kelas, guru mata pelajaran, ahli bidang studi (dosen matematika), dan penulis. Penyusunan kisi-kisi dan penulisan soal dilakukan oleh tim yang terdiri dari dua orang guru senior mata pelajaran matematika Sekolah Dasar. Validitas isi dan keterbacaan soal melibatkan dua ahli (*expert*) dalam bidang pendidikan matematika dan pengukuran. Validitas isi dianalisis menggunakan teknik Gregory (2000, p.123). Keterbacaan soal melibatkan 6 guru dan 12 orang siswa kelas VI Sekolah Dasar. Revisi soal dilakukan setelah hasil uji keterbacaan dan uji validitas isi dari *expert* diperoleh. Instrumen (tes) yang telah diperbaiki selanjutnya diujicoba di sebelas Sekolah Dasar. Data hasil ujicoba dianalisis menggunakan pendekatan Klasik dan IRT menggunakan program *Parscale* (Muraki & Bock, 1977) dengan estimasi *Marginal Maximum Likelihood (MML)* untuk mengetahui karakteristik butir tes yang disusun sehingga dapat dipilih butir-butir yang baik. Butir-butir yang telah diseleksi pada masing-masing paket tes, selanjutnya dirakit menjadi tes akhir semester yang digunakan sebagai instrumen pengambilan data penelitian.

Pengumpulan data penelitian dilakukan dengan *stratified random sampling* melibatkan sampel sebanyak 1034 terdiri dari 257

siswa kelas III, 257 siswa kelas IV, 260 siswa kelas V, dan 260 siswa kelas VI. Penerapan random dilakukan pada tingkat sekolah, sedangkan penentuan strata sekolah dengan memperhatikan letak sekolah dan katagori sekolah.

Pemenuhan asumsi unidimensi dan validitas konstruk dari tes yang digunakan dilakukan dengan analisis faktor eksploratori dan konfirmatori. Banyaknya dimensi yang diukur oleh keempat tes, dilihat dari *Scree-plot* nilai *Eigen*. Hal ini sesuai dengan pendapat Demars (2010, p. 39) bahwa *eigenvalue* dari inter-item matriks korelasi adalah salah satu metode yang *simple* untuk mengetahui dimensionalitas data. Pengujian kecocokan model hipotetik pengukuran terhadap data empiris menggunakan analisis faktor konfirmatori yang dilakukan menurut peringkat kelas terhadap kelompok butir soal yang terdapat pada tes tersebut. Program yang digunakan adalah Lisrel 8.54 dengan indikator *goodness of fit* (Joreskog & Sorbom, 1996, p. 27) yang meliputi *Non Normed Fit Index (NNFI)*, *Comparative Fit Index (CFI)*, dan *Root Mean Square Error of Approximation (RMSEA)*, dan secara umum dengan statistik χ^2 . Indikator *NNFI* dan *CFI* dengan kriteria lebih dari 0,90 menunjukkan kecocokan yang baik, dan nilai *RMSEA* kurang dari 0,08 menunjukkan kecocokan yang layak. Statistik χ^2 dengan rasio $\chi^2/df \leq 3$.

Desain penyetaraan menggunakan *common items non-equivalent groups* desain atau desain tes *anchor* dengan metode linear dan ekipersentil pada pendekatan Klasik dan dengan metode *mean & mean* dan *mean & sigma* pada pendekatan IRT.

Akurasi metode linear dan ekipersentil diuji dengan *Standard Error of Equating (SEE)* yang dihitung dengan rumus (Linn, 1989, p. 255),

$$SEE [Y(X)] = \{[(\sigma_Y^2/2)] [2 + (X_i - \mu_X)^2/\sigma_X^2] [(1/N_X) + (1/N_Y)]\}^{1/2}$$

dengan,

SEE [Y(X)] adalah standar kesalahan penyetaraan dari tes X ke Y, σ_Y^2 adalah varians skor tes Y, σ_X^2 adalah varians skor tes X, X_i

adalah skor ke i tes X, μ_X adalah rerata skor tes X, N_X adalah jumlah peserta tes X, N_Y adalah jumlah peserta tes Y.

Metode *mean & mean* dan *mean & sigma* akurasi diuji dengan *Root Mean Square Different* (RMSD) kemampuan siswa (θ) sebelum dan sesudah penyetaraan, yang dihitung dengan rumus (Kim & Cohen, 2002, p. 31),

$$RMSD(\theta) = \sqrt{\frac{\sum_{i=1}^N (\hat{\theta}_i - \theta_i)^2}{N}}$$

dengan,

N adalah banyaknya *examinee*, $\hat{\theta}_i$ adalah kemampuan peserta ke i setelah disetarakan, dan θ_i adalah kemampuan peserta ke i sebelum disetarakan.

Informasi tentang perkembangan kemampuan siswa dalam pelajaran matematika dapat diperoleh dari hasil penyetaraan skor dan skala kemampuan siswa pada keempat kelas yang dilibatkan.

Hasil Penelitian dan Pembahasan

Data penelitian yang berupa skor siswa pada tes prestasi belajar matematika model campuran pada ujian akhir semester kelas III, IV, V, dan IV Sekolah Dasar dikoreksi oleh dua orang *rater* untuk mendapatkan skor yang baik. Untuk menjamin konsistensi penilaian, skor dari dua orang *rater* tersebut diestimasi reliabilitasnya dengan menggunakan formula *inter-rater* dengan pendekatan *Hoyt* (Mardapi, 2012, p. 86).

Koefisien reliabilitas (r_{11}) *inter-rater* skor tes kelas III, IV, V, dan IV diperoleh berturut-turut sebesar 0,992, 0,823, 0,923 dan 0,926. Nilai-nilai koefisien tersebut semuanya berada di atas 0,700. Hal tersebut berarti unsur subjektivitas dari masing-masing *rater* (penilai) tidak berpengaruh terhadap pemberian skor pada ke empat tes, sehingga skor yang diberikan dapat digunakan sebagai data penelitian.

Uji Unidimensi dan Validitas Konstruk

Nilai *Chi-Square* pada uji Bartlet tes kelas III, sebesar 1217,326 dengan derajat kebebasan 595 dan nilai-p kurang dari 0,01. Hasil ini menunjukkan bahwa ukuran sampel sebesar 257 yang digunakan pada penelitian telah mencukupi. *Scree-plot* nilai *Eigen* tes kelas III menunjukkan nilai *Eigen* mulai landai pada faktor ke dua. Hal ini berarti terdapat satu faktor yang dominan pada tes kelas III. Faktor-faktor lain yang berpengaruh berkaitan dengan faktor dominan tersebut. Hasil *running* dengan program Lisrel mendapatkan nilai *chi-square* sebesar 94,63 dengan $df = 74$ dan $p = 0,053$, *Root Mean Square Error Approximation* (RMSEA) = 0,053, *Non-Normed Fit Index* (NNFI) = 0,949, *Comparative Fit Index* (CFI) = 0,958 dan $\chi^2/df = 94,63/74 = 1,28 < 3$ yang berarti model tes kelas III *fit* dengan data.

Nilai *Chi-Square* pada uji Bartlet tes kelas IV sebesar 1300,659 dengan derajat kebebasan 595 dan nilai-p kurang dari 0,01. Hasil ini menunjukkan bahwa ukuran sampel sebesar 257 yang digunakan pada penelitian juga telah cukup. *Scree-plot* nilai *Eigen* tes kelas IV menunjukkan nilai *Eigen* mulai landai pada faktor ke dua. Hasil *running* dengan program Lisrel mendapatkan nilai *chi-square* diperoleh sebesar 94,58 dengan $df = 74$ dan $p = 0,054$, *Root Mean Square Error Approximation* (RMSEA) = 0,053, *Non-Normed Fit Index* (NNFI) = 0,953, *Comparative Fit Index* (CFI) = 0,962 dan $\chi^2/df = 94,58/74 = 1,28 < 3$. Hal ini menunjukkan model tes kelas IV *fit* dengan data.

Nilai *Chi-Square* pada uji Bartlet tes kelas V sebesar 1733,177 dengan derajat kebebasan 595 dan nilai-p kurang dari 0,01. Hasil ini menunjukkan bahwa ukuran sampel sebesar 260 yang digunakan pada penelitian telah cukup. *Scree-plot* tes kelas V menunjukkan nilai *Eigen* pada tes kelas V mulai landai pada faktor ke dua. Hasil *running* dengan program Lisrel mendapatkan nilai *chi-square* sebesar 93,24 dengan $df = 74$ dan $p = 0,065$, *Root Mean Square Error Approximation* (RMSEA) = 0,053, *Non-*

Normed Fit Index (NNFI) = 0,951, *Comparative Fit Index* (CFI) = 0,960 dan $\chi^2/df = 93,24/74 = 1,26 < 3$. Hal ini menunjukkan model tes kelas V *fit* dengan data.

Nilai *Chi-Square* pada uji Bartlett tes kelas VI sebesar 1746,443 dengan derajat kebebasan 595 dan nilai-p kurang dari 0,01. Hal ini berarti ukuran sampel sebesar 260 yang digunakan pada penelitian telah cukup. *Scree-plot* tes kelas VI menunjukkan bahwa nilai *Eigen* pada tes kelas VI mulai landai pada faktor ke dua, yang berarti hanya satu faktor yang dominan yaitu kemampuan matematika. Hasil *running* dengan program Lisrel memperoleh nilai *chi-square* sebesar 122,07 dengan $df = 100$ dan $p = 0,06613$, *Root Mean Square Error Approximation* (RMSEA) = 0,045, *Non-Normed Fit Index* (NNFI) = 0,963, *Comparative Fit Index* (CFI) = 0,969 dan $\chi^2/df = 122,07/100 = 1,22 < 3$. Hal ini juga berarti model tes kelas VI *fit* dengan data.

Analisis Butir Soal

Analisis butir dengan pendekatan Klasik meliputi korelasi *Pearson* (r_p) dan *Polyserial* (r_{ps}) yang dibaca pada *output Parscale Fase 1 (PH1)*. Analisis butir tes dengan pendekatan IRT meliputi parameter *slope*, *location*, dan *probability* yang dibaca pada *output Parscale Fase 2 (PH2)*. Rangkuman hasil analisis keempat paket tes disajikan pada Tabel 1, Tabel 2, dan Tabel 3.

Tabel 1 menunjukkan, rerata skor dari tes kelas III, IV, V, dan VI berturut-turut

adalah 25,961, 26,770, 27,25 dan 27,660 berarti di atas rerata ideal (24,50). Varians skor keempat tes cukup besar ini berarti distribusi skor cukup besar. Distribusi skor siswa untuk tes kelas III, IV, dan V membentuk *curve* sedikit juling ke kanan, hal ini dapat dilihat dari nilai rerata sedikit lebih besar dari median. Distribusi skor untuk tes kelas VI membentuk *curve* sedikit juling ke kiri, karena nilai reratanya lebih rendah dari median. Hal ini menunjukkan sebagian besar siswa kelas VI mendapat skor tinggi. Nilai rerata koefisien korelasi *Pearson* dan rerata koefisien korelasi *Polyserial* keempat tes juga berada di atas 0,2 yang berarti keempat tes secara umum memiliki daya beda yang baik.

Tabel 1. Analisis Butir Tes Kelas III, IV, V, dan VI dengan Pendekatan Klasik

Komponen	Tes Kls III	Tes Kls IV	Tes Kls V	Tes Kls VI
Mean	25,96	26,770	27,25	27,660
Varians	30,045	31,146	46,127	46,822
St. Dev.	5,481	5,581	6,792	6,843
<i>Skewness</i>	0,431	0,285	-0,001	-0,090
<i>Kurtosis</i>	-0,034	-0,086	-0,585	-0,614
Skor Min	12	13	10	10
Skor Maks	41	45	42	42
Median	25	26	27	28
Rerata (r_p)	0,298	0,317	0,369	0,373
Rerata (r_{ps})	0,367	0,391	0,455	0,460
Panjang Tes	35	35	35	35
Subjek	257	257	260	260

Tabel 2. Analisis Parameter Butir Tes Kelas III dan IV dengan Pendekatan IRT

Komponen	Tes Kelas III			Tes Kelas IV		
	<i>Slope</i>	<i>Location</i>	<i>Prob.</i>	<i>Slope</i>	<i>Location</i>	<i>Prob.</i>
Mean	0,537	-0,070	0,412	0,558	-0,228	0,513
St.Dev.	0,234	0,330	0,231	0,250	0,380	0,277
Varians	0,055	0,109	0,055	0,062	0,145	0,077
Min	0,247	-0,951	0,077	0,295	-1,333	0,052
Maks	1,022	0,525	0,915	1,360	0,536	0,956
<i>Skewness</i>	0,784	-0,498	0,388	1,480	-0,880	-0,188
<i>Kurtosis</i>	-0,610	0,099	-0,920	1,956	1,532	-1,092
Median	0,464	-0,033	0,377	0,475	-0,183	0,498
Panjang Tes	35			35		
Responden	257			257		

Tabel 2 dan Tabel 3 menunjukkan, nilai rerata *location* dari keempat tes berturut-turut sebesar -0,070, -0,228, -0,324, dan -0,430, semuanya berada di sekitar titik nol yang berarti keempat tes memiliki tingkat kesukaran butir yang sedang. Nilai rerata parameter *slope* tes kelas III, IV, V dan VI berturut-turut adalah 0,537, 0,558, 0,639, dan 0,644 ($>0,2$), berarti keempat tes memiliki daya beda butir yang baik. Nilai minimum *probability* semua butir pada keempat

tes $>0,05$ yang berarti semua butir tes *fit* dengan model.

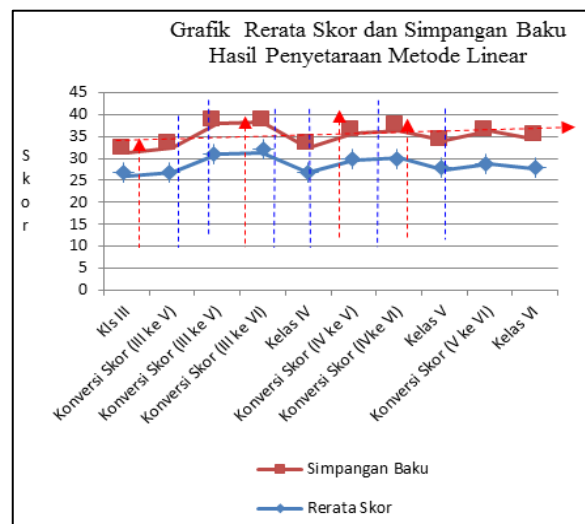
Informasi hubungan antara fungsi informasi tes dengan *Standard Error of Measurement* pada keempat tes memiliki *error* rendah pada rentang skala dari -2 sampai dengan +2, artinya tes akan memberikan informasi yang optimal jika digunakan untuk mengukur kemampuan siswa pada rentang kemampuan antara -2 sampai dengan +2. Hal ini sesuai dengan pendapat Hambleton, Swaminathan, & Rogers (1991, p. 13).

Tabel 3. Hasil Analisis Parameter Butir Tes Kelas V dan VI dengan IRT

Komponen	Tes Kelas V			Tes Kelas VI		
	<i>Slope</i>	<i>Location</i>	<i>Prob.</i>	<i>Slope</i>	<i>Location</i>	<i>Prob.</i>
Rerata	0,639	-0,324	0,417	0,644	-0,430	0,590
St.Dev,	0,310	0,310	0,229	0,317	0,309	0,218
Varians	0,096	0,096	0,053	0,100	0,095	0,047
Min.	0,252	-1,189	0,055	0,255	-1,243	0,198
Maks.	1,447	0,081	0,841	1,502	0,029	0,962
<i>Skewness</i>	1,375	-0,786	0,068	1,364	-0,881	-0,001
<i>Kurtosis</i>	1,216	0,146	-1,147	1,242	0,575	-0,893
Median	0,539	-0,237	0,421	0,533	-0,351	0,586
Panjang Tes	35			35		
Responden	260			260		

Hasil Penyetaraan

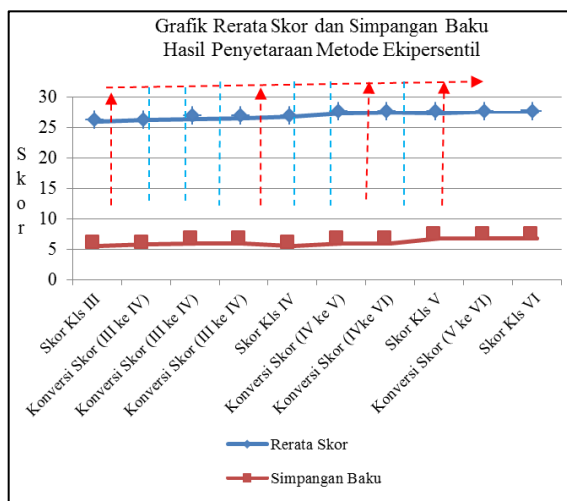
Telah dijelaskan sebelumnya, bahwa penyetaraan vertikal dengan pendekatan klasik menggunakan metode linear dan ekipersentil. Penyetaraan vertikal dengan metode linear mendapatkan persamaan konversi: (1) dari kelas III ke kelas IV adalah $I_Y(X) = 1,027 X + 0,094$; (2) dari kelas III ke kelas V adalah $I_Y(X) = 1,286 X - 2,014$; (3) dari kelas III ke kelas VI adalah $I_Y(X) = 1,262 X - 1,656$; (4) dari kelas IV ke kelas V adalah $I_Y(X) = 1,144 X - 0,553$; (5) dari kelas IV ke kelas VI adalah $I_Y(X) = 1,120 X - 0,411$; dan dari kelas V ke kelas VI adalah $I_Y(X) = 1,098 X - 1,034$. Rerata skor dan simpangan baku hasil penyetaraan disajikan dengan grafik seperti tampak pada Gambar 1. Besarnya *SEE* penyetaraan dengan metode linear diperoleh sebesar 0,608.



Gambar 1. Hasil Penyetaraan dengan Metode Linear

Penyetaraan skor dengan metode ekipersentil pada dasarnya adalah mendudukkan skor siswa dari keempat tes pada per-

sentil yang sama. Hasil penyetaraan dan simpangan bakunya disajikan dengan grafik pada Gambar 2. Grafik pada Gambar 2 menunjukkan, rerata skor siswa mengalami peningkatan seiring dengan meningkatnya peringkat kelas. Hal ini berarti makin tinggi peringkat kelas kemampuan siswa semakin meningkat. Besarnya nilai *SEE* pada penyetaraan dengan metode ekipersentil diperoleh sebesar 0,505.



Gambar 2. Hasil Penyetaraan Metode Ekipersentil

Penyetaraan dengan pendekatan *IRT* model *GPCM* meliputi penyetaraan parameter daya beda butir atau *slope* (a), parameter tingkat kesukaran butir atau *location* (b) dan parameter kemampuan siswa atau *ability* (θ). Konstanta penyetaraan a dan β pada metode *mean & mean* dihitung berdasarkan *mean* dari daya beda butir dan tingkat kesukaran butir, sedangkan pada metode *mean & sigma* konstanta a dan β dihitung dengan *mean* dan simpangan baku dari parameter *location*.

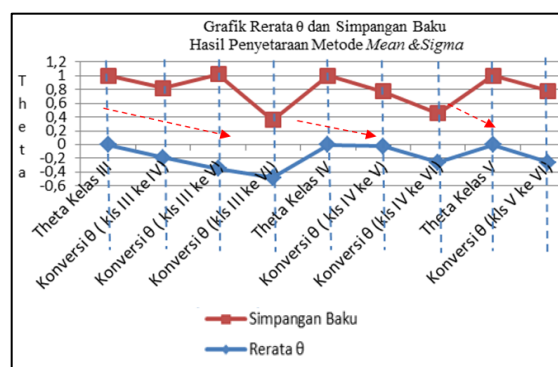
Hasil penyetaraan parameter *slope* dan *location* metode *mean & mean* disajikan pada tabel 4 dan Tabel 5, sedangkan hasil penyetaraan parameter kemampuan θ disajikan dengan grafik pada Gambar 3. Akurasi penyetaraan dihitung berdasarkan nilai *RMSD* kemampuan sebelum dan sesudah penyetaraan. Besarnya nilai *RMSD* pada penyetaraan vertikal dengan metode *mean & mean* diperoleh sebesar 0,394.

Tabel 4. Rerata *Slope* dan Simpangan Baku hasil Penyetaraan Tes Kelas III, IV, V, dan VI Metode *Mean & Mean*

Penyetaraan	Rerata <i>Slope</i>	Simpangan Baku
<i>Slope</i> Kls III	0,537	0,234
Konversi <i>Slope</i> (III ke IV)	0,518	0,222
Konversi <i>Slope</i> (III ke V)	0,624	0,277
Konversi <i>Slope</i> (III ke VI)	0,637	0,280
<i>Slope</i> Kls IV	0,558	0,250
Konversi <i>Slope</i> (IV ke V)	0,753	0,367
Konversi <i>Slope</i> (IV ke VI)	0,799	0,365
<i>Slope</i> Kelas V	0,639	0,310
Konversi <i>Slope</i> (V ke VI)	0,672	0,326
<i>Slope</i> Kls VI	0,644	0,316

Tabel 5. Rerata *Location* dan Simpangan Baku Tes Kelas III, IV, V, dan VI Hasil Penyetaraan Metode *Mean & Mean*

Penyetaraan	Rerata <i>location</i>	Simpangan Baku
Tes Kelas III	-0,070	0,330
Konversi Loc. (III ke IV)	-0,267	0,338
Konversi Loc. (III ke V)	-0,482	0,286
Konversi Loc. (III ke VI)	-0,543	0,272
Tes Kelas IV	-0,228	0,380
Konversi Loc. (IV ke V)	-0,207	0,180
Konversi Loc. (IV ke VI)	-0,416	0,254
Tes Kelas V	-0,324	0,310
Konversi Loc. (V ke VI)	-0,585	0,296
Tes Kelas VI	-0,430	0,305



Gambar 3. Grafik Konversi Kemampuan Siswa antar-Peringkat Kelas Metode *Mean & Mean*

Hasil penyetaraan vertikal parameter *location* dan *slope* pada metode *mean & sigma* disajikan pada Tabel 6 dan Tabel 7, sedang-

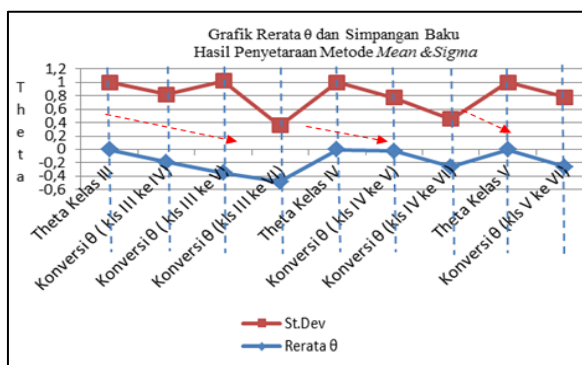
kan hasil penyetaraan parameter kemampuan θ disajikan dengan grafik pada Gambar 4. Besarnya nilai *RMSD* metode *mean & sigma* adalah 0,344474.

Tabel 6. Rerata *Location* dan Simpangan Baku Tes Kelas III, IV, V, dan VI Hasil Penyetaraan Metode *Mean & Sigma*

Penyetaraan	Rerata <i>Location</i> (b)	Simpangan Baku
<i>Location</i> Kls III	-0,070	0,330
Konversi <i>Location</i> (III-IV)	-0,262	0,330
Konversi <i>Location</i> (III-V)	-0,448	0,442
Konversi <i>Location</i> (III-VI)	-0,543	0,272
<i>Location</i> Kls IV	-0,207	0,294
Konversi <i>Location</i> (IV-V)	-0,228	0,380
Konversi <i>Location</i> (IV-VI)	-0,390	0,297
<i>Location</i> Kls V	-0,324	0,310
Konversi <i>Location</i> (V-VI)	-0,595	0,321
<i>Location</i> Kls VI	-0,430	0,308

Tabel 7. Rerata *Slope* dan Simpangan Baku Tes Kelas III, IV, V, dan VI Hasil Penyetaraan Metode *Mean & Sigma*

Penyetaraan	Rerata <i>Slope</i> (a)	Simpangan Baku
<i>Slope</i> Kls III	0,537	0,234
Konv. <i>Slope</i> (III-IV)	0,682	0,442
Konv. <i>Slope</i> (III-V)	0,416	0,188
Konv. <i>Slope</i> (III-VI)	0,637	0,280
<i>Slope</i> Kls IV	0,558	0,250
Konv. <i>Slope</i> (IV-V)	0,738	0,307
Konv. <i>Slope</i> (IV-VI)	0,790	0,361
<i>Slope</i> Kls V	0,639	0,310
Konv. <i>Slope</i> (V-VI)	0,622	0,298
<i>Slope</i> Kls VI	0,638	0,317



Gambar 4. Grafik Rerata θ dan Simpangan Baku Hasil Penyetaraan Metode *Mean & Sigma*

Pembahasan

Pada penyetaraan dengan metode linier (Gambar 1) tampak bahwa rerata skor siswa akan naik jika dikonversi ke tingkat kelas di atasnya. Hal tersebut juga berarti bahwa kemampuan siswa mengalami peningkatan seiring dengan meningkatnya peringkat kelas. Namun, jika dicermati persamaan konversi dari kelas III ke V, kelas III ke VI, kelas IV ke V, kelas IV ke VI dan kelas V ke VI maka dapat dilihat siswa yang memiliki skor rendah dan skor tinggi mempunyai arah konversi yang berbeda. Misalnya konversi skor kelas III ke kelas VI memberi informasi bahwa siswa yang mendapat skor 6 ke bawah pada kelas III akan menjadi lebih rendah jika skor tersebut dikonversi ke kelas VI, sedangkan siswa yang mendapat skor 7 ke atas, skornya akan menjadi lebih besar. Persoalan yang sama bisa diselidiki pada persamaan konversi yang lain. Hal ini menunjukkan ketidakkonsistenan terhadap hasil konversi dengan menggunakan metode linear.

Penyetaraan skor dengan metode ekipersentil memberikan informasi yang lebih konsisten dibandingkan metode linear. Hal tersebut dapat dilihat dari grafik rerata skor hasil konversi yang semuanya mempunyai kecenderungan yang sama setelah dilakukan penyetaraan. Rerata skor kelas VI lebih tinggi dari rerata skor kelas V, lebih tinggi dari rerata skor kelas IV, dan lebih tinggi dari rerata skor kelas III, hal yang sama juga berlaku pada kelas V dan IV. Hal ini dapat diartikan bahwa kemampuan siswa mengalami peningkatan sesuai dengan meningkatnya peringkat kelas. Hal tersebut tentu sangat wajar, karena pengetahuan siswa semestinya mengalami peningkatan sesuai dengan pengalaman belajarnya.

Penyetaraan dengan pendekatan *IRT* adalah penyetaraan skala parameter butir suatu tes terhadap skala parameter butir tes lainnya. Hasil penelitian menunjukkan bahwa konversi nilai parameter *location* atau tingkat kesukaran butir menunjukkan hasil yang konsisten baik pada metode *mean & mean* maupun metode *mean & sigma*. Informasi tersebut dapat dilihat pada grafik per-

bandingan nilai parameter *location* sebelum dan sesudah penyetaraan. Nilai rerata parameter *location* menurun jika dikonversi ke skala peringkat di atasnya. Hal ini dapat diartikan bahwa tes kelas III menjadi lebih mudah jika dikerjakan oleh siswa kelas IV, V atau VI. Tes kelas IV menjadi lebih mudah jika dikerjakan oleh siswa kelas V atau VI dan tes kelas V akan lebih mudah jika dikerjakan oleh siswa kelas VI. Sebaliknya, tes kelas VI akan lebih sulit jika dikerjakan oleh siswa kelas V, IV atau III dan seterusnya.

Pada umumnya *slope* dari keempat tes tidak jauh berubah jika dikonversi pada skala tes yang lainnya. Pada metode *mean & mean* indeks daya beda tes kelas III sedikit menurun jika dikonversi ke kelas IV, sedikit lebih baik jika dikonversi ke kelas V atau VI. Indeks daya beda tes kelas IV sedikit lebih baik jika dikonversi ke kelas V, atau VI. Demikian juga tes kelas V daya bedanya sedikit lebih baik jika dikonversi ke kelas VI. Pada metode *mean & sigma*, indeks daya beda tes kelas III sedikit lebih baik jika dikonversi ke kelas IV, sedikit lebih menurun jika dikonversi ke kelas V dan sedikit lebih baik jika dikonversi ke kelas VI. Indeks daya beda tes kelas IV sedikit lebih baik jika dikonversi ke kelas V maupun kelas VI. Indeks daya beda tes kelas V sedikit menurun jika dikonversi ke kelas VI. Secara umum rerata nilai parameter daya beda dari keempat tes mempunyai nilai rata-rata relatif sama. Rerata *slope* tes kelas III, IV, V, dan VI setelah konversi berturut-turut sebesar 0,545, 0,554, 0,523 dan 0,557.

Persamaan konversi kemampuan (θ) mengikuti persamaan ciri butir. Dijelaskan di atas baik pada metode *mean & mean* maupun pada metode *mean & sigma* kemampuan siswa akan menurun jika dikonversi pada skala kemampuan di atasnya. Informasi tersebut dapat dilihat dari grafik konversi rerata kemampuan siswa yang selalu menurun jika dikonversi pada skala di atasnya. Informasi tersebut juga dapat diartikan bahwa rerata kemampuan siswa pada kelas VI lebih tinggi dari rerata kemampuan siswa kelas V. Rerata kemampuan siswa kelas V lebih ting-

gi dari rerata kemampuan siswa kelas IV dan rerata kemampuan siswa kelas IV lebih tinggi dari rerata kemampuan siswa kelas III. Hal ini diperkuat dengan perbandingan rerata kemampuan siswa hasil kalibrasi *concurrent* yang menunjukkan besarnya rerata kemampuan siswa kelas III, IV, V, dan VI berturut-turut sebesar: -0,104, -0,076, 0,004, dan 0,004.

Hasil perhitungan *SEE* pada metode linear mendapatkan besarnya kesalahan standar penyetaraan (*SEE*) sebesar 0,608 sedangkan penyetaraan dengan metode ekipersentil didapat kesalahan standar penyetaraan (*SEE*) sebesar 0,505. Perbedaan nilai *SEE* sebesar 0,102 tidak begitu besar sehingga kebermaknaannya untuk membedakan akurasi kedua masih perlu diteliti.

Hasil perhitungan *RMSD* pada metode *mean & smean* nilai *RMSD* sebesar 0,394 sedangkan penghitungan *RMSD* pada penyetaraan dengan metode *mean & sigma* mendapatkan nilai sebesar 0,344. Kedua nilai *RMSD* tersebut juga tidak jauh berbeda (hanya berselisih 0,049). *RMSD* metode *mean & sigma* sedikit lebih kecil dibandingkan metode *mean & mean*. Kebermaknaan perbedaan nilai *RMSD* untuk membedakan akurasi kedua metode tersebut masih perlu diteliti lebih lanjut melalui penelitian simulasi.

Simpulan dan Saran

Simpulan

Penyetaraan vertikal dengan pendekatan Klasik memberikan informasi bahwa *Standart Error of Equating (SEE)* pada penyetaraan vertikal dengan metode ekipersentil sebesar 0,505 dan *SEE* penyetaraan vertikal dengan metode linear sebesar 0,608. Metode ekipersentil menghasilkan *SEE* lebih kecil dibandingkan metode linear, namun perbedaan nilai *SEE* tersebut sangat kecil sehingga tidak bermakna dalam membedakan akurasi metode linear dan ekipersentil.

Penyetaraan vertikal dengan pendekatan *IRT* memberikan informasi bahwa *Root Mean Square Different (RMSD)* kemampuan (θ) pada penyetaraan vertikal dengan

metode *mean & sigma* sebesar 0,344 dan Nilai *RMSD* kemampuan pada penyetaraan vertikal dengan metode *mean & mean* sebesar 0,394. Nilai *RMSD* yang dihasilkan metode *mean & sigma* lebih kecil dibandingkan metode *mean & mean*, namun, perbedaan nilai *RMSD* tersebut juga sangat kecil sehingga tidak bermakna dalam membedakan akurasi kedua metode.

Penelitian ini menghasilkan empat perangkat tes yaitu tes prestasi belajar matematika semester dua Sekolah Dasar untuk kelas III, kelas IV, kelas V, dan kelas VI yang telah memenuhi syarat sebagai tes yang baik setelah melalui analisis kualitatif maupun kuantitatif. Keempat tes memiliki nilai *location (b)* atau tingkat kesukaran butir kategori sedang ($-2 < b < 2$) dengan nilai *location* berturut-turut sebesar -0,228, -0,324, -0,430, dan -0,430, memiliki daya beda atau *slope (a)* yang baik ($a > 0,2$) dengan nilai rerata *slope* berturut-turut sebesar 0,537, 0,558, 0,639, dan 0,644, serta memiliki *items fit statistics (p)* yang baik ($p > 0,05$) dengan nilai *p* minimum berturut-turut sebesar 0,077, 0,052, 0,055, dan 0,198.

Keempat metode penyetaraan yang digunakan dalam penelitian ini yaitu metode linear, metode ekipersentil, metode *mean & mean*, dan metode *mean & sigma* menunjukkan bahwa rerata kemampuan siswa mengalami peningkatan seiring dengan meningkatnya peringkat kelas. Hal tersebut diperkuat oleh hasil kalibrasi *concurrent* yang menunjukkan perbandingan rerata kemampuan siswa kelas III, IV, V, dan VI mengalami peningkatan setelah dilakukan penyetaraan dengan nilai θ berturut-turut sebesar -0,104, -0,076, 0,004, dan 0,173. Namun, kenaikan nilai θ dari suatu peringkat ke peringkat di atasnya sangat kecil sehingga masih perlu diteliti kebermaknaannya. Kenaikan tertinggi terjadi pada kelas VI yang disebabkan oleh faktor latihan menjelang Ujian Sekolah.

Saran

Hasil penelitian ini memberikan sumbangan pikiran berkaitan dengan kualitas penyetaraan vertikal yang menggunakan dua pendekatan yaitu pendekatan Klasik dan pendekatan *IRT*. Penyetaraan dengan pen-

dekatan Klasik yang membandingkan metode linear dan ekipersentil menunjukkan hasil yang tidak terlalu jauh berbeda. Demikian juga penyetaraan dengan pendekatan *IRT*, yang membandingkan akurasi metode *mean & mean* dan *mean & sigma* hasilnya tidak terlalu jauh berbeda. Perlu dilakukan penelitian simulasi untuk mengetahui signifikansi perbedaan akurasi metode-metode tersebut.

Penyetaraan vertikal dengan pendekatan Klasik dan *IRT* dapat dipertimbangkan penggunaannya dalam tes model campuran. Hal itu dapat digunakan untuk mengamati pertumbuhan kemajuan belajar siswa dari tahun ke tahun (sesuai periode pengukuran), mengamati perbandingan tingkat kemampuan siswa, mengamati perbandingan tingkat kesukaran butir melalui kelompok butir *anchor* menurut jenjang kelas berbeda khususnya bidang studi matematika dan IPA.

Keempat metode penyetaraan memberikan informasi bahwa peningkatan kemampuan siswa dari satu tingkat ke tingkat di atasnya tidak besar sehingga kemungkinan hanya disebabkan oleh faktor umur dan pengalaman belajar. Oleh karena itu, perlu dilakukan perbaikan kualitas program pembelajaran untuk memperoleh peningkatan yang lebih baik.

Penyetaraan vertikal dengan pendekatan Klasik dan *IRT* dapat dipertimbangkan dalam pengembangan bank soal matematika. Hal itu sekaligus sebagai pelapis keamanan dalam pelaksanaan tes.

Penerapan metode linear, ekipersentil, *mean & mean* dan *mean & sigma* memberikan pertimbangan pemilihan metode yang lebih baik. Pemilihan metode melibatkan berbagai program untuk menunjang analisis data, ke-tersediaan dan penguasaan aplikasi program sangat diperlukan.

Daftar Pustaka

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehard and Winston Inc.

- DeMars, C. (2002). Incomplete data and item parameter estimates under JM-LE and MML estimation. *Applied Measurement in Education*, 15, 15-31.
- Gregory, R.J. (2007). *Psychological testing. History, principles, and applications* (4th ed). Boston: Pearson Education, Inc.
- Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer Inc.
- Hambleton, R.K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication Inc.
- Joreskog, K.G. & Sorbom, D. (1996). *LISREL 8: Structural Equation Modeling*. Chicago: Scientific Software International.
- Kemendikbud. (2013). *Kurikulum 2013. Kompetensi dasar (SD)/madrasah ibtidaiyah (MI)*
- Kim, S.H., & Cohen, A.S. (2002). A comparison of linking and concurrent calibration under the grade response model. *Applied Psychological Measurement*, 26, 25-61.
- Kim, S., & Hanson, B.A. (2002). Test equating under the multiple-choice model. *Applied Psychological Measurement*, 26, 255-270.
- Kim, S., Walker, M.E., & Frederick, M. (2010). Comparisons among designs for equating mixed-format tests in large-scale assessments. *Journal of Educational Measurement*, 47, 36-53.
- Kim, J. et. al, (2009, April). *Investigation of vertical scaling using the rasch model*. Paper, presented at the annual meeting of the National Council on Measurement in Education.
- Kolen, M.J., & Brennan, R.L (1995). *Test equating: Methods and practices*. New York: Springer-verlag.
- Mardapi, D. (2008). *Teknik penyusunan tes dan nontes*. Yogyakarta: Mitra Cendikia
- Mardapi, D. (2012). *Pengukuran, penilaian, & evaluasi pendidikan*. Yogyakarta: Nuha Medika
- Muraki, E., & Bock, R.D. (1997). *Parscale : IRT item analysis and test scoring for rating-scale data*. Chicago: Scientific Software International.
- Naga. D.S. (1992). *Pengantar teori skor pada pengukuran pendidikan*. Jakarta: Guna-darma.
- Pang, X., et.al, (2010). A comparison of four test equating methods. *Report Prepared*. Ontario: Education Quality and Accountability office
- Rasiman. (2008, November). *Pengembangan bahan ajar berdasarkan perkembangan kognitif untuk meningkatkan hasil belajar matematika siswa SD*. Makalah disajikan dalam Seminar Nasional Matematika dan Pendidikan Matematika di Universitas Negeri Yogyakarta
- Sugeng. (2010). *Penyetaraan vertikal model kredit parsial soal matematika SMP*. Disertasi Doktor, tidak diterbitkan. Program Pascasarjana, UNY Yogyakarta.
- Swediati, N. (1997). *Equating test under the GPCM*. Disertasi Doktor, tidak diterbitkan. University of Massachusetts, Amherst.
- Syaipuddin. (2005). *Penyetaraan tes model respons berjenjang*. Disertasi Doktor, tidak diterbitkan. Yogyakarta: Program Pascasarjana Universitas Negeri Yogyakarta.

EVALUASI PROGRAM PEMBELAJARAN IPA SMP MENGUNAKAN MODEL COUNTENANCE STAKE

Astin Lukum

Universitas Negeri Gorontalo
astin.lukum@yahoo.com

Abstrak

Tujuan penelitian ini adalah mengevaluasi program pembelajaran IPA SMP di Kabupaten Bone Bolango berdasarkan Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 65 Tahun 2013 tentang Standar Proses Pendidikan Dasar dan Menengah. Penelitian ini merupakan penelitian evaluasi dengan model Countenance Stake. Pengumpulan data menggunakan teknik, observasi, wawancara dan dokumentasi. Teknik analisis yang digunakan adalah deskriptif kualitatif. Kesimpulan hasil penelitian sebagai berikut: (1) perencanaan pembelajaran IPA termasuk kategori cukup (68%), ditemukan belum adanya kesesuaian antara RPP dengan standar proses pembelajaran; (2) pelaksanaan pembelajaran IPA termasuk kategori cukup (57%), belum adanya kesesuaian antara pelaksanaan pembelajaran dengan standar proses pelaksanaan pembelajaran; dan (3) hasil belajar peserta didik belum memenuhi Kriteria Ketuntasan Minimal (KKM) dengan aktualitas ketercapaian 65% kategori cukup. 4) terdapat *contingency* antara perencanaan, pelaksanaan dengan hasil belajar IPA, yang perencanaan pembelajaran dengan kategori yang cukup menyebabkan guru melaksanakan pembelajaran belum sesuai standar proses sehingga hasil belajar peserta didik belum memenuhi KKM.

Kata Kunci: *Evaluasi Program, Countenance Stake, Pembelajaran IPA*

THE EVALUATION OF SCIENCE TEACHING ON JUNIOR HIGH SCHOOL USING STAKE'S COUNTENANCE MODEL

Astin Lukum

Universitas Negeri Gorontalo
astin.lukum@yahoo.com

Abstract

The purpose of the study was to describe the science learning program on junior high school in Bone Bolanga district based on the Regulation of Minister of Education and Culture of the Republic of Indonesia, Number 65 of 2013 about Processing Standard of Primary and Secondary Education. This study used Stake's Countenance evaluation model. The data were collected using observation, interview and documentation techniques. The conclusion was: (1) the planning of science learning was categorized *fair* (68%), it was found that lesson plan was not in accordance with the learning processing standard. (2) The implementation of science learning was categorized fair (57%), that unconformitted with learning processing implementation standard. (3) Student learning outcomes have not met the completeness of minimum criteria (KKM) that categorized enough (65%) and (4) There were the contingency of planing learning proces and outcome.

Keywords: *Program Evaluation, Stake's Countenance, Science Learning*

Pendahuluan

Penyelenggaraan pembelajaran IPA merupakan tugas utama guru IPA yang didesain berdasarkan kurikulum dan silabus yang bertujuan untuk menjadikan peserta didik berubah pola pikir, sikap dan keterampilannya. Dalam proses pembelajaran aktivitas peserta didik merupakan hal utama yang menjadi fokus perhatian guru. Namun, belum semua guru IPA melaksanakan pembelajaran yang berpusat pada peserta didik, masih banyak guru melakukan proses pembelajaran yang berpusat pada guru, peserta didik menjadi pendengar yang baik, peserta didik takut bertanya tentang IPA.

IPA merupakan kumpulan pengetahuan tentang objek atau gejala-gejala tentang alam. IPA sebagai proses yang dikenal dengan metode ilmiah. Di samping itu, IPA juga memiliki nilai-nilai ilmiah atau *value of science* yang melekat pada pengetahuan ilmiah (Paramata, 2001, p.2). Lukum (2013, p.64) mengemukakan bahwa IPA merupakan proses ilmiah yang bersifat empiris, sistematis, dan logis serta sikap ilmiah seperti sikap ingin tahu, menghargai pembuktian, sabar, kritis, tidak putus asa, kreatif dan berdaya cipta. IPA juga memiliki karakteristik dalam cara mempelajarinya yang berbeda dengan cara-cara mempelajari ilmu pengetahuan lainnya (Rustaman, 2010, p.5). Hal ini memberikan pengertian bahwa IPA merupakan pengetahuan yang dibangun berdasarkan pengamatan dan klasifikasi data yang melibatkan aplikasi penalaran matematis dan analisis data terhadap gejala-gejala alam.

Untuk melaksanakan pembelajaran IPA sesuai dengan hakikatnya telah diupayakan melalui metode penyelidikan ilmiah (NSTA, 2004, p.1). Namun demikian menurut McBride (2004, p.3), bahwa penerapan pembelajaran yang berbasis inkuiri belum optimal di lapangan. Hal ini disebabkan oleh waktu pembelajaran, penguasaan konten, penilaian hanya berorientasi pada kemampuan kognitif dibanding sikap dan keterampilan. Proses pembelajaran inkuiri, mengandung revolusi mental yang lebih tinggi tingkatannya, misalnya merumuskan proble-

ma sendiri, merancang eksperimen, melakukan eksperimen, mengumpulkan dan menganalisis data, menarik kesimpulan, mempunyai sikap-sikap objektif, jujur, hasrat ingin tahu, terbuka, dan sebagainya.

Oleh sebab itu, pembelajaran IPA yang dikembangkan dalam Kurikulum 2013 adalah IPA sebagai mata pelajaran *integrative science*, bukan sebagai pendidikan disiplin ilmu. Pendidikan yang berorientasi aplikatif, pengembangan kemampuan berpikir, kemampuan belajar, rasa ingin tahu, dan pengembangan sikap peduli dan bertanggung jawab terhadap lingkungan sosial dan alam (Depdikbud, 2013, p.10). Pembelajaran IPA Terpadu pada hakikatnya merupakan suatu pendekatan pembelajaran yang dapat membiasakan peserta didik secara individual ataupun kelompok dengan aktif mengeksplorasi, mengelaborasi, mengkonfirmasi, dan mengomunikasikan hasilnya.

Dalam kurikulum yang terintegrasi, guru sebaiknya secara eksplisit dapat mengasimilasi konsep lebih dari satu disiplin ilmu dalam proses pembelajaran IPA (Harrel, 2010, p.147). Selanjutnya, standar pembelajaran IPA menurut *National Science Education Standards* meliputi (1) pembelajaran berbasis inkuiri, (2) membimbing dan memfasilitasi belajar, (3) penilaian, (4) pengembangan lingkungan untuk pembelajaran, (5) membentuk komunitas belajar, (6) perencanaan dan pengembangan pembelajaran di sekolah (NSES, 1996, p.43). Standar proses pembelajaran IPA di Indonesia mengacu pada standar proses pembelajaran sesuai Permendikbud Nomor 65 Tahun 2013, Standar Proses dijabarkan sebagai suatu kriteria mengenai pelaksanaan pembelajaran pada satuan pendidikan untuk mencapai Standar Kompetensi Lulusan. Oleh sebab itu, pada proses pembelajaran setiap satuan pendidikan dituntut untuk mampu melakukan perencanaan pembelajaran dengan baik, sehingga pelaksanaan proses pembelajaran dapat berjalan semaksimal mungkin, serta penilaian proses pembelajaran bisa diarahkan untuk meningkatkan efisiensi dan efektivitas ketercapaian kompetensi lulusan.

Permasalahan yang ditemui bahwa guru belum paham hakikat IPA dan cara membelajarkannya dalam kelas. Guru IPA pada umumnya merancang pembelajaran IPA belum sesuai dengan hakikatnya yang berakibat pada kesalahan konsep pada anak, dan tidak memperhatikan keadaan psikologis anak dari pembukaan sampai evaluasi di akhir pembelajaran, dan pada gilirannya pembelajaran IPA menjadi kurang bermakna. Kebanyakan guru IPA kesulitan dalam memunculkan minat peserta didik terutama dalam memfasilitasi peserta didik untuk selalu merasa ingin tahu apa, bagaimana dan mengapa fenomena alam muncul. Guru kurang optimal dalam menerapkan metode pembelajaran, kesulitan dalam menentukan alat peraga yang sesuai dengan materi yang diajarkan, kesulitan menanamkan konsep yang benar pada peserta didik dan sering bersifat verbalistik.

Kesulitan lainnya adalah guru kurang memberikan motivasi pada awal pembelajaran untuk menghilangkan pola duduk, dengar, catat dan hapal sehingga peserta didik dan guru kurang berinteraksi yang menyebabkan peserta didik apatis, kesulitan memahami konsep materi IPA. Hal ini menyebabkan munculnya trauma peserta didik untuk belajar IPA dan pada akhirnya daya serap peserta didik dalam mengikuti pembelajaran masuk dalam ingatan jangka pendek dan menjadikan prestasi belajar IPA menurun.

Berdasarkan studi awal yang dilakukan bahwa masih ada beberapa Kompetensi Dasar (KD) untuk mata pelajaran IPA di Kota Gorontalo daya serapnya di bawah 60%. Bahkan ada yang 15,08% yaitu pada KD menjelaskan proses organ tubuh, ada juga yang 17,09% untuk menentukan aktivitas yang dipengaruhi oleh saraf tertentu dan yang lainnya adalah pada KD menentukan organ yang rusak akibat ketergantungan zat adiktif dan psikotropika berdasarkan kasus, mengidentifikasi peristiwa dalam kehidupan sehari-hari yang berkaitan dengan pemuian, memprediksi pengaruh kasus kepadatan populasi manusia, menjelaskan manfaat bioteknologi dalam kehidupan manusia, me-

mentukan besarnya gaya pada perpindahan benda berdasarkan energi yang dimilikinya, menentukan teknologi reproduksi yang tepat dalam meningkatkan kualitas dan kuantitas pada organisme tertentu, menghitung salah satu parameter dari Hukum Ohm pada suatu rangkaian tertutup bila besaran lainnya diketahui (Lukum, 2012, pp.149-150). Masalah lain yang menarik adalah sebanyak 36% guru IPA di Kota Gorontalo belum bersedia untuk disupervisi dalam melaksanakan proses pembelajaran (Lukum, 2012, p.156). Supervisi yang dilaksanakan oleh pengawas dan kepala sekolah sangat berpengaruh dalam menghasilkan guru IPA yang bermutu dalam melaksanakan pembelajaran di kelas. Dalam menghadapi perubahan kurikulum yang meliputi tujuan, isi, metode, dan evaluasi pengajarannya, sudah sewajarnya kalau para guru mengharapkan bantuan bimbingan dari pengawas dan kepala sekolah.

Daya serap tersebut menunjukkan bahwa program pembelajaran IPA bersifat penting untuk dievaluasi, karena fungsi pembelajaran IPA dapat merubah pola pikir peserta didik menjadi manusia yang kritis, kreatif dan berdaya cipta serta mengagumi kebesaran Tuhan Yang Maha Esa. Seyogyanya, optimalisasi pembelajaran IPA tersebut dapat dicapai apabila perencanaan dan pelaksanaan pembelajaran IPA dilakukan dan berjalan dengan maksimal.

Ada indikasi pembelajaran IPA di sekolah berjalan tidak sebagaimana mestinya. Hal ini disebabkan oleh faktor kesiapan guru dalam merancang dan melaksanakan pembelajaran, serta kurangnya dukungan kepala sekolah dalam menyelenggarakan program pembelajaran di sekolah.

Itulah sebabnya pembelajaran IPA ini perlu dipantau dan mendapatkan perhatian dari semua pihak yang terkait. Hal inilah yang menjadi dasar peneliti untuk melaksanakan evaluasi terhadap pembelajaran IPA SMP di Kabupaten Bone Bolango Provinsi Gorontalo.

Evaluasi pembelajaran IPA bertujuan untuk mendeskripsikan program pembelajaran IPA SMP di Kabupaten Bone Bola-

ngo. Tolok ukurnya adalah Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 65 Tahun 2013 tentang Standar Proses Pendidikan Dasar dan Menengah.

Evaluasi dapat didefinisikan sebagai penilaian pencapaian tujuan melalui pengumpulan dan analisis data yang berguna untuk membuat keputusan dari suatu program. Model evaluasi berguna dalam membimbing pengelolaan, pengumpulan data dan analisis (Wood, 2001, p.18). Evaluasi pembelajaran adalah kegiatan mengevaluasi hal-hal yang dilakukan dalam proses pembelajaran meliputi perencanaan, pelaksanaan dan proses penilaian serta dampaknya terhadap peserta didik. Evaluasi ini dilakukan dengan tujuan dapat memperbaiki kekurangan dalam pembelajaran, dan dapat dijadikan dasar untuk proses pembelajarannya selanjutnya.

Mengacu pada hal ini maka model evaluasi yang tepat dalam penelitian ini adalah model evaluasi *countenance stake*. Model *countenance stake* terdiri atas dua matriks. Matriks pertama dinamakan matriks deskripsi dan yang kedua dinamakan matriks pertimbangan. Matriks pertimbangan baru dapat dikerjakan oleh evaluator setelah matriks deskripsi diselesaikan (Arikunto, 2008, p.43). Evaluasi model Stake memberikan deskripsi dan pertimbangan sepenuhnya mengenai pembelajaran IPA yang dilakukan oleh guru IPA. Dalam model ini stake sangat menekankan peran evaluator dalam mengembangkan tujuan pembelajaran IPA menjadi tujuan khusus dan terukur.

Matriks deskripsi terdiri atas kategori tujuan (*intent*) dan observasi. Matriks pertimbangan terdiri atas kategori standar dan pertimbangan. Pada setiap kategori terdapat tiga fokus: (a) *antecedent* (konteks) yaitu sebuah kondisi yang ada sebelum instruksi yang berhubungan dengan hasil, (b) *transaction* (proses) yang merupakan proses instruksi kegiatan, dan (c) *outcomes* (hasil) yaitu efek dari pengalaman, pengamatan dan hasil kerja (Stake, 1967, pp.1-12). Matriks deskripsi berhubungan dengan kategori *intent* atau tujuan yang direncanakan dalam pengem-

banan program pembelajaran yang digunakan oleh guru IPA. Guru sebagai pengembang program merencanakan keadaan atau persyaratan yang diinginkan untuk suatu kegiatan pembelajaran di kelas. Kategori observasi berhubungan dengan apa yang terjadi sesungguhnya dalam program pembelajaran sebagai implementasi yang diinginkan pada *intent* atau tujuan. Kategori standar dalam matriks pertimbangan adalah kriteria yang harus dipenuhi oleh suatu program pembelajaran yang dijadikan fokus evaluasi. Kategori pertimbangan, dalam hal ini evaluator melakukan pertimbangan dari apa yang telah dievaluasi pada matriks deskripsi, selanjutnya dilakukan pemberian pertimbangan. Keseluruhan matriks yang mendukung model Stake ini terdiri dari 12 kotak.

Worthen & Sanders, 1987, dalam Wood, (2001, p.19) mengemukakan bahwa Stake menciptakan kerangka evaluasi untuk membantu evaluator dalam mengumpulkan, mengatur, dan menafsirkan data kuantitatif dan kualitatif. Inti dari kegiatan evaluasi adalah proses dihasilkannya informasi sebagai alternatif keputusan (Stufflebeam, 1973, pp.3-5; Hamm, 1985, pp.256-622; Stake, 1967, pp.2-4; Stufflebeam & Shinkfield, 1985, p.159). Tahapan evaluasi Stake yang relevan adalah *Input (Antecedent)*, *Proses (Transaction)*, dan *Produk (Outcomes)*, (Provus, 1969, pp.2,3; Kaufman & Thomas, 1980, p.137). Gambaran *Outcome* model Stake adalah dampak dari pelaksanaan program pembelajaran.

Hal yang menarik pada evaluasi ini terletak pada perbedaan antara deskripsi tindakan dan keputusan yang sesuai dengan program pendidikan pada *antecedent*, *transaction* dan *outcomes* (Popham 1993, dalam Wood, 2001, p.19). Berdasarkan hal tersebut, keuntungan evaluasi model *countenance stake* adalah penilaiannya didasarkan atas kebutuhan program yang dievaluasi, sehingga dapat mengdeskripsikan secara kompleks program pembelajaran yang dilaksanakan oleh guru serta memiliki potensi besar untuk memperoleh pengalaman dan teori terhadap program pembelajaran yang dievaluasi.

Antecedent adalah sebuah kondisi yang ada sebelum pembelajaran berlangsung yang berhubungan dengan hasil, sedangkan *transaction* merupakan proses pengalaman pembelajaran. Ada dua cara yang dapat digunakan untuk memproses data evaluasi deskriptif pada program pendidikan yaitu menemukan *contingency* (keterhubungan) antara anteseden, transaksi, dan *outcome* serta *congruence* (kesesuaian) antara tujuan yang diharapkan dan kondisi yang diobservasi (Stake, 1977, pp. 372-390). Perlu diperhatikan apakah rencana pelaksanaan pembelajaran sesuai dengan pelaksanaan pembelajaran di kelas atau terjadi penyimpangan.

Data kurikulum dikatakan *congruence* apabila tujuan yang diharapkan secara nyata terpenuhi. Dalam satu garis matriks data, evaluator dapat membandingkan sel-sel yang berisi tujuan dan observasi, untuk mencatat kesenjangan dan menjelaskan sebuah kongruens dengan garis tersebut. Hubungan atau *contingencies* antara variabel-variabel perlu mendapat perhatian. Dalam evaluasi program pendidikan yang mencari hubungan yang mungkin untuk peningkatan program pendidikan, tugas evaluator adalah tergantung pada hasil identifikasi kondisi yang terjadi pada anteseden dan kondisi yang terjadi pada transaksi. Observasi *contingencies* ini tergantung pada bukti empiris (Wood, 2001, pp.19-20). Analisis *contingencies* meliputi analisis logis dan analisis empiris.

Analisis logis dilakukan untuk pemberian pertimbangan mengenai hubungan antara *antecedent*, *transaction*, dan hasil dari kotak-kotak tujuan. Evaluator harus dapat menentukan apakah *antecedent* yang telah dikemukakan dalam pengembangan program pembelajaran akan tercapai dengan rencana transaksi yang diajukan. Atau sebetulnya ada model transaksi lain yang lebih efektif. Demikian pula mengenai hubungan antara transaksi dengan hasil yang diharapkan. Analisis kedua adalah analisis empirik. Dasar bekerjanya sama dengan analisis logis tetapi data yang digunakan adalah data empirik. Hamm (1985, pp. 256-622) menyatakan bahwa Model Stake berguna untuk memberikan bukti yang luas untuk mendu-

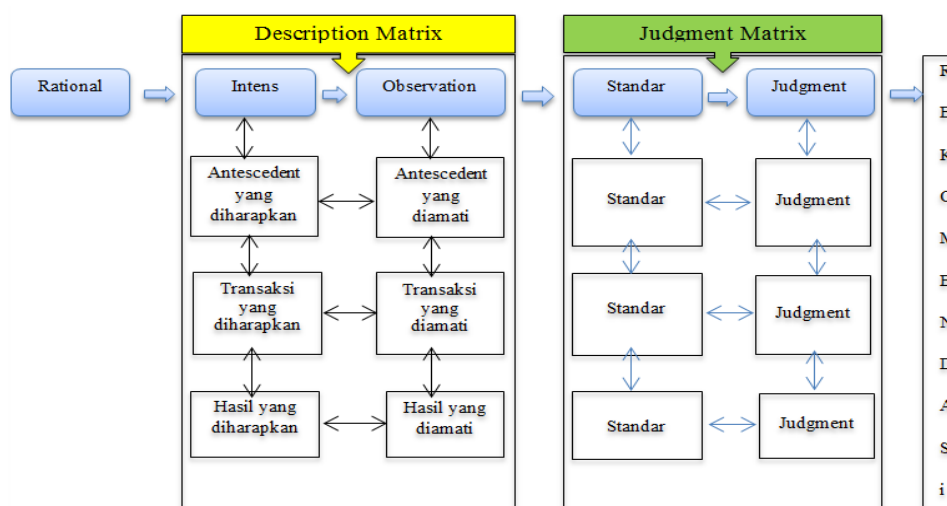
kung suatu keberhasilan program tertentu dengan cara mendokumentasikan hubungan yang masuk akal antara semua komponen.

Berdasarkan uraian tersebut maka yang dimaksud dengan *antecedents* dalam program pembelajaran IPA adalah perencanaan pembelajaran dalam hal ini rencana pelaksanaan pembelajaran (RPP) yang dibuat guru. *Transaction* atau proses adalah pelaksanaan pembelajaran yang dilaksanakan sesuai dengan rencana pembelajaran yang ditetapkan. *Outcomes* atau produk yaitu hasil pelaksanaan pembelajaran IPA berupa hasil belajar peserta didik.

Berdasarkan uraian yang telah disampaikan, penelitian ini bertujuan untuk mengevaluasi program pembelajaran IPA SMP di Kabupaten Bone Bolango. Tolok ukurnya berdasarkan Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 65 Tahun 2013 tentang Standar Proses Pendidikan Dasar dan Menengah.

Metode Penelitian

Penelitian ini menggunakan metode deskriptif dengan model evaluasi yang digunakan adalah Coutenance Stake. Model evaluasi Stake merupakan analisis proses evaluasi yang menekankan pada dua jenis operasi yaitu deskripsi (*descriptions*) dan pertimbangan (*judgments*) serta membedakan tiga fase dalam evaluasi program yaitu: (1) persiapan (*antecedents*) dalam penelitian ini adalah perencanaan pembelajaran; (2) transaksi adalah pelaksanaan pembelajaran; dan (3) *outcome* dari program ini yakni hasil belajar peserta didik. Matriks deskripsi berhubungan dengan *intens* program pembelajaran IPA dan hasil *observations* dari program ini di sekolah. Matriks *judgement* berhubungan dengan standar atau kriteria dalam hal ini adalah Permendikbud No 65 Tahun 2013 tentang Standar Proses Pembelajaran dan *judgement* (pertimbangan) evaluator. Penekanan paling besar pada model ini adalah pendapat bahwa evaluator membuat keputusan tentang program yang sedang dievaluasi. Desain penelitian ini menggunakan model evaluasi *coutenance* yang dikembangkan Stake seperti pada Gambar 1.

Gambar 1. Desain Penelitian dengan model *Countenance Stake*

Alur evaluasi model *Countenance Stake* terdiri dari empat langkah, yaitu langkah awal, mengumpulkan data, analisis logis, dan analisis empiris. Setiap langkah dijelaskan sebagai berikut.

Langkah awal yang dilakukan adalah menyusun rasional dari program pembelajaran IPA SMP di Kabupaten Bone Bolango. Pada bagian ini dikumpulkan data awal tentang program yang telah dilaksanakan oleh guru yang meliputi perencanaan, pelaksanaan, dan hasil belajar peserta didik berdasarkan kajian teoretis, dukungan peraturan yang berlaku, serta kondisi nyata sekolah.

Tahap pengumpulan data mengenai *intents* ini dilakukan untuk memperoleh informasi tentang tujuan dari program pembelajaran IPA SMP dan efek yang diharapkan dari program tersebut. Analisis tujuan ini dilakukan pula pada tiga bagian komponen evaluasi yaitu *antecedent* berupa RPP, *transaction* (proses) yakni pelaksanaan, dan juga hasil belajar IPA sebagai *outcomes* dalam program ini. Analisis dilakukan dengan memperhatikan kondisi objektif program tersebut kemudian dilakukan pengolahan data matriks deskripsi, dengan dua konsep yaitu *contingency* dan *congruence*. Kedua konsep ini berbeda dalam penggunaannya. *Contingency* dipergunakan untuk menganalisis data secara vertikal, mencari keterhubungan/keselarasan antara *antecedent*, *transaksi*,

dan juga *outcome*. Analisis *Contingency* ini dilakukan dengan dua cara yaitu keterhubungan secara logika dan keterhubungan secara empirik.

Analisis logis terhadap data dalam penelitian ini digunakan untuk memberikan pertimbangan mengenai keterhubungan antara *antecedent* (RPP), transaksi (pelaksanaan pembelajaran), dan hasil belajar IPA yang ada di dalam matrik *intents*. Hasil analisis ini menemukan apakah RPP yang dibuat guru IPA sebagai persyaratan awal dalam program pembelajaran IPA akan tercapai dengan rencana transaksi yang dikemukakan. Demikian pula mengenai hubungan antara pelaksanaan pembelajaran dengan hasil belajar IPA yang diharapkan.

Analisis empiris dilakukan untuk mempertimbangkan keterhubungan antara *antecedent* (RPP), *transaksi* (pelaksanaan pembelajaran), dan juga hasil belajar. Analisis ini berdasarkan data empirik yang diperoleh di lapangan.

Selain mencari kontingensi peneliti kemudian memberikan pertimbangan mengenai *congruence* atau perbedaan yang terjadi antara apa yang direncanakan dengan apa yang terjadi di lapangan. Analisis *congruence* dilakukan terlebih dahulu dengan cara menyusun standar pengukuran keterlaksanaan program pada semua tahap evaluasi dengan menyusun kriteria-kriteria yang jelas dan terukur. Standar yang akan digunakan ber-

dasarkan pertimbangan teoretis dan pertimbangan praktis pada kondisi lapangan penelitian. Analisis terhadap kesesuaian standar dengan data hasil penelitian akan menjadi dasar dalam pengambilan keputusan (*judgment*). Pengambilan keputusan ini dilaksanakan untuk ketiga komponen evaluasi yaitu *antecedent*, *transaksi*, dan juga *outcome*. Langkah terakhir dari penelitian ini adalah memberikan rekomendasi dan pertimbangan berdasarkan hasil evaluasi.

Teknik pengumpulan data dalam penelitian ini menggunakan teknik observasi, dokumentasi dan wawancara. Observasi digunakan pada pelaksanaan pembelajaran di kelas meliputi kegiatan awal, kegiatan inti dan kegiatan penutup dengan menggunakan lembar observasi sesuai dengan standar proses. Observasi dilakukan oleh observer yakni peneliti dan kepala sekolah. Hal ini dilakukan agar kredibilitas observer dapat dipercaya, dengan mempertimbangkan kepala sekolah lebih mengetahui kondisi sekolah. Dokumentasi digunakan untuk melakukan penilaian RPP yang dibuat guru IPA. Penilaian RPP dilakukan dengan cara memberi skor sesuai kuantifikasi ketersediaan RPP yang dimiliki guru dengan menggunakan instrumen penilaian sesesuai dengan standar proses. Di samping itu studi dokumen digunakan untuk data nilai harian, ujian tengah semester dan ujian akhir semester. Instrumen yang digunakan adalah instrumen yang tercantum dalam standar proses pembelajaran Permendikbud No. 65 Tahun 2013 yang terdiri dari standar perencanaan, pelaksanaan, dan penilaian hasil pembelajaran. Wawancara dilakukan kepada guru dan kepala sekolah untuk memperoleh data kesiapan guru dalam merencanakan dan melaksanakan pembelajaran serta upaya kepala sekolah dalam melaksanakan supervisinya pada pembelajaran IPA.

Analisis data dalam penelitian ini dilaksanakan secara deskriptif kualitatif. Data kualitatif dianalisis menggunakan analisis tematik yaitu membandingkan data pada tiga tahapan Stake yaitu: *antecedent*, *transaction* dan *outcomes* pada matriks deskripsi dengan standar yang ada pada matriks pertimbangan,

kemudian disimpulkan. Dalam analisis tematik ini ditempuh alur analisis yang terdiri dari pengumpulan data, reduksi data dan kesimpulan verifikasi.

Data penilaian RPP dan pelaksanaan pembelajaran dianalisis secara statistik deskriptif dengan menggunakan rumus persentase sebagai berikut.

$$\text{Nilai perolehan} = \frac{\text{Skor Mentah}}{\text{Skor Maksimum}} \times 100$$

Nilai perhitungan persentase (%) kemudian dikonversikan dalam bentuk kualitatif untuk menentukan aktualitas ketercapaian. Rentang nilai disajikan pada Tabel 1.

Tabel 1. Konversi Hasil Perhitungan Ketercapaian Program

No	Skala Persentase	Kategori Nilai	Predikat Hasil Evaluasi
1	$85 \leq \text{NA} \leq 100$	A	Amat Baik
2	$70 \leq \text{NA} < 85$	B	Baik
3	$56 \leq \text{NA} < 70$	C	Cukup
4	$\text{NA} < 56$	D	Kurang

(Arikunto, 2008, Hal. 8)

Persentase capaian skor menunjukkan seberapa besar keterlaksanaan program pembelajaran IPA di Kabupaten Bone Bolango dapat dicapai. Besarnya persentase capaian skor selanjutnya digunakan untuk mendeskripsikan kesesuaian antara standar proses pembelajaran yang telah ditetapkan dengan hasil yang ditemukan di lapangan.

Hasil Penelitian dan Pembahasan

Hasil penelitian untuk setiap tahapan evaluasi disajikan pada matriks *Countenance Stake* pada tabel yang meliputi intens, observasi, standar dan judgment untuk masing-masing 3 komponen program yang dikelompokkan dalam tabel menurut *antecedent*, *transaction*, dan *outcomes*. Selanjutnya hal itu dianalisis *congruence* dan *contingency*.

Congruence

Komponen Antecedent

Komponen yang dievaluasi pada *antecedent* ini adalah RPP yang dibuat guru IPA

pada empat sekolah yang menjadi sasaran Kurikulum 2013.

Berikut ini disajikan *countenance matrix* komponen *antecedent* pada Tabel 2.

Tabel 2. *Countenance Matrix* Komponen Antecedent

<i>Description Matrix</i>		<i>Judgment Matrix</i>	
Intens	Observasi	Standar	<i>Judgments</i>
RPP			
RPP yang dibuat guru IPA sesuai dengan standar proses permendikbud No 65 tahun 2013.	Aktualitas ketercapaian RPP yang dibuat guru IPA sebanyak 68% kategori cukup. Belum semua guru IPA merencanakan pembelajaran sesuai dengan kriteria yang telah ditetapkan dalam standar proses permendikbud No 65 tahun 2013.	Komponen RPP berdasarkan Standar Proses Pembelajaran meliputi: 1) Identitas mata pelajaran; 2) materi pokok; 3) alokasi waktu; 4) tujuan pembelajaran yang dirumuskan berdasarkan KD; 5) kompetensi dasar dan indikator pencapaian kompetensi; 6) materi pembelajaran, memuat fakta, konsep, prinsip, dan prosedur yang relevan; 7) metode pembelajaran yang digunakan pendidik guna mencapai KD yang disesuaikan dengan karakteristik peserta didik; 8) media pembelajaran yang membantu proses menyampaikan materi pelajaran; 9) sumber belajar dapat berupa buku, media cetak dan elektronik, alam sekitar, atau sumber belajar lain yang relevan; 10) langkah-langkah pembelajaran dilakukan melalui tahapan pendahuluan, inti, dan penutup; 11) penilaian hasil pembelajaran	Sebagian RPP yang dibuat guru IPA belum sesuai dengan Permendiknas No 65 tahun 2013 tentang standar proses.

Berdasarkan Tabel 2, RPP yang dibuat guru IPA di sekolah-sekolah sasaran Kurikulum 2013 termasuk dalam kategori cukup (68%). Kesesuaian intens dengan observasi, pada matriks deskripsi ditemukan belum adanya kesesuaian antara ketersediaan RPP yang dibuat guru IPA dengan Standar Proses Pembelajaran, terutama dalam komponen pemilihan sumber belajar pada indikator kesesuaian dengan pendekatan saintifik, dan karakteristik peserta didik; komponen pemilihan media belajar pada indikator kesesuaian dengan pendekatan saintifik dan menggunakan metode pembelajaran yang bervariasi sehingga membuat peserta didik aktif belajar; komponen metode pembelajaran dan skenario pembelajaran pada indikator kegiatan pembelajaran dirancang membuat peserta didik aktif belajar, sedangkan pada komponen penutup yaitu dalam indikator membuat rangkuman/simpulan pelajaran, melakukan penilaian dan/atau refleksi terhadap kegiatan yang sudah dilaksanakan, memberikan umpan

balik terhadap proses dan hasil belajar, dan merencanakan kegiatan tindak lanjut (remedi, pengayaan, konseling, dan/atau tugas) dan menyampaikan rencana pembelajaran pada pertemuan berikutnya.

Beberapa faktor yang mempengaruhi rendahnya capaian tersebut meliputi latar belakang guru IPA. Mereka tidak memiliki kualifikasi bidang pendidikan IPA melainkan dari Kimia, Biologi, dan Fisika, sehingga dalam merencanakan pembelajaran IPA terpadu mengalami kesulitan. Hal ini yang menyebabkan guru kurang kreatif dalam memilih sumber dan media belajar yang berkesesuaian dengan pendekatan saintifik sehingga pembelajaran dirancang tidak dapat membuat peserta didik aktif bahkan peserta didik enggan untuk mengemukakan pertanyaan yang kritis. Hal ini diperkuat oleh Winaryati. E., Suyata, & Sumarno, (2013, p.248) mengemukakan bahwa belum adanya kesesuaian antara RPP dengan kesiapan guru mengajar, kesiapan peserta didik dan kesiapan peralatan yang ada di SMP Negeri.

Dengan demikian, guru IPA diberi kesempatan untuk bekerja sama dengan guru utama untuk mengawali pengalamannya dari merancang pembelajaran, praktik mengajar dan merefleksikannya, serta belajar bagaimana membelajarkan peserta didik dengan beragam kepentingan dan pengalaman dalam memahami ide-ide ilmiah.

Komponen Transaction

Komponen yang dievaluasi pada *Transaction* ini adalah kegiatan pelaksanaan pembelajaran IPA pada empat sekolah yang menjadi sasaran Kurikulum 2013. Berikut ini disajikan *countenance matrix* komponen *transaction* pada Tabel 3.

Tabel 3. Countenance Matrix Komponen Transaction

<i>Description Matrix</i>		<i>Judgment Matrix</i>	
Intens	Observasi	Standar	Judgments
Pelaksanaan pembelajaran			
Guru IPA dapat melaksanakan pembelajaran sesuai dengan Standar Proses Permendikbud No 65 Tahun 2013	Aktualitas ketercapaian pelaksanaan pembelajaran IPA adalah 57% kategori cukup. Belum semua guru IPA melaksanakan pembelajaran sesuai dengan kriteria yang telah ditetapkan dalam standar proses permendiknas No 65 tahun 2013.	Pelaksanaan pembelajaran berdasarkan standar proses meliputi: (1) pendahuluan: a) menyiapkan peserta didik secara psikis dan fisik untuk mengikuti proses pembelajaran; b) memberi motivasi belajar peserta didik secara kontekstual sesuai manfaat dalam kehidupan sehari-hari; c) mengajukan pertanyaan yang mengaitkan pengetahuan sebelumnya dengan materi yang akan dipelajari; d) menjelaskan tujuan pembelajaran atau kompetensi dasar yang akan dicapai; dan e) menyampaikan cakupan materi dan penjelasan uraian kegiatan sesuai silabus. (2) Kegiatan Inti: menggunakan model pembelajaran, metode pembelajaran, media pembelajaran, dan sumber belajar yang disesuaikan dengan karakteristik peserta didik dan mata pelajaran. Pemilihan pendekatan tematik dan/atau tematik terpadu dan/atau saintifik dan/atau inkuiri, dan penyingkapan (<i>discovery</i>) dan/atau pembelajaran yang menghasilkan karya berbasis pemecahan masalah (<i>project based learning</i>) disesuaikan dengan karakteristik kompetensi dan jenjang pendidikan. Kegiatan Penutup Dalam kegiatan penutup, guru bersama peserta didik baik secara individual maupun kelompok melakukan refleksi untuk mengevaluasi: a) seluruh rangkaian aktivitas pembelajaran dan hasil-hasil yang diperoleh untuk selanjutnya secara bersama menemukan manfaat langsung maupun tidak langsung dari hasil pembelajaran yang telah berlangsung; b) memberikan umpan balik terhadap proses dan hasil pembelajaran; c) melakukan kegiatan tindak lanjut dalam bentuk pemberian tugas, baik tugas individual maupun kelompok; dan d) menginformasikan rencana kegiatan pembelajaran untuk pertemuan berikutnya.	Keterlaksanaan pembelajaran IPA SMP belum sepenuhnya sesuai dengan standar proses. Guru IPA masih perlu meningkatkan profesionalismenya melalui kegiatan kelompok kerja guru (KKG) dan bimbingan dari pengawas sekolah.

Tabel 3 menjelaskan bahwa aktualitas ketercapaian pelaksanaan pembelajaran termasuk dalam kategori cukup (57%). Dari tabel tersebut juga ditemukan belum adanya kesesuaian antara pelaksanaan pembelajaran yang ada di sekolah dengan standar proses pelaksanaan pembelajaran yang ada pada tujuan. Ketidaksesuaian ini terdapat pada komponen pendahuluan yakni dalam indikator mengajukan pertanyaan yang menantang, menyampaikan manfaat materi pembelajaran, mendemonstrasikan sesuatu yang terkait dengan tema, dan mengecek perilaku awal (*entry behavior*). Demikian halnya dalam kegiatan inti pembelajaran belum semua guru IPA melaksanakan pembelajaran kontekstual yang memungkinkan tumbuhnya kebiasaan positif, sebagai dampak pengiring hasil pembelajaran (*nurturant effect* atau suasana kondusif yang tercipta dengan sendirinya (*hidden curriculum*)).

Selain itu, ditemukan pula kesulitan guru IPA dalam menerapkan pendekatan pembelajaran saintifik (mengamati, menanya, menalar, mencoba, mengkomunikasikan). Temuan lainnya pada indikator kemampuan mengkaitkan materi dengan pengetahuan lain yang relevan, perkembangan IPTEK, dan kehidupan nyata, mengelola pembahasan materi pembelajaran dan pengalaman belajar dengan tepat, memfasilitasi kegiatan yang memuat komponen eksplorasi, elaborasi dan konfirmasi serta memberikan pertanyaan peserta didik untuk menalar (proses berpikir yang logis dan sistematis). Pada kegiatan penutup ditemukan ketidaksesuaian pada indikator melakukan refleksi atau membuat rangkuman dengan melibatkan peserta didik, mengumpulkan hasil kerja sebagai bahan portofolio, serta melaksanakan tindak lanjut dengan memberikan arahan kegiatan berikutnya dan tugas pengayaan.

Ketidaksesuaian pelaksanaan pembelajaran IPA dengan standar proses berpangkal dari RPP, guru serta faktor-faktor pendukung pembelajaran seperti media dan metode mengajar. RPP yang dibuat guru

IPA belum optimal, karena keterbatasan kemampuan guru dalam melakukan pembaruan metode dan strategi pembelajaran, serta guru sulit dalam mengelola waktu. Hal ini berdampak pada cara mengajar guru yang selalu hanya membentuk budaya menghafal dibanding dengan membentuk pola berpikir kritis anak. Padahal pelaksanaan pembelajaran IPA dapat membantu peserta didik mengembangkan pemahaman konseptual dan kemampuan untuk menyelidiki (membuat pertanyaan, menjawab pertanyaan-pertanyaan ilmiah), dapat berkomunikasi dan membenarkan temuan, produknya diperlukan untuk membangun warga negara yang produktif (Davis, 2008, pp.1-8).

Penelitian Raharja & Retnowati, (2013, p.297) melaporkan bahwa beberapa kemampuan guru yang masih kurang dalam pelaksanaan pembelajaran yakni menggunakan beragam pendekatan, media dan sumber belajar; kurangnya kemampuan guru dalam memfasilitasi interaksi antar peserta didik serta belum maksimal dalam melibatkan peserta didik secara aktif dalam kegiatan pembelajaran termasuk, dalam melibatkan peserta didik mencari informasi luas tentang materi atau topik yang diajarkan. Dengan demikian pelaksanaan pembelajaran di kelas menghendaki guru IPA yang paham dalam merancang pembelajaran yang inovatif dan memiliki kemampuan dalam mengimplementasikan RPP di kelas.

Komponen Outcome

Komponen yang dievaluasi pada *outcome* ini adalah hasil belajar IPA pada empat sekolah yang menjadi sasaran Kurikulum 2013. Berikut ini disajikan *countenance* matriks komponen *outcome* pada Tabel 4. Tabel 4 menjelaskan bahwa belum ada kesesuaian antara hasil belajar IPA dengan Kriteria Ketuntasan Minimal (KKM). Hal ini ditunjukkan oleh aktualitas ketercapaian hasil belajar peserta didik 65% atau pada kategori cukup. Faktor penyebab ketidaksesuaian ini adalah pada penentuan proses penilaian.

Tabel 4. *Countenance Matrix* Komponen *Outcomes*

<i>Description Matrix</i>		<i>Judgment Matrix</i>	
Intens	Observasi	Standar	<i>Judgments</i>
Hasil Belajar Peserta didik			
Hasil belajar IPA memenuhi Kriteria Ketuntasan Minimal (KKM) sebesar 75%.	Aktualitas ketercapaian hasil belajar IPA adalah 65% kategori cukup. Ditemukan masih ada peserta didik yang tidak tuntas pada ulangan harian, tugas, ujian tengah semester dan ujian akhir semester.	Nilai peserta didik pada ulangan harian, tugas, ujian tengah semester, ujian akhir semester dan buku rapor memenuhi KKM.	Hasil belajar peserta didik belum memenuhi KKM. Guru IPA seyogyanya menggunakan pendekatan penilaian otentik (<i>authentic assesment</i>) yang menilai kesiapan peserta didik, proses, dan hasil belajarsecara utuh.

Belum semua guru menguasai penilaian otentik dan tidak tersedianya dokumen penilaian. Penilaian pembelajaran IPA yang dilakukan oleh guru belum memadukan sikap, pengetahuan, dan keterampilan melalui proses penilaian otentik.

Hal yang sama ditemukan oleh Winaryati. E., Suyata, & Sumarno (2013, p.248) bahwa belum adanya kesesuaian ketersediaan dokumen pendukung pembelajaran dengan realisasi pelaksanaannya, mencakup: jurnal guru, modul/bahan ajar, LKS, bank soal. Terkait dengan kreativitas peserta didik ditemukan belum adanya kesesuaian antara perencanaan dengan proses pembelajaran terkait dengan kreativitas peserta didik. Temuan ini sejalan dengan penelitian Raharja, dan Retnowati (2013, p.302) bahwa ada kesenjangan antara pelaksanaan komponen perencanaan, pelaksanaan, penilaian, dan pengawaan pembelajaran dengan standar minimal pelaksanaan proses pembelajaran sebagaimana termuat dalam Permendiknas No 41 Tahun 2007.

Contingency

Keterhubungan antara *antecedent* dengan *transaction*, *transaction* dengan *outcome* dan *antecedent*, *transaction* dan *outcomes*, baik pada intens dan *observation*, semua hasil evaluasi dalam kategori cukup. Hal ini sesuai dengan fenomena hasil observasi bahwa masih ada sebagian guru IPA yang belum memahami cara menyusun RPP yang baik serta melaksanakan pembelajaran yang se-

suai dengan RPP yang dibuat. Sebagian guru IPA masih melakukan “*copy paste*” RPP dan ini berdampak pada tidak berhasilnya pembelajaran yang dilakukan guru. Hal ini memberikan gambaran bahwa terdapat *contingency* antara perencanaan, pelaksanaan dengan hasil belajar IPA.

RPP yang dibuat guru menggambarkan kemampuan guru IPA dalam merencanakan pembelajaran termasuk dalam kategori cukup. Hal ini dipengaruhi oleh sebagian guru tidak paham menyusun RPP sehingga rancangan pembelajaran sulit untuk diimplementasikan di kelas, Hal ini berdampak pada pelaksanaan pembelajaran di kelas belum optimal.

Pelaksanaan pembelajaran menggambarkan kemampuan guru dalam mengelola pembelajaran belum sesuai dengan standar proses. Faktor ini dipengaruhi oleh guru kesulitan dalam merencanakan pembelajaran terutama dalam indikator menyebabkan peserta didik aktif dan peserta didik mampu mengajukan pertanyaan yang menantang dalam kelas. Akibatnya, semua yang direncanakan sulit untuk diimplementasikan di kelas. Jika guru dapat menyusun RPP yang baik maka pelaksanaan pembelajaran di kelas baik pula sehingga berdampak pada hasil belajar peserta didik yang baik.

Hasil belajar IPA dalam kategori cukup menggambarkan RPP dan pelaksanaan pembelajaran IPA belum sepenuhnya sesuai dengan standar proses pembelajaran. Akibatnya, ditemukan masih ada peserta didik

yang tidak tuntas dalam ulangan harian, ujian tengah semester dan ujian akhir semester.

Hal ini sejalan dengan penelitian Winaryati, E., Suyata, & Sumarno (2013, p.4248) yang menemukan bahwa dimensi yang memiliki nilai kurang baik yakni faktor kesiapan realisasi RPP pada pembelajaran yang berbasis 5 Domain Sains. Hal ini berarti belum adanya kesesuaian antara RPP dengan kesiapan guru mengajar, kesiapan siswa dan kesiapan peralatan yang ada.

Simpulan

Simpulan

Simpulan penelitian ini sebagai berikut: (1) perencanaan pembelajaran IPA termasuk dalam kategori cukup (68%), kesesuaian intens dengan observasi, pada matriks deskripsi ditemukan belum adanya kesesuaian antara RPP yang dibuat guru IPA dengan Permendikbud No 65 Tahun 2013 tentang Standar Proses Pembelajaran; (2) pelaksanaan pembelajaran IPA termasuk dalam kategori cukup (57%), ditemukan belum adanya kesesuaian antara pelaksanaan pembelajaran yang ada di sekolah dengan standar proses pelaksanaan pembelajaran; (3) hasil belajar peserta didik belum memenuhi Kriteria Ketuntasan Minimal (KKM) dengan aktualitas ketercapaian 65% kategori cukup; dan (4) terdapat *contingency* antara perencanaan, pelaksanaan dengan hasil belajar IPA, perencanaan pembelajaran dengan kategori yang cukup menyebabkan guru melaksanakan pembelajaran belum sesuai standar proses sehingga hasil belajar peserta didik belum memenuhi KKM.

Saran

Berdasarkan simpulan penelitian maka disarankan hal-hal sebagai berikut: (1) hendaknya guru IPA meningkatkan kemampuannya dalam strategi pembelajaran yang meliputi: penggunaan media, metode pembelajaran dan pendekatan pembelajaran; (2) pelaksanaan pembelajaran hendaknya dirancang sehingga dapat mendorong pelaksanaan pembelajaran yang aktif, efektif dan menyenangkan; 3) hendaknya guru IPA

menggunakan penilaian otentik (*authentic assesment*) dalam pembelajaran.

Rekomendasi

Berdasarkan analisis *congruences* dan *contingency*, direkomendasikan sebagai berikut.

Pertama, perencanaan pembelajaran yang baik perlu mempertimbangkan pencapaian tujuan melalui penggunaan strategi pembelajaran yang sesuai dengan perkembangan peserta didik. Strategi pembelajaran meliputi: media, metode pembelajaran dan pendekatan pembelajaran.

Kedua, pelaksanaan pembelajaran hendaknya mengacu pada RPP yang dibuat guru. RPP yang berkualitas menggunakan strategi pembelajaran aktif, efektif dan menyenangkan, yang akan menimbulkan pertanyaan-pertanyaan yang menantang dari peserta didik.

Ketiga, hasil belajar IPA seharusnya diperoleh peserta didik melalui proses pembelajaran yang menggunakan penilaian otentik. Penilaian ini memadukan penilaian sikap, pengetahuan, dan keterampilan secara utuh.

Daftar Pustaka

- Arikunto, S. (2008). *Evaluasi program pendidikan* (2nded). Jakarta: PT. Bumi Aksara.
- Depdikbud. (2013). *Materi pelatihan guru implementasi kurikulum 2013 SMP, MTs Ilmu Alam* (pp.1-366). Jakarta: Badan Pengembangan Sumber Daya Manusia Pendidikan dan Kebudayaan Penjaminan Mutu Pendidikan, Ke-mendikbud.
- Hamm, R.W. (1985). *A Systematic evaluation of an environmental investigation course*. (Doctoral dissertation. Georgia State University). ERIC Document. Reproduction Service No ED-256-622.
- Harrell, P.E. (2010). Teaching an Integrated Science Curriculum: Linking Teacher Knowledge and Teaching Assignments. *Issues in Teacher Education*. Vol. 19, No. 1, pp. 145-165.

- Kaufman, R., & Thomas, S. (1980). *Evaluation without fear*. New York London: New Viewpoints
- Kementerian Pendidikan dan Kebudayaan. (2013). *Peraturan menteri pendidikan dan kebudayaan no. 65 tahun 2013 tentang Standar Proses*. Jakarta: BSNP.
- Lukum, A. (2012). *Evaluasi program supervisi pembelajaran IPA*. (Disertasi Doktor. Universitas Negeri Jakarta).
- Lukum, A. (2013). "Evaluation of Science Learning Supervision on Secondary School", *International Journal of Education* Vol. 5, no:74, pp. 61-81.
- McBride, J.W., Bhatti, M.I., Hannan M.A., & Fein Berg, M. (2004). *Using an inquiry approach to teach science to secondary school science teachers (pp.1-6)*. Physic Education 39 (5), IOP Publishing Ltd.
- National Science Teachers Association (NSTA). (2004). *Scientific inquiry*. <http://www.nsta.org/docs/PositionStatementScientificInquiry.pdf>. (pp.1-3). Diakses tanggal 08 April 2014.
- National Science Education Standards. (1996). *Science study and teaching standards*. National Committee on Science Education Standards and Assessment, National Research Council. (<http://www.nap.edu/catalog/4962.html>). (pp.1-262). Diakses 07 April 2014.
- Paramata, Y. (2001). *Pengembangan model sosialisasi inovasi dan supervisi pembelajaran ilmu pengetahuan alam*. (Disertasi Doktor. Universitas Pendidikan Indonesia). Hal 2.
- Probus, M., Malcolm. (1969). *The discrepancy evaluation models. An approach to local program improvement and development*. Pitaburgh Public School.
- Raharja, J. T., & Retnowati, T. H. (2013). Evaluasi Pelaksanaan Pembelajaran Seni Budaya SMA di Kabupaten Lombok Timur, NTB. *Jurnal Penelitian dan Evaluasi Pendidikan*. Vol. 17, No. 2, (pp.287-258).
- Rustaman, N.Y. (2010). Kemampuan Dasar Bekerja Ilmiah dalam Pendidikan Sains dan Asessmentnya. *Makalah Universitas Indonesia*. http://file.upi.edu/direktori/sps/prodi.pendidikanipa/195012311979032_nuryani_rustaman/kdbi_dalamdiksainsfinal.pdf (diakses 08 April 2014)
- Scriven, M. (1967). *The methodology of evaluation*. In R.W.Tyler. R M. Gagne, & M Scriven (Eds). *Perspectrives of curriculum evaluation*. (pp.39-83). Chicago: Rand McNally.
- Stake, R E. (1967). *Forward technology for the evaluation of educational programs*. In R W Tyler, R M Gagne, & M Scriven. (Eds). *Perspectives of curriculum evaluation*. (pp.1-12). Chicago: Rand McNally.
- Stake, R E. (1967). The countenance of educational evaluation. *Teacher's Coolege Record*. Vol. 68, no:7.
- Stake, R E. (1977). *The Countenance of educational evaluation*. In A.A. Bellack & H.M Kliebard. Eds 1. *Curriculum and evaluation* (pp. 372-390). Berkeley.CA McCutchan.
- Stufflebeam, D. L., & Shinkfield, A.J. (1984). *Systematic evaluation a self-instructional guide to theory and practice*. Boston: Kluwer-Nijhoff Publishing.
- Stufflebeam, D. L. (1973). Educational evaluation: theory and practice. In B.R. Worthen & J.R. Sanders (Eds), *Evaluation as Enligbtenment for Decision Marking* (pp.3-5). Boston: Kluwer-Nijhoff Publishing.
- Winaryati. E., Suyata, & Sumarno. (2013). Model evaluasi dalam supervisi pembelajaran IPA berbasis lima domain sains. *Jurnal Penelitian dan Evaluasi Pendidikan*. Vol. 17, No. 2, (pp.241-258).
- Wood, B.B. (2001). Stake's Countenance Model: Evaluating an Enveronmental Education Proffesional Development Course. *The Journal of Environmental Education*, Vol.32, No.2, pp. 18-27.

MENENTUKAN KRITERIA KETUNTASAN MINIMAL BERBASIS PESERTA DIDIK

¹⁾Djemari Mardapi, ²⁾Samsul Hadi, ³⁾Heri Retnawati

^{1, 2, 3)}Universitas Negeri Yogyakarta

¹⁾djemarimardapi@gmail.com, ²⁾samsul.hd@gmail.com, ³⁾retnawati.heriuny1@gmail.com

Abstrak

Penelitian ini bertujuan untuk menentukan kriteria ketuntasan minimal (KKM) yang sering disebut pula dengan *standard setting* atau *cut of score*, berbasis peserta didik untuk mata pelajaran matematika SMP di Provinsi DI Yogyakarta antarwaktu. Penelitian ini merupakan studi longitudinal selama 3 tahun yaitu tahun 2010, 2011, dan 2012. Metode pengumpulan data yakni dokumentasi untuk memperoleh data statistik ujian akhir yang akan ditentukan KKM-nya dan *Focus Group Discussion* (FGD) untuk menentukan kelompok master dan kelompok nonmaster. Analisis data dilakukan secara kuantitatif dengan metode grup kontras. Hasil penelitian menunjukkan bahwa KKM mata pelajaran matematika di Provinsi DI Yogyakarta tahun 2010 sebesar 6,75, tahun 2011 mengalami penurunan sebesar 0,05 menjadi 6,70 dan tahun 2012 naik lagi menjadi 6,90 (untuk skala 0-10).

Kata kunci: KKM (*standard setting* atau *cut of score*), mata pelajaran matematika SMP

DETERMINING THE CUT OF SCORE BASED ON THE TEST PARTICIPANTS

¹⁾Djemari Mardapi, ²⁾Samsul Hadi, ³⁾Heri Retnawati

^{1, 2, 3)}Universitas Negeri Yogyakarta

¹⁾djemarimardapi@gmail.com, ²⁾samsul.hd@gmail.com, ³⁾retnawati.heriuny1@gmail.com

Abstract

This study aims to determine the minimum completeness criteria which is often called the standard setting or the cut of a score, based on test participants for mathematics lesson in Yogyakarta province. This study is a longitudinal study using data for 3 years ie 2010, 2011, and 2012. The data were collected through documentation to get the statistics data of national final examination that would be determined its minimum completeness criteria and Focus Group Discussion (FGD) to determine master and non master groups. The data were analyzed quantitatively by the method of contrasting group. The results showed that cut of score in mathematics lesson in Yogyakarta province in 2010 was at 6.75, but decreased by 0.05 to 6.7 in 2011 and rose again in 2012 to 6.9 (for a 0-10 scale).

Keywords: *standard setting or the cut of the score, junior school mathematics lesson*

Pendahuluan

Saat ini, kurikulum yang digunakan pemerintah yakni kurikulum tingkat satuan pendidikan (KTSP) ataupun kurikulum 2013 merupakan kurikulum berbasis kompetensi. Pada kurikulum ini, siswa dikatakan berhasil jika telah menguasai kompetensi tertentu yang telah ditetapkan dalam kurikulum. Indikator bahwa siswa telah menguasai kurikulum yakni kemampuan hasil belajar yang diukur telah mencapai kriteria ketuntasan minimal (KKM) yang telah ditetapkan, bahkan sebaiknya melampaui KKM. Dengan KKM ini, siswa yang telah berhasil dapat melanjutkan belajar untuk dapat menguasai kompetensi selanjutnya, dan yang belum menguasai dapat memperdalam yang belum dikuasai melalui remidi. Hal ini menunjukkan pentingnya KKM dalam menentukan keberlanjutan belajar peserta didik (Mardapi, Hadi & Retnawati, 2014a, Mardapi, Hadi & Retnawati, 2014b).

Selama ini, kriteria ketuntasan minimal atau dalam istilah pengukuran sering disebut dengan batas lulus (*cut of score*) ditentukan menggunakan kebijakan. Sebagai contoh pada ujian nasional (UN). Pada tahun 2005, batas lulus yang ditentukan 4,01. Untuk tahun 2006 dan 2007, batas lulus ditingkatkan menjadi 4,26, yang kemudian di tahun 2009 ditingkatkan lagi menjadi 5,26. Nilai 4,01, 4,26 maupun 5,26 merupakan batas yang relatif rendah dibandingkan batas lulus negara-negara lainnya. Namun demikian masyarakat meresponnya dengan penuh kecemasan dan keresahan, dan batas ini dianggap terlalu tinggi.

Di sisi lain, sekolah menentukan KKM yang cukup tinggi ketuntasan ulangan harian atau untuk kelulusan ujian sekolah, misalnya 7,5. Nilai ini ditetapkan oleh musyawarah guru berdasarkan intake, kompleksitas, dan daya dukung yang dimiliki sekolah. Akibatnya, KKM di suatu sekolah berbeda dengan sekolah lain. Akibatnya, KKM untuk suatu daerah misalnya daerah kabupaten/kota, belum dapat ditentukan karena sekolah-sekolah yang terdapat di suatu kabupaten memiliki KKM yang ber-

beda-beda, sehingga standar kelulusan daerah satu dengan yang lain sulit dibandingkan.

Penentuan kriteria ketuntasan minimal tidak hanya dapat ditentukan melalui kebijakan pemerintah seperti kriteria kelulusan dengan ujian nasional ataupun dengan memperhatikan *intake*, kompleksitas, dan daya dukung seperti yang dilakukan di sekolah. Penentuan batas lulus yang lain, dapat ditentukan dengan berbasis peserta tes dan yang berbasis alat ukur atau perangkat tesnya. Penentuan kriteria ketuntasan minimal dengan basis peserta didik dan alat ukurnya selama ini masih belum dilakukan. Padahal hal ini menjamin KKM yang valid dan dapat digunakan untuk menentukan KKM tingkat daerah, misalnya kecamatan atau kabupaten/kota.

Terkait dengan hal tersebut, diperlukan suatu penelitian tentang model penentuan kriteria ketuntasan minimal berbasis peserta didik dan berbasis perangkat tes yang digunakan sehingga dapat menentukan kriteria ketuntasan minimal yang tepat untuk menentukan keberhasilan pembelajaran yang dicapai peserta didik. Model penentuan kriteria ketuntasan minimal berbasis peserta didik dan perangkat tes menjadi hal yang urgen untuk ditemukan. Pada penelitian ini akan dikembangkan model penentuan kriteria ketuntasan minimal berbasis peserta didik dan perangkat tes, yang mendukung pelaksanaan evaluasi yang dilakukan oleh pendidik atau sekolah mengetahui keberhasilan pendidikan.

Pelaksanaan UAN memberikan pengaruh yang besar terhadap pelaksanaan pembelajaran di sekolah. Guru dan peserta didik hanya memfokuskan perhatian pada mata pelajaran yang diujikan saja (Mardapi, dkk. 2001). Demikian pula halnya dengan adanya batas kelulusan yang ditetapkan pemerintah. Selain siswa dan guru hanya terfokus pada mata pelajaran yang diujikan, batas kelulusan ini menyebabkan siswa merasa cemas dan takut tidak lulus (Mardapi, dkk., 2001). Studi pendahuluan menunjukkan, batas kelulusan yang ditentukan oleh UAN saja menyebabkan banyak siswa yang tidak lulus,

sehingga direspons berlebihan oleh masyarakat.

Ada beberapa istilah terkait dengan batas kelulusan. Di Indonesia yang paling populer adalah kriteria ketuntasan minimal, *cut of scoredan standard setting*. Dalam menentukan batas lulus, ada 3 jenis metode yang dapat digunakan, yakni metode yang terfokus pada *judgement*, tes dan peserta didik. Di Indonesia, batas kelulusan hanya ditentukan berdasarkan *judgement* dari pemerintah. Penentuan batas lulus berdasarkan perangkat tes yang digunakan dan pola respons peserta didik belum dilakukan. Penelitian-penelitian yang telah ada/dilakukan yakni penerapan teori penentuan batas lulus di negara-negara bagian Amerika.

Batas lulus untuk ujian nasional berbasis perangkat tes telah diteliti Mardapi, dkk. (2009). Hasil penelitian menunjukkan bahwa dengan menggunakan perangkat tes, batas lulus bisa lebih tinggi dibandingkan yang ditetapkan oleh pemerintah. Hal ini dapat dijadikan masukan untuk pembuat kebijakan dalam memperbaiki kualitas pendidikan, terkait dengan hasil evaluasi dan pemanfaatannya.

Ada beberapa metode penentuan batas lulus yang disarankan oleh ahli. Metode yang terpusat pada tes misalnya metode Nedelsky, metode Angoff dan metode Ebel (Algina dan Crocker, 1997). Metode yang terpusat pada siswa misalnya metode grup kontras (*contrasting group*) (Algina dan Crocker, 1997), sedangkan yang merupakan gabungan dari metode yang terpusat pada siswa, namun memanfaatkan pola respons siswa yakni metode pemetaan butir (*item mapping*) yang melibatkan teori respons butir unidimensi. Efektifitas perbandingan metode untuk menentukan batas lulus belum dilakukan, terlebih dengan menggunakan data tes dan peserta tes di Indonesia. Demikian pula stabilitas penentuan batas lulus berdasarkan metode tertentu perlu dilakukan, juga tentang dinamika batas lulus di wilayah tertentu antar waktu.

Kriteria ketuntasan minimal (KKM) disebut pula dengan batas lulus atau *standard setting*. *Standard* dapat diartikan sebagai ukur-

an atau patokan yang disepakati, dan *standard setting* adalah proses menentukan *cut score* terhadap instrumen pendidikan. Komponen esensial dari *standard setting* melalui *judgment* seperti yang dikemukakan oleh Angoff (1971), Nedelsky (1954), dan Plake, Melican, & Mills (1991) adalah panelis atau penilai ahli. Panelis ini sebaiknya memiliki delapan kualifikasi ahli bidang studi (*Subject Matter Expert*, SME) yakni (1) terbaik dalam bidang spesialisasinya; (2) memiliki wawasan yang luas dalam bidang keahliannya; (3) memiliki kemampuan menyelesaikan masalah dengan cepat sesuai bidangnya; (4) mampu mengkaji secara mendalam level konseptual dalam bidangnya dibandingkan orang baru; (5) menganalisis problem-problem dalam bidangnya secara kualitatif; (6) menilai problem secara lebih akurat dibandingkan orang baru; dan (8) mempunyai daya ingat semantik yang lebih kompleks.

Standard setting adalah proses yang digunakan untuk menentukan atau memilih suatu *passing score* pada suatu ujian. Dari semua langkah-langkah di dalam proses pengembangan tes, *standard setting* merupakan tahapan yang lebih dekat pada seni daripada sains (ilmu pengetahuan); sedang metode statistik yang sering digunakan di dalam pelaksanaan suatu *standard setting*, juga lebih banyak melalui pertimbangan dan atau kebijakan. Hattie & Brown (2003) menyatakan bahwa *setting performance standard* merupakan suatu proses meminta pertimbangan rasional dari para ahli yang (a) memiliki pengetahuan tentang kebutuhan akan tes dan asesmen yang ingin ditetapkan standarnya; (b) memahami makna skor pada level yang bervariasi pada skala yang digunakan untuk menyimpulkan performansi peserta tes; dan (c) memahami sepenuhnya batasan tentang prestasi yang berhubungan dengan standar performansi yang dimintakan kepada mereka untuk ditetapkan.

Ada beberapa cara menetapkan *standard setting*. Metode menentukan standar setting berbasis perangkat tes yakni metode Angoff tradisional dan Angoff perluasan (*Extended Angoff*) (Impara & Plake, 1997; Stahl, 2008; Ricker, 2009). Metode penentu-

an *standard setting* berbasis campuran peringkat tes dan peserta tes salah satunya adalah metode Bookmark (Lord, 1980; Lin, tth; Mitzel, Lewis, Patz & Green, 2001). Salah satu metode untuk menentukan *standard setting* berbasis peserta tes adalah metode grup kontras (Crocker & Algina, 1986).

Metode grup kontras (*contrasting group*) dari Nedelsky dilakukan dengan mengelompokkan grup yang dianggap kelompok master dan nonmaster, secara grafik digambarkan distribusinya, dan titik potong distribusi merupakan *cut of score*. Langkah-langkah menentukan *cut score* menggunakan metode grup kontras dari Nedelsky sebagai berikut:

- 1) mengumpulkan panel pakar berkualifikasi yang sudah mengenal karakteristik sekolah dan siswa yang diuji
- 2) para pakar mengidentifikasi dua kelompok siswa, kelompok yang jelas menguasai wilayah mata pelajaran yang dikriteriakan (kelompok master) dan kelompok siswa yang tidak menguasai wilayah mata pelajaran yang tidak menguasai (kelompok nonmaster)
- 3) menguji kedua kelompok tersebut
- 4) melukiskan grafik yang sama distribusi frekuensi skor ujian siswa-siswa pada kedua kelompok tersebut
- 5) mencari titik potong kedua distribusi frekuensi itu menjadi batas penguasaan minimum wilayah kriteria (Crocker & Algina, 1986).

Metode

Penelitian ini merupakan studi longitudinal selama 3 tahun untuk mengetahui kriteria ketuntasan minimal (KKM) mata pelajaran matematika di kota Yogyakarta. Data yang digunakan yakni data statistik ujian nasional mata pelajaran matematika tahun 2010, 2011, 2012 dari Puspendik Balitbang Kemendikbud Indonesia. Analisis data penentuan KKM dilakukan dengan metode grup kontras. Penentuan kelompok master dan nonmaster dalam grup kontras dilakukan dengan *focus group discussion* (FGD), yang melibatkan pengurus organisasi musyawarah guru mata pelajaran matematika

(MGMP) di Provinsi DI Yogyakarta. Hasil analisis KKM antartahun kemudian dibandingkan dengan menggunakan grafik.

Hasil

Langkah pertama menentukan KKM berbasis peserta didik adalah menentukan kelompok master dan nonmaster. Penentuan kelompok master dan nonmaster berdasarkan kriteria tertentu. Kriteria ini ditentukan dengan *Focus Group Discussion* (FGD). FGD yang melibatkan 3 ahli pengukuran, guru-guru dan pengurus MGMP Matematika se-Provinsi DI Yogyakarta sebagai pakarnya. Tiap wilayah kabupaten dan kota yakni kabupaten Bantul, Sleman, Kulon Progo, Gunungkidul dan Kota Yogyakarta diwakili oleh 1 pengurus MGMP dan 2 guru matematika SMP yang senior. Kriteria sekolah master dan nonmaster pada awalnya ditawarkan oleh forum yakni dengan rerata sekolah pada pencapaian UN untuk matematika sebesar 7,5, rerata prestasi peserta didik pada mata pelajaran tahun tertentu, dan juga rerata prestasi secara nasional. Namun, pakar dalam FGD menyepakati untuk kriteria sekolah master dan nonmaster ditentukan dengan rerata UN tingkat provinsi.

Dengan menggunakan kriteria yang telah disepakati, dikelompokkan sekolah-sekolah yang ada di Provinsi DI Yogyakarta sebagai kelompok master atau kelompok nonmaster. Tiap kelompok kemudian dibuat interval kemudian digambarkan polygon frekuensinya. Pertemuan polygon frekuensi kelompok master dan nonmaster merupakan KKM mata pelajaran matematika.

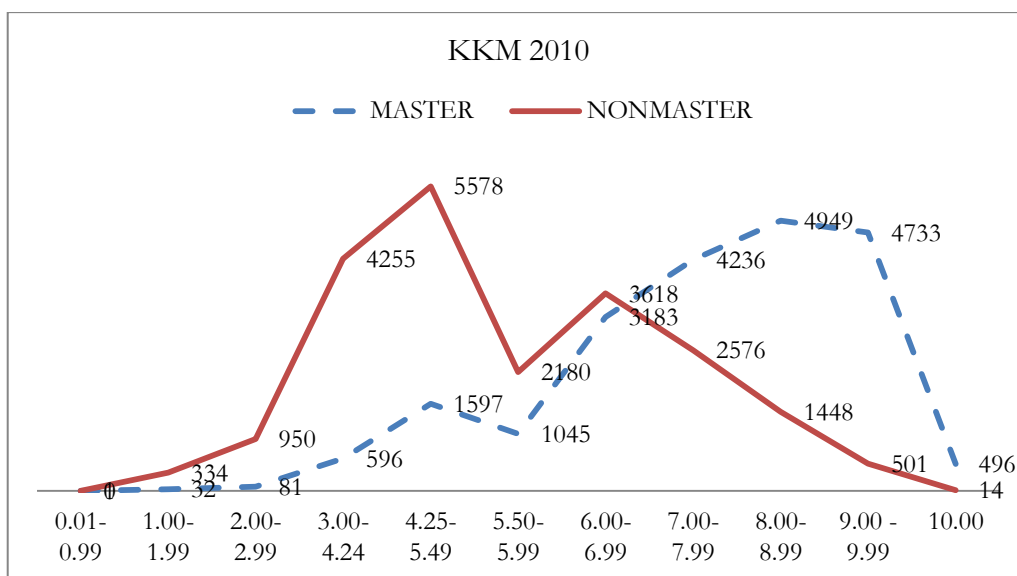
Rerata Ujian Nasional Matematika SMP di Provinsi Yogyakarta tahun 2010 yakni sebesar 6,46. Berdasarkan rerata ini, diklasifikasikan sekolah-sekolah menjadi kelompok master yaitu sekolah yang rata-rata nilai Ujian Nasional Matematikanya lebih dari atau sama dengan rata-rata nilai ujian nasional Provinsi DI Yogyakarta dan nonmaster yaitu kelompok sekolah dengan nilai rata-ratanya di bawah rata-rata kota. Setelah diketahui kelompok-kelompok sekolah master dan nonmaster, kemudian dilihat dis-

tribusi nilai UN matematika siswanya untuk setiap sekolah. Kelompok master terdiri dari 146 sekolah dan kelompok nonmaster sebanyak 276 sekolah. Selanjutnya, dirangkum hasil perhitungan distribusi nilai siswa dalam UN mata pelajaran matematika dari setiap kelompok. Hasil dari kelompok master dan nonmaster disajikan pada Tabel 1.

Berdasarkan Tabel 1, dibuat satu grafik untuk kelompok master dan nonmaster. Dari grafik tersebut, terdapat sebuah titik potong distribusi (*cut of score*) yang merupakan KKM yang disajikan pada Gambar 1. Berdasarkan gambar tersebut, diperoleh *cut of score* dari grafik adalah 6,75.

Tabel 1. Distribusi Nilai UN Matematika Provinsi DI Yogyakarta Tahun 2010 Berdasarkan Kelompok Master dan Nonmaster

	Master	Nonmaster
0,01- 0,99	0	1
1,00- 1,99	32	334
2,00- 2,99	81	950
3,00- 4,24	596	4255
4,25- 5,49	1597	5578
5,50- 5,99	1045	2180
6,00- 6,99	3183	3618
7,00- 7,99	4236	2576
8,00- 8,99	4949	1448
9,00 - 9,99	4733	501
10,00	496	14



Gambar 1. Grafik Distribusi Frekuensi Kelompok Master dan Nonmaster Provinsi di Yogyakarta Tahun 2010

Jumlah siswa sekolah yang masuk ke dalam kelompok master lebih sedikit dibandingkan siswa yang masuk dalam kelompok nonmaster. Hal tersebut dikarenakan rata-rata nilai matematika provinsi Yogyakarta cenderung tinggi yaitu 6,75, sedangkan sekolah yang masuk ke dalam kelompok nonmaster kebanyakan adalah sekolah swasta dengan jumlah murid yang hanya sedikit dan rata-rata nilainya cukup kecil. Setelah dianalisis menggunakan grafik

nilai KKM Provinsi DI Yogyakarta cukup tinggi yaitu 7,50. Nilai ini cukup tinggi jika dibandingkan dengan nilai rata-ratanya.

Rerata Ujian Nasional Matematika di provinsi DI Yogyakarta tahun 2011 adalah sebesar 6,87. Berdasarkan rerata ini, diperoleh sekolah kelompok master dan nonmaster. Dari 418 sekolah, terdapat 133 sekolah yang masuk kelompok master dan 285 sekolah masuk kelompok nonmaster. Hasil rangkuman distribusi nilai UN siswa

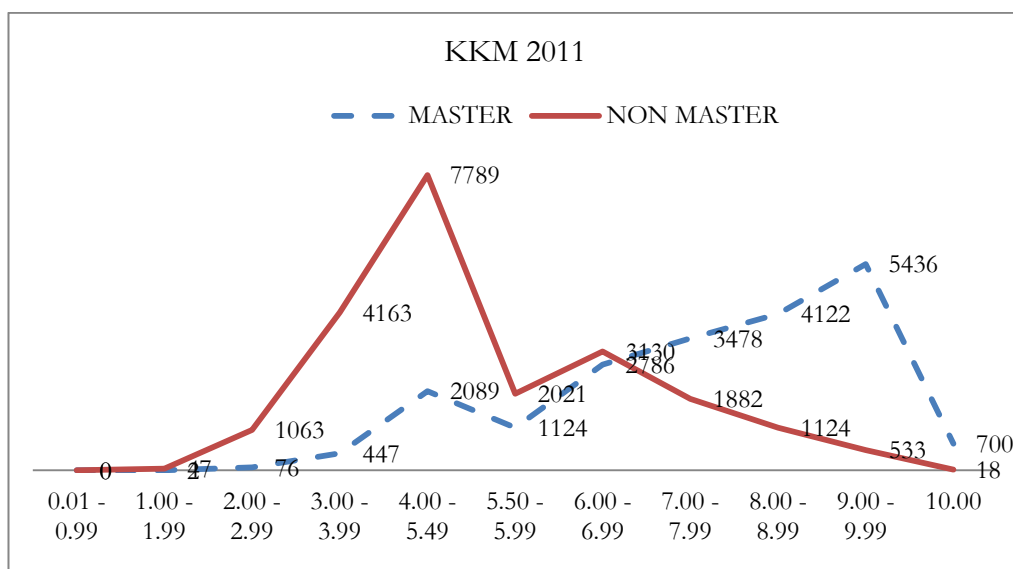
dalam mata pelajaran matematika dari kelompok master dan nonmaster yang disajikan pada Tabel 2.

Setelah itu dibuat satu grafik untuk kelompok master dan nonmaster untuk mengetahui perpotongannya yang berupa *cut of score* yaitu 6,70. Perpotongan tersebut merupakan nilai KKM yang diperoleh untuk mata pelajaran matematika tahun 2011. Grafik tersebut disajikan pada Gambar 2.

Kelompok nonmaster didominasi oleh sekolah-sekolah swasta dan banyak sekolah yang jumlah siswanya sangat sedikit. Nilai KKM yang diperoleh dari perpotongan grafik distribusi frekuensi kelompok master dan nonmaster tidak jauh berbeda dengan rata-rata Provinsi DI Yogyakarta yaitu lebih rendah dari 7.

Tabel 2. DIstribusi Nilai UN Matematika Kota Yogyakarta Tahun 2011 Berdasarkan Kelompok Master dan Nonmaster

	Master	Nonmaster
0,01 - 0,99	0	0
1,00 - 1,99	2	47
2,00 - 2,99	76	1063
3,00 - 3,99	447	4163
4,00 - 5,49	2089	7789
5,50 - 5,99	1124	2021
6,00 - 6,99	2786	3130
7,00 - 7,99	3478	1882
8,00 - 8,99	4122	1124
9,00 - 9,99	5436	533
10,00	700	18



Gambar 2. Grafik Distribusi Frekuensi Kelompok Master dan Nonmaster Kota Yogyakarta Tahun 2011

Pada tahun 2012, rerata Ujian Nasional di Kota Yogyakarta yakni sebesar 6,46. Berdasarkan rerata ini, diperoleh 17 sekolah kelompok master dan 40 sekolah kelompok nonmaster. Hasil rekap distribusi nilai siswa dalam setiap kelompok sekolah kemudian dirangkum. Hasil rangkuman distribusi nilai matematika dari kelompok master dan nonmaster yang disajikan pada Tabel 3. Berdasarkan hasil rangkuman pada Tabel 3 kemudian dibuat grafik. Untuk dis-

tribusi nilai kelompok master dan nonmaster dibuat dalam satu grafik, untuk mengetahui perpotongannya yang berupa *cut of score* yaitu 6,9. Hasil selengkapnya disajikan pada Gambar 3.

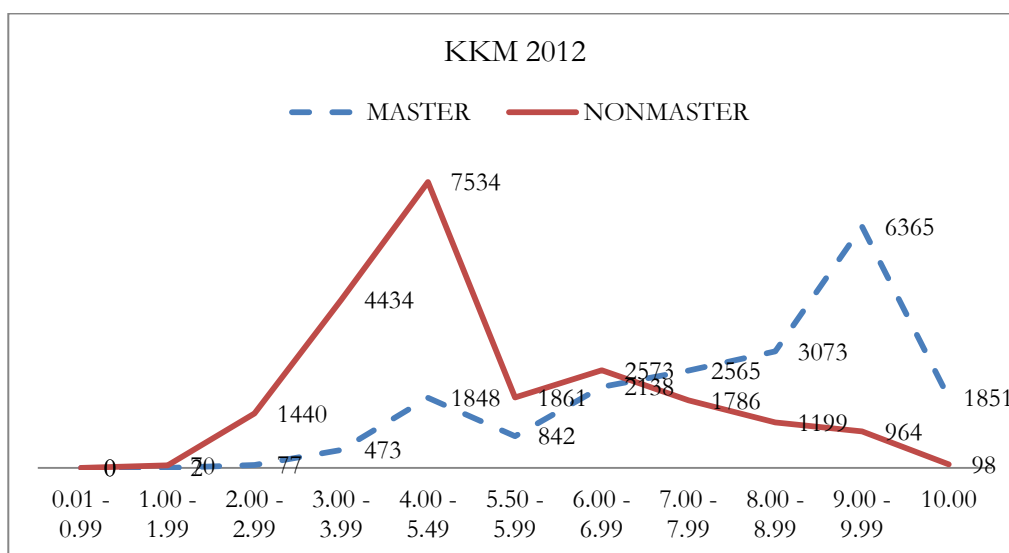
Pada tahun 2012 jumlah sekolah yang masuk dalam kelompok master lebih sedikit dibandingkan kelompok nonmaster, namun nilai rata-ratanya cenderung tinggi. Kelompok nonmaster juga masih didominasi sekolah swasta dan untuk nilai KKM yang

diperoleh cukup tinggi yaitu 6,9. Hasil penentuan KKM dengan menggunakan data tahun 2010, 2011, dan 2012 kemudian direkap. Hasilnya kemudian disajikan dengan grafik pada Gambar 4.

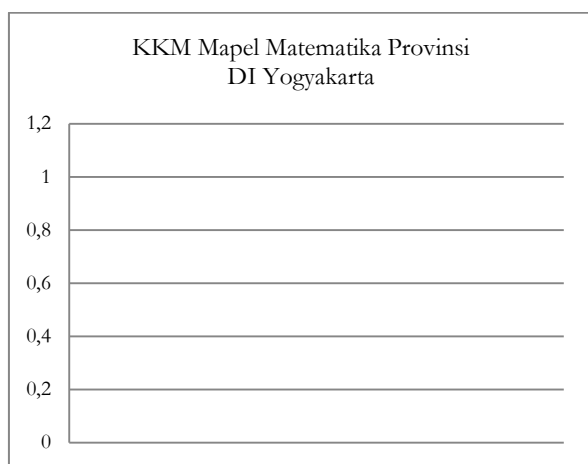
Berdasarkan Gambar 3 tersebut, tampak bahwa kriteria ketuntasan minimal provinsi DI Yogyakarta tergolong cukup, yaitu mendekati 7,00. Pada tahun 2010 KKM sempat berada pada skor 6,75 namun mengalami penurunan sebesar 0,05 pada tahun 2011. Dari tahun 2011 nilai KKM kembali naik pada tahun 2012 menjadi sebesar 6,90. Nilai KKM Provinsi DI Yogyakarta tersebut selama 3 tahun tertinggi diperoleh pada data tahun 2012.

Tabel 3. Distribusi Nilai UN Matematika Kota Yogyakarta Tahun 2012 Berdasarkan Kelompok Master dan Nonmaster

	MASTER	NONMASTER
0,01 - 0,99	0	0
1,00 - 1,99	2	70
2,00 - 2,99	77	1440
3,00 - 3,99	473	4434
4,00 - 5,49	1848	7534
5,50 - 5,99	842	1861
6,00 - 6,99	2138	2573
7,00 - 7,99	2565	1786
8,00 - 8,99	3073	1199
9,00 - 9,99	6365	964
10,00	1851	98



Gambar 3. Grafik Distribusi Frekuensi Kelompok Master dan Nonmaster Provinsi Daerah Istimewa Yogyakarta Tahun 2012



Gambar 4. Grafik Nilai KKM DI-Yogyakarta Berbasis Peserta Didik

Simpulan

Hasil penelitian menunjukkan bahwa KKM mata pelajaran matematika tahun 2010 sebesar 6,75, tahun 2011 menurun 0,05 menjadi 6,70 dan tahun 2012 naik kembali menjadi sebesar 6,90 (untuk skala 0-10).

Daftar Pustaka

Angoff, W. H. (1971). Scale, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, DC: American Council on Education.

- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: CBS College Publishing.
- Depdikbud. (1998). *Laporan Hasil Evaluasi Pelaksanaan Program Direktorat Jenderal Pendidikan Dasar dan Menengah tahun 1994/1995 s.d. 1997/1998*. Jakarta: Depdikbud.
- Ebel, Robert L. (1972). *Essentials of educational measurement*. Englewood Cliffs: Prentice-Hall.
- Hattie, J.A., & Brown, G. T. L. (2003, August). *Standard setting for asTTle reading: A comparison of methods*. asTTle Technical Report #21, University of Auckland/Ministry of Education.
- Impara, J.C., & Plake, B.S. (1997). Standard-setting: An alternative approach. *Journal of Educational Measurement*, 34, 353–366.
- Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 485–514). New York: American Council on Education/Macmillan.
- Kane, M.T. (2002a). Practice-Based standards Setting. *The Bar Examiner*, August 2002.
- Lin, J. (tth). *The Bookmark Standard Setting Procedure: Strengths and Weaknesses*. Alberta: The Centre for Research in Applied Measurement and Evaluation, The University of Alberta. Diambil pada tanggal 30 Oktober 2009 dari http://www.education.ualberta.ca/educ/psych/crame/files/standard_setting.pdf
- Mardapi, D. (2001). Ebtanas dalam tinjauan evaluasi pendidikan. *Bahan kuliah umum mahasiswa baru pascasarjana Universitas Muhammadiyah Prof. Dr. Hamka, tanggal 8 September 2001*.
- Mardapi, D., Hadi, S., & Retnawati, H. (2009). Standard setting. *Laporan penelitian*. Yogyakarta: Universitas Negeri Yogyakarta.
- Mardapi, D., Hadi, S., & Retnawati, H. (2014a). Determining the Standard Setting Based on the Students in the Junior High School in Yogyakarta Indonesia. *Proceeding International Conference of Research and Measurement of Thailand*, Burapha university 2014.
- Mardapi, D., Hadi, S., & Retnawati, H. (2014b). Menentukan Kriteria Ketuntasan Minimal (KKM) Berbasis Peserta Didik Mata pelajaran Matematika SMP di Kota Yogyakarta. *Prosiding Seminar Nasional LPPM UNY 2014*.
- Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The Bookmark Procedure: Psychological Perspectives. In G.J. Cizek (Ed.), *Setting Performance Standards*. Mahwah, NJ.
- Nedelsky, L. (1954). Absolute grading standards for objective test. *Educational and Psychological Measurement*, 14, 3-19.
- Plake, B. S., Melican, G. J., & Mills, C. N. (1991). Factors influencing intrajudge consistency during standard-setting. *Educational measurement: Issues and Practice*, 10(2), 15-16, 22, 25.
- Ricker, K. L. (2009). Setting Cut Scores: Critical Review of Angoff and Modified-Angoff Methods. Edmonton (Alberta, Canada): Centre for Research in Applied Measurement and Evaluation University of Alberta.
- Stahl, J. A. (2008). *Standard Setting Methodologies: Strengths and Weaknesses*. Illinois: Pearson VUE. Diambil pada tanggal 28 September 2009 dari www.iaea2008.cambridgeassessment.org.uk/ca

MODEL *AUTHENTIC SELF-ASSESSMENT* DALAM PENGEMBANGAN *EMPLOYABILITY SKILLS* MAHASISWA PENDIDIKAN TINGGI VOKASI

¹⁾I Made Suarta, ²⁾Nyoman Sentosa Hardika, ³⁾I Gusti Ngurah Sanjaya, & ⁴⁾I Wayan Basi Arjana
^{1, 2, 3, 4)}Politeknik Negeri Bali

¹⁾i_md_suarta@yahoo.co.id; ²⁾sen_hardika@yahoo.co.id; ³⁾ignsanjaya@yahoo.com;
⁴⁾basiarjana@hotmail.com

Abstrak

Tujuan penelitian ini adalah mengembangkan perangkat penilaian guna mengevaluasi capaian *employability skills* yang diintegrasikan dalam pembelajaran aplikasi basis data. Model penilaian yang dikembangkan merupakan kombinasi dari penilaian diri sendiri dan penilaian otentik, diusulkan dengan sebutan model *authentic self-assessment*. Langkah-langkah pengembangan model *authentic self-assessment* meliputi penentuan standar, penentuan tugas otentik, pembuatan kriteria, dan pembuatan rubrik. Hasil pengembangan perangkat penilaian terdiri atas: (1) model penilaian kemampuan pemecahan masalah; (2) model penilaian kemampuan mengelola diri sendiri; dan (3) model penilaian capaian pembelajaran aplikasi basis data. Model ini dapat digunakan untuk menilai aspek kognitif, afektif, dan psikomotorik. Hasil uji coba mengindikasikan capaian kemampuan pemecahan masalah dan kemampuan mengelola diri termasuk kategori baik, serta kompetensi merancang aplikasi basis data secara konseptual dan logikal termasuk kategori tinggi. Model ini juga telah memenuhi prinsip-prinsip dasar penilaian, yakni prinsip validitas, prinsip reliabilitas, terfokus pada kompetensi, komprehensif, prinsip objektivitas, dan prinsip mendidik.

Kata kunci: *penilaian otentik, penilaian diri, keterampilan pemecahan masalah, keterampilan mengelola diri, pendidikan vokasi*

AUTHENTIC SELF-ASSESSMENT MODEL FOR DEVELOPING EMPLOYABILITY SKILLS STUDENT IN HIGHER VOCATIONAL EDUCATION

¹⁾I Made Suarta, ²⁾Nyoman Sentosa Hardika, ³⁾I Gusti Ngurah Sanjaya, & ⁴⁾I Wayan Basi Arjana
^{1, 2, 3, 4)}Politeknik Negeri Bali

¹⁾i_md_suarta@yahoo.co.id; ²⁾sen_hardika@yahoo.co.id; ³⁾ignsanjaya@yahoo.com;
⁴⁾basiarjana@hotmail.com

Abstract

The purpose of this research is to develop assessment tools to evaluate achievement of employability skills which are integrated in the learning database applications. The assessment model developed is a combination of self-assessment and authentic assessment, proposed as models of authentic self-assessment. The steps of developing authentic self-assessment models include: identifying the standards, selecting an authentic task, identifying the criteria for the task, and creating the rubric. The results of development assessment tools include: (1) problem solving skills assessment model, (2) self-management skills assessment model, and (3) competence database applications assessment model. This model can be used to assess the cognitive, affective, and psychomotor achievement. The results indicate: achievement of problem solving and self-management ability was in good category, and competencies in designing conceptual and logical database was in high category. This model also has met the basic principles of assessment, i.e.: validity, reliability, focused on competencies, comprehensive, objectivity, and the principle of educating.

Keywords: *authentic assessment, self-assessment, problem solving skills, self-management skills, vocational education*

Pendahuluan

Proses pembelajaran meliputi tiga hal pokok yaitu merencanakan kegiatan pembelajaran, melaksanakan kegiatan pembelajaran, dan melakukan evaluasi pembelajaran. Evaluasi merupakan hal yang esensial dalam melihat apakah program pembelajaran yang direncanakan telah tercapai atau belum (Eseryel, 2002, Mavin et al., 2010). Sehubungan dengan evaluasi pembelajaran, maka sangat penting untuk dilakukan penilaian pembelajaran. Tujuan penilaian di antaranya untuk mengetahui tingkat ketercapaian tujuan pembelajaran dan melihat keefektifan proses belajar mengajar (Newton, 2007), serta meningkatkan prestasi dan memberikan kesempatan untuk mengembangkan peserta didik secara mandiri (Mc Sweeney, 2012). Sementara itu, Earl & Katz (2006) mengidentifikasi tiga tujuan penilaian yang berbeda namun saling berhubungan antara satu dengan lainnya, yakni penilaian untuk pembelajaran (*assessment for learning*), penilaian sebagai pembelajaran (*assessment as learning*), dan penilaian pembelajaran (*assessment of learning*).

Pendidikan vokasi merupakan jalur pendidikan untuk menyiapkan peserta didik mendapatkan pekerjaan melalui penguasaan keterampilan-keterampilan yang relevan dengan dunia kerja. Tujuan pembelajaran di politeknik bukan sekadar penguasaan konten materi yang diajarkan, namun peserta didik harus mampu menggunakan pengetahuan dan keterampilannya dalam menghadapi persoalan kehidupan sehari-hari. Cassidy (2006) menyatakan keterampilan yang dipelajari peserta didik selama mereka menempuh pendidikan dikelompokkan menjadi dua yakni keterampilan teknis dan keterampilan nonteknis. Keterampilan teknis terkait dengan keahlian dalam bidang pekerjaan tertentu. Keterampilan nonteknis atau dikenal dengan sebutan *employability skills* merupakan sejumlah keterampilan yang dapat digunakan dalam kehidupan sehari-hari di tempat kerja serta dapat ditransfer pada berbagai bidang pekerjaan dan profesi, seperti kerja sama dalam tim, kemampuan berkomunikasi, kemampuan pemecahan

masalah, kemampuan beradaptasi, kemampuan mengelola diri (Hager & Holland, 2006; Tome, 2007; Lowden et al., 2011).

Terdapat sejumlah faktor yang berkontribusi pada pengembangan keterampilan mahasiswa, seperti metode atau strategi pembelajaran, konteks pembelajaran yang relevan, serta aktivitas dan keterlibatan mahasiswa dalam pembelajaran. Cotton (2001) menyatakan perlu adanya penerapan praktik-praktik pendidikan yang bersifat khusus dalam menangani pengembangan keterampilan kerja. Cumming & Maxwell (1999) berpendapat konstruksi otentisitas dalam pendidikan berbasis kompetensi tidak diambil dari teori-teori belajar, tapi dari teori bahwa pendidikan vokasi harus difokuskan pada aplikasi langsung ke tempat kerja dan karenanya harus mencerminkan sedekat mungkin kinerja keterampilan kerja. Terkait dengan pembelajaran pada pendidikan vokasi, sangatlah penting dilakukan penilaian secara langsung atas unjuk kerja dan capaian keterampilan baik keterampilan teknis maupun keterampilan nonteknis.

Penilaian secara langsung atau lebih dikenal dengan asesmen otentik pada dasarnya merupakan penilaian unjuk kerja yang ditunjukkan sebagai akibat dari suatu proses belajar yang komprehensif. Menurut Mueller (2014) penilaian otentik merupakan suatu bentuk penilaian yang para peserta didik menunjukkan kompetensi, atau kombinasi pengetahuan, keterampilan dan sikap melalui tugas-tugas esensial pada situasi kerja yang sesungguhnya. Asesmen otentik juga dikenal dengan sebutan asesmen kinerja (*performance assessment*), asesmen alternatif, dan asesmen langsung (*direct assessment*). Asesmen otentik berpusat pada peserta didik, karena prosesnya dilakukan oleh pendidik dengan melibatkan peserta didik sendiri (Gulikers et al., 2004).

Asesmen otentik sangat dianjurkan dalam implementasi kurikulum berbasis kompetensi, karena penilaiannya dapat mencakup ke tiga aspek pembelajaran, yakni kognitif, afektif, dan psikomotorik. Menurut Jönsson (2008) kekuatan penilaian otentik terletak pada kesimpulan tentang kompe-

tensi siswa yang ditunjukkan melalui kinerja. Penilaian otentik dapat memberikan data yang lebih valid tentang kompetensi peserta didik serta memiliki dampak positif pada proses belajar peserta didik. Melalui perbaikan proses belajar, model penilaian ini juga dapat mengembangkan kemampuan peserta didik dalam menerapkan pengetahuan dan keterampilannya untuk pemecahan masalah-masalah nyata.

Salah satu prinsip asesmen otentik adalah mengukur kompetensi peserta didik dengan berbagai cara dan berbagai sumber. Sumber penilaian dapat berupa proses dan produk. Penilaian proses menekankan pada kemampuan mengelola diri sendiri dalam menghasilkan solusi atas permasalahan yang ditangani. Sementara itu, Mintah (2003) mengidentifikasi sejumlah teknik asesmen otentik, diantaranya: asesmen portofolio (*portfolio*), kinerja (*performance*), esai, proyek, dan penilaian diri (*self-assessment*).

Penilaian diri sendiri (*self-assessment*) menjadi visi baru dalam evaluasi pembelajaran untuk kemajuan studi peserta didik. Model penilaian ini menghendaki peserta didik menilai pekerjaan mereka sendiri, berdasarkan bukti dan kriteria yang jelas, untuk tujuan memperbaiki kinerja masa depan (McMillan & Hearn, 2008). Menurut Rolheiser & Ross (2014) asesmen diri adalah suatu cara untuk melihat ke dalam diri sendiri. Melalui asesmen diri peserta didik dapat melihat kelebihan maupun kekurangannya, untuk selanjutnya kekurangan ini menjadi tujuan perbaikan (*improvement goal*). Dengan demikian, peserta didik lebih bertanggung jawab terhadap proses dan pencapaian tujuan belajarnya.

Terkait evaluasi capaian keterampilan teknis dan nonteknis, penting untuk dikembangkan praktik-praktik penilaian yang relevan dengan tujuan tersebut. Cotton (2001) menyatakan penilaian yang dilakukan oleh teman sejawat merupakan contoh praktik-praktik pendidikan yang berkontribusi positif dalam pengembangan *employability skills* mahasiswa. Sementara itu, Sluijsmans et al (2008) mengintegrasikan asesmen otentik dalam pembelajaran berbasis kompetensi

pada pendidikan vokasi. Model *protocol portfolio scoring* (PPS) yang dikembangkan, memberikan nilai tambah dalam konteks paradigma "demand-driven" di mana tugas-tugas dan penilaian disesuaikan dengan kebutuhan dan persyaratan peserta didik.

Penelitian ini bertujuan untuk mengembangkan perangkat penilaian guna mengevaluasi capaian *employability skills* yang diintegrasikan dalam pembelajaran aplikasi basis data. *Employability skills* yang diintegrasikan meliputi keterampilan pemecahan masalah dan keterampilan mengelola diri sendiri. Kedua jenis keterampilan nonteknis ini merupakan indikator-indikator *employability skills* yang sangat penting bagi lulusan pendidikan vokasi untuk memasuki dunia kerja (Pellegrino & Hilton, 2012; CBI, 2012). Model penilaian yang dikembangkan merupakan kombinasi dari penilaian diri sendiri dan penilaian otentik, diusulkan dengan nama model penilaian diri secara nyata (*authentic self-assessment*).

Pengembangan model *authentic self-assessment* berangkat dari penerapan pembelajaran berbasis kompetensi. Menurut Mueller (2014) langkah-langkah pengembangan *authentic self-assessment* meliputi: (1) mengidentifikasi standar; (2) memilih tugas otentik; (3) mengidentifikasi kriteria untuk tugas-tugas otentik; dan (4) membuat rubrik. Identifikasi standar mengacu pada standar kompetensi mata kuliah aplikasi basis data. Tugas-tugas yang dibuatkan solusinya dalam bentuk aplikasi basis data diangkat dari permasalahan nyata di lapangan, khususnya yang terkait dengan kebutuhan perusahaan atau organisasi dalam pengolahan data. Kegiatan pembelajaran aplikasi basis data dilaksanakan dalam *setting* pembelajaran praktik di laboratorium komputer. Alasan pemilihan *setting* ini karena penguasaan *employability skills* mahasiswa akan efektif apabila melalui proses pembiasaan yang membutuhkan situasi nyata. *Setting* pembelajaran praktek merupakan perwujudan dari integrasi pengembangan keterampilan teknis dalam bidang aplikasi basis data dan *employability skills*.

Manfaat penelitian ini terutama bagi dosen adalah menambah wawasan, penge-

tahuan, kemampuan, dan keterampilan dalam merancang, mengembangkan serta menerapkan strategi penilaian hasil belajar yang inovatif. Model penilaian ini juga diharapkan dapat mengembangkan kemampuan peserta didik menerapkan pengetahuan dan keterampilannya dalam pemecahan masalah-masalah melalui proses internalisasi situasi kerja yang mendekati situasi nyata. Dengan demikian, penilaian diri secara langsung dapat memberikan informasi yang valid dan akurat tentang apa yang diketahui serta dapat dilakukan oleh peserta didik.

Metode Penelitian

Penelitian ini mengembangkan perangkat penilaian *employability skills* mahasiswa yang terintegrasi dengan pembelajaran Aplikasi Basis data. Prosedur penelitian dilaksanakan dalam tiga tahapan, yaitu: (1) persiapan, (2) pelaksanaan, dan (3) evaluasi. Kegiatan pokok pada tahapan persiapan meliputi: menentukan topik pembelajaran, menyusun panduan pelaksanaan pembelajaran, dan menyusun rubrik penilaian. Tahapan pelaksanaan penelitian diwujudkan dalam bentuk proses pembelajaran praktik di laboratorium komputer, melakukan evaluasi proses pembelajaran, dan mengevaluasi hasil belajar mahasiswa. Sementara tahapan akhir penelitian dilakukan kegiatan pengumpulan data, tabulasi, analisis data, serta menafsirkan hasil-hasil analisis.

Subjek ujicoba penelitian ini adalah mahasiswa Program Studi D3 Akuntansi, Jurusan Akuntansi, Politeknik Negeri Bali semester 4 tahun akademik 2013/2014 yang mengambil mata kuliah Aplikasi Basis Data. Jumlah mahasiswa sebanyak 73 orang terbagi ke dalam tiga kelas paralel. Evaluasi hasil belajar mahasiswa dilakukan dengan metode *authentic self-assessment*. Pada metode ini mahasiswa menilai dirinya sendiri secara jujur, objektif, serta penuh tanggung jawab atas capaian kemampuan *employability skills* dan capaian pembelajaran praktik dalam setiap subjek yang sudah dipelajari.

Perangkat penilaian disusun dalam bentuk rubrik yang terdiri atas: (1) rubrik

kemampuan *employability skills* mahasiswa, dan (2) rubrik capaian pembelajaran praktik aplikasi basis data. Berdasarkan *need analysis* (Suarta, dkk., 2013) keterampilan pemecahan masalah (*problem solving skills*) dan keterampilan mengelola diri sendiri (*self-management skills*) merupakan dua konstruk *employability skills* dengan tingkat kepentingan tertinggi menurut persepsi alumni, dosen, dan supervisor industri. Dengan demikian, pengembangan rubrik kemampuan *employability skills* mahasiswa difokuskan untuk menilai kemampuan pemecahan masalah dan kemampuan mengelola diri sendiri. Atribut-atribut kemampuan pemecahan masalah dan kemampuan mengelola diri disajikan pada Tabel 1, diadaptasi dan dikembangkan dari berbagai sumber (Mourtos, et al., 2004; Honey, 2008; Crebert, et al., 2011; Syed Hassan, et al., 2012; Hains-Wesson, 2013, p.6).

Tabel 1. Butir-Butir Penilaian Kemampuan *Employability Skills* Mahasiswa

No	Atribut-Atribut <i>Employability Skills</i>
Kemampuan pemecahan masalah (<i>problem solving</i>)	
1.	Mengidentifikasi permasalahan serta menggali informasi yang relevan dengan permasalahan
2.	Mendefinisikan permasalahan yang akan dicari-kannya solusi
3.	Mengevaluasi berbagai alternatif solusi serta mengurutkannya
4.	Mengambil keputusan atau mengambil tindakan berdasarkan kriteria yang sudah ditetapkan
5.	Melihat konsekuensi atas putusan yang diambil
6.	Menerapkan solusi yang telah diputuskan serta memeriksa hasilnya
7.	Memecahkan permasalahan secara mandiri
Kemampuan mengelola diri sendiri (<i>self-management</i>)	
8.	Bertanggungjawab terhadap tindakan-tindakan yang diambil
9.	Membuat rencana kerja (pembelajaran) secara sistematis
10.	Melaksanakan rencana kerja (pembelajaran) secara konsisten
11.	Bersikap tenang dalam menghadapi situasi penuh tekanan
12.	Melakukan evaluasi diri dan mengupayakan perbaikan guna peningkatan kinerja (prestasi belajar)
13.	Memiliki keyakinan akan kemampuan untuk menyelesaikan pekerjaan (tugas-tugas perkuliahan)
14.	Mengelola berbagai sumberdaya yang dimiliki untuk menghasilkan unjuk kerja terbaik

Penilaian dilakukan dengan memberikan skor antara 1 (sangat tidak baik/sangat rendah) hingga 5 (sangat baik/sangat tinggi). Sementara itu, rubrik capaian pembelajaran praktik, disesuaikan dengan kompetensi dasar pada setiap topik pembelajaran aplikasi basis data.

Produk-produk hasil pengembangan model *authentic self-assessment* divalidasi oleh pakar bidang pendidikan teknologi dan kejuruan (*expert judgment*). Data hasil uji coba model *authentic self-assessment* dianalisis secara deskriptif kuantitatif. Analisis deskriptif kuantitatif dimaksudkan untuk mendapatkan gambaran atau deskripsi kemampuan *employability skills* mahasiswa serta capaian pembelajaran praktik mahasiswa.

Hasil dan Pembahasan

Pengembangan Model Authentic Self-Assessment

Model *authentic self-assessment* yang dikembangkan dalam penelitian ini meliputi model untuk menilai: (1) kemampuan *employability skills* mahasiswa, dan (2) capaian pembelajaran praktik aplikasi basis data. Dalam aplikasinya, kedua model penilaian ini digunakan secara bersama-sama. Hal ini dilakukan karena pengembangan *employability skills* mahasiswa dilakukan secara terintegrasi dengan pembelajaran aplikasi basis data. Pengembangan model *authentic self-assessment* dilakukan dalam empat tahapan, yaitu: (1) penentuan standar; (2) penentuan tugas otentik; (3) pembuatan kriteria; dan (4) pembuatan rubrik.

Pada bagian berikut ini dijelaskan hasil pengembangan model penilaian capaian pembelajaran praktik aplikasi basis data:

Penentuan Standar

Standar yang dimaksud adalah sebuah pernyataan tentang apa yang harus diketahui atau dapat dilakukan peserta didik. Dalam hal ini, standar mengacu pada standar kompetensi mata kuliah aplikasi basis data yang dirumuskan sebagai berikut: “*Mahasiswa mampu merancang aplikasi basis data dalam bidang akuntansi.*” Standar kompetensi ini di-

jabarkan menjadi beberapa kompetensi dasar. Standar kompetensi dan kompetensi dasar menjadi acuan dan tujuan yang ingin dicapai dalam keseluruhan proses pembelajaran. Pada bagian ini, contoh kompetensi dasar yang akan dinilai yakni: “*Mahasiswa mampu membuat rancangan basis data secara konseptual dan logikal.*”

Penentuan Tugas Otentik

Tugas otentik adalah tugas-tugas yang secara nyata dibebankan kepada peserta didik untuk kemudian diukur capaian kompetensinya. Pemberian tugas-tugas otentik disesuaikan dengan standar kompetensi dan kompetensi dasar yang telah ditetapkan, serta relevan dengan kehidupan nyata. Sebelum mahasiswa diminta untuk mengerjakan tugas-tugas otentik sesuai dengan indikator-indikator kompetensi dasar, terlebih dahulu dijelaskan berbagai hal terkait konsep dasar basis data dan perancangan basis data. Pengetahuan ini menjadi dasar bagi mahasiswa dalam membuat rancangan basis data, baik bersifat konseptual maupun logikal.

Setiap mahasiswa diberi kebebasan mengangkat tema aplikasi basis data, diangkat dari permasalahan nyata di lapangan. Berdasarkan tema aplikasi basis data yang sudah dipilih oleh setiap mahasiswa, selanjutnya diminta untuk: (1) mengidentifikasi entitas-entitas yang terdapat dalam aplikasi basis data yang dirancang, (2) mengidentifikasi atribut-atribut dari setiap entitas dalam aplikasi basis data yang dirancang, (3) menentukan atribut-atribut sebagai kunci utama, (4) mengidentifikasi hubungan yang ada di antara entitas-entitas, (5) membuat diagram relasi antarentitas, (6) mengidentifikasi kunci tamu dari setiap relasi yang ada di antara entitas, dan (7) merancang model data logikal menggunakan basis data relasional.

Pembuatan Kriteria

Kriteria merupakan pernyataan yang menggambarkan tingkat capaian dan bukti-bukti nyata capaian belajar peserta didik dengan kualitas tertentu yang diinginkan. Kriteria atau juga dikenal dengan indikator,

merupakan penjabaran dari kompetensi dasar sehingga jelas kemampuan, keterampilan, atau kinerja apa yang menjadi sasaran penilaian. Contoh rumusan kompetensi dasar: “*Mahasiswa mampu merancang basis data secara konseptual dan logikal*” dijabarkan menjadi tujuh indikator, sebagai berikut: (1) mengidentifikasi entitas-entitas yang terdapat dalam aplikasi basis data yang dirancang, (2) mengidentifikasi atribut-atribut dari setiap entitas dalam aplikasi basis data yang dirancang, (3) menentukan atribut-atribut sebagai kunci utama, (4) mengidentifikasi hubungan yang ada di antara entitas-entitas, (5) membuat diagram relasi antarentitas (*entity-relationship diagram*), (6) mengidentifikasi kunci tamu dari setiap relasi yang ada di antara entitas, dan (7) merancang model data logikal menggunakan basis data relasional.

Pembuatan Rubrik

Rubrik dapat dipahami sebagai sebuah skala penyekoran yang dipergunakan untuk menilai kinerja peserta didik untuk tiap kriteria terhadap tugas-tugas tertentu. Rubrik yang dibuat adalah rubrik analitis, yang penilaiannya diberikan untuk tiap kriteria. Dua hal pokok yang dibuat pada setiap rubrik, yaitu kriteria dan tingkat capaian kinerja tiap kriteria. Kriteria mengacu kepada indikator-indikator kompetensi dasar. Tingkat capaian kinerja ditunjukkan dalam angka-angka, diukur menggunakan skala 1 – 7. Skor 1 menunjukkan tidak ada kinerja yang ditandai dengan tingkat capaian pembelajaran sangat rendah atau sangat tidak kompeten, sedang skor 7 menunjukkan kinerja sangat meyakinkan dan bermakna yang ditandai dengan tingkat capaian pembelajaran sangat baik atau sangat kompeten.

Gambar 1 menunjukkan contoh rubrik penilaian kemampuan merancang aplikasi basis data secara konseptual dan logikal. Penilaian hasil rancangan (kinerja) dilakukan sendiri oleh mahasiswa, sehingga model penilaian ini dinamakan *authentic self-assessment*.

Sementara itu, model penilaian kemampuan *employability skills* mahasiswa difokuskan pada keterampilan pemecahan masa-

lah (*prob-lem solving*) dan keterampilan mengelola diri sendiri (*self-management*). Indikator-indikator penilaian mengacu kepada butir-butir sebagaimana dirumuskan pada Tabel 1. Penilaian kemampuan *employability skills* dilakukan bersamaan dengan penilaian capaian pembelajaran pada setiap kompetensi dasar. Hal ini dilakukan dengan maksud untuk dapat mengukur capaian pengembangan *employability skills* yang diintegrasikan dengan kompetensi tertentu.

Berilah penilaian secara jujur, objektif, serta penuh tanggung jawab terhadap diri sendiri, atas kemampuan yang saudara miliki dalam merancang basisdata secara konseptual dan logikal. Penilaian atas capaian kemampuan yang saudara miliki dilakukan dengan memberikan skor 1 – 7 (skor 1 = capaian pembelajaran sangat rendah / sangat tidak kompeten; skor 7 = capaian pembelajaran sangat baik / sangat kompeten).				
RUBRIK PENILAIAN KEMAMPUAN MERANCANG BASISDATA SECARA KONSEPTUAL DAN LOGIKAL				
Kode MK : MKB-46016		Nama Mhs :		
Mata Kuliah : Aplikasi Basis Data		NIM :		
Program Studi : D3 Akuntansi		Kelas/Smt :		
Jurusan : Akuntansi		Nilai :		
NO	KOMPONEN PENILAIAN	BOBOT (%)	SKOR	NILAI (B X S)
1	Kelengkapan identifikasi entitas	10		
2	Kelengkapan identifikasi atribut	10		
3	Identifikasi kunci utama	5		
4	Identifikasi bentuk hubungan antar entitas	10		
5	Rancangan E-R Diagram	20		
6	Penurunan kunci tamu dari setiap relasi	5		
7	Rancangan data model relasional	20		
8	Akurasi pemecahan masalah	10		
9	Dokumentasi tugas	10		
TOTAL NILAI				

Gambar 1. Model *Authentic Self-Assessment* untuk Kemampuan Merancang Basis Data secara Konseptual dan Logikal.

Validitas dan Reliabilitas Rubrik Kemampuan Employability Skills

Terdapat sejumlah prinsip dasar penilaian hasil belajar yang perlu dipedomani, diantaranya: (1) prinsip validitas, (2) prinsip reliabilitas, (3) terfokus pada kompetensi, (4) prinsip komprehensif, (5) prinsip objektivitas, dan (6) prinsip mendidik (Brookhart, 2003; Hemispheric Project, 2006; Jönsson, 2008). Terkait dengan prinsip-prinsip validitas dan reliabilitas, validitas instrumen penilaian dibuktikan dengan validasi berdasarkan pendapat ahli. Sedangkan reliabilitas instrumen penilaian dibuktikan dengan *Cronbach's Alpha*. Hasil-hasil estimasi reliabilitas disajikan pada Tabel 2.

Berdasarkan hasil-hasil tersebut, reliabilitas skor instrumen kemampuan peme-

cahan masalah dan kemampuan mengelola diri sendiri termasuk kategori rendah.

Tabel 2. Validitas dan Reliabilitas Rubrik Kemampuan *Employability Skills* Mahasiswa

No	Atribut Employability Skills	Reliabilitas
Kemampuan Pemecahan Masalah		0,657
1	Mengidentifikasi permasalahan serta menggali informasi yang relevan dengan permasalahan	
2	Mendefinisikan permasalahan yang akan dicarikan solusinya	
3	Mengevaluasi berbagai alternatif solusi serta mengurutkannya	
4	Mengambil keputusan atau mengambil tindakan berdasarkan kriteria yang sudah ditetapkan	
5	Melihat konsekuensi atas putusan yang diambil	
6	Menerapkan solusi yang telah diputuskan serta memeriksa hasilnya	
7	Memecahkan permasalahan secara mandiri	
Kemampuan Mengelola Diri Sendiri		0,421
8	Bertanggungjawab terhadap tindakan-tindakan yang diambil	
9	Membuat rencana kerja (pembelajaran) secara sistematis	
10	Melaksanakan rencana kerja (pembelajaran) secara konsisten	
11	Bersikap tenang dalam menghadapi situasi penuh tekanan	
12	Melakukan evaluasi diri dan mengupayakan perbaikan guna peningkatan kinerja (prestasi belajar)	
13	Memiliki keyakinan akan kemampuan untuk menyelesaikan pekerjaan (tugas-tugas perkuliahan)	
14	Mengelola berbagai sumberdaya yang dimiliki untuk menghasilkan unjuk kerja terbaik	

Prinsip terfokus pada kompetensi dan komprehensif tercermin pada kompetensi dasar dan penjabarannya dalam bentuk indikator-indikator. Sementara itu, dari sisi objektivitas menunjukkan sebanyak 89% mahasiswa menyatakan menilai dirinya secara jujur, serta sebanyak 89% menyatakan penilaian dilakukan secara objektif. Dengan demikian, nilai yang didapatkan mahasiswa menunjukkan tingkat kemampuan sesungguhnya dalam menyelesaikan suatu permasalahan. Penilaian yang dilakukan sendiri (*self-assessment*) merupakan bentuk penilaian yang mendidik, karena mahasiswa melakukan evaluasi atas capaian pembelajaran secara mandiri. Hasil evaluasi ini dijadikan dasar untuk melakukan perbaikan guna peningkatan kinerja. Kemampuan untuk melakukan evaluasi diri merupakan cerminan dari kemampuan mengelola diri sendiri.

Deskripsi Hasil-Hasil Uji Coba

Berikut ini disajikan hasil-hasil uji coba model *authentic self-assessment*. Perangkat model penilaian yang diujicoba meliputi: (1) rubrik kemampuan pemecahan masalah dan kemampuan mengelola diri, dan (2) rubrik capaian pembelajaran membuat rancangan aplikasi basis data secara konseptual dan logikal. Subjek uji coba adalah mahasiswa Program Studi D3 Akuntansi, Jurusan Akuntansi, Politeknik Negeri Bali semester 4 tahun akademik 2013/ 2014 yang mengambil mata kuliah Aplikasi Basis data. Jumlah mahasiswa sebanyak 73 orang terbagi ke dalam tiga kelas paralel.

Deskripsi Kemampuan Pemecahan Masalah

Statistik deskriptif menunjukkan jumlah skor minimal kemampuan pemecahan masalah adalah 19, jumlah skor maksimal 31, rerata sebesar 24.36, dan simpangan baku 2.28. Selanjutnya dengan menggunakan nilai rerata kriteria ($\bar{X}_k = 21$) dan simpangan baku kriteria ($\sigma_k = 4,67$) dapat disusun klasifikasi kemampuan pemecahan masalah dalam lima kategori seperti pada Tabel 3.

Berdasarkan statistik deskriptif tersebut, secara umum kemampuan pemecahan

masalah mahasiswa termasuk dalam kategori tinggi. Terdapat 31,5% mahasiswa mempunyai kemampuan pemecahan masalah dalam kategori sedang, dan sebanyak 68,5% termasuk dalam kategori tinggi dan tinggi sekali.

Tabel 3. Klasifikasi Kemampuan Pemecahan Masalah

Kategori	Rentang Skor	Jumlah responden	Persentase dari total
Rendah sekali	7,00 – 14,00	0	0,0%
Rendah	14,01 – 18,67	0	0,0%
Sedang	18,68 – 23,33	23	31,5%
Tinggi	23,34 – 28,00	47	64,4%
Tinggi sekali	28,01 – 35,00	3	4,1%
Total		73	100%

Deskripsi Kemampuan Mengelola Diri

Statistik deskriptif menunjukkan jumlah skor minimal kemampuan mengelola diri sendiri adalah 19, jumlah skor maksimal 31, rerata sebesar 25.48, dan simpangan baku 2,07. Selanjutnya dengan menggunakan nilai rerata kriteria ($\bar{X}_k = 21$) dan simpangan baku kriteria ($\sigma_k = 4,67$) dapat disusun klasifikasi kemampuan mengelola diri sendiri dalam lima kategori seperti pada Tabel 4.

Berdasarkan statistik deskriptif tersebut, secara umum kemampuan mahasiswa dalam mengelola diri termasuk dalam kategori tinggi. Terdapat 11,0% mahasiswa mempunyai kemampuan mengelola diri dalam kategori sedang, dan 89,0% termasuk dalam kategori tinggi dan tinggi sekali.

Tabel 4. Klasifikasi kemampuan mengelola diri

Kategori	Rentang Skor	Jumlah responden	Persentase dari total
Rendah sekali	7,00 – 14,00	0	0,0%
Rendah	14,01 – 18,67	0	0,0%
Sedang	18,68 – 23,33	8	11,0%
Tinggi	23,34 – 28,00	60	82,2%
Tinggi sekali	28,01 – 35,00	5	6,8%
Total		73	100%

Deskripsi Capaian Pembelajaran

Capaian pembelajaran yang dimaksud merupakan capaian kemampuan dalam membuat rancangan aplikasi basis data secara konseptual dan logikal. Total hasil penilaian dikonversikan dalam skala 100. Nilai terendah adalah 32.1, tertinggi 100, rerata nilai 77.88, dan simpangan baku 17.54. Klasifikasi capaian pembelajaran disusun dalam lima kategori seperti pada Tabel 5.

Berdasarkan statistik deskriptif tersebut, secara umum kemampuan mahasiswa dalam merancang basis data secara konseptual dan logikal cukup baik. Sebanyak 71,3% mahasiswa mempunyai kemampuan dalam kategori tinggi dan tinggi sekali, sementara sisanya sebanyak 28,7% mempunyai kemampuan dalam kategori rendah sekali hingga sedang.

Tabel 5. Klasifikasi Kemampuan Merancang Basis Data secara Konseptual dan Logikal

Kategori	Rentang Skor	Jumlah responden	Persentase dari total
Rendah sekali	≤ 45	4	5,5%
Rendah	46 – 55	7	9,6%
Sedang	56 – 70	10	13,7%
Tinggi	71 – 80	11	15,1%
Tinggi sekali	≥ 81	41	56,2%
Total		73	100%

Pembahasan

Pengembangan *employability skills* mahasiswa yang diintegrasikan dengan pembelajaran aplikasi basis data memberikan implikasi pentingnya model penilaian yang mampu mengukur kedua aspek keterampilan tersebut. Penilaian secara langsung terhadap kinerja dalam membuat aplikasi basis data menjadi sebuah model alternatif dalam mengukur kompetensi mahasiswa. Sementara itu, penilaian yang dilakukan sendiri oleh mahasiswa terhadap kinerja atau produk yang telah dihasilkan merupakan bentuk evaluasi diri dalam upaya melakukan perbaikan-perbaikan untuk mencapai kompetensi yang diharapkan. Hasil-hasil pengem-

bangun model *authentic self-assessment* yang meliputi model penilaian (1) kemampuan *employability skills* mahasiswa, dan (2) capaian pembelajaran praktik aplikasi basis data menunjukkan hasil yang positif dalam menilai aspek kognitif, afektif, dan psikomotorik. Hal ini sejalan dengan pendapat Allen & Velden (2005) yang menyatakan bahwa pada *self-assessment*, individu diminta untuk menilai tingkat keterampilan mereka sendiri pada domain yang berbeda-beda. Jönsson (2008) juga menyatakan penilaian otentik dapat memberikan data yang lebih valid tentang kompetensi peserta didik.

Hasil-hasil uji coba menunjukkan, sebanyak 68,5% mahasiswa menunjukkan kemampuan pemecahan masalah dalam kategori tinggi dan tinggi sekali, serta sebanyak 89,0% mahasiswa menunjukkan kemampuan mengelola diri sendiri dalam kategori tinggi dan tinggi sekali. Secara rata-rata, hasil ini dapat dimaknai sebagai indikasi kemampuan *employability skills* mahasiswa dalam kategori baik. Sejalan dengan pendapat McMillan & Hearn (2008) yang menyebutkan penilaian diri sendiri (*self-assessment*) menghendaki peserta didik menilai pekerjaan mereka sendiri, berdasarkan bukti dan kriteria yang jelas, untuk tujuan memperbaiki kinerja masa depan. Melalui asesmen diri peserta didik dapat melihat kelebihan maupun kekurangannya, untuk selanjutnya kekurangan ini menjadi tujuan perbaikan (*improvement goal*). Melalui perbaikan proses belajar, model penilaian ini dapat mengembangkan kemampuan peserta didik dalam menerapkan pengetahuan dan keterampilannya untuk pemecahan masalah-masalah nyata, serta membuat peserta didik lebih bertanggungjawab terhadap proses dan pencapaian tujuan belajarnya.

Capaian pembelajaran dalam bentuk kemampuan membuat rancangan aplikasi basis data secara konseptual dan logikal juga menunjukkan hasil yang sejalan dengan kemampuan *employability skills* mahasiswa. Sebanyak 71,3% mahasiswa mempunyai kemampuan dalam kategori tinggi dan tinggi sekali. Hasil ini memang belum bisa dibandingkan peningkatannya karena uji coba di-

lakukan pada satu kompetensi dasar saja. Namun demikian, dengan hasil di mana 70% lebih mahasiswa mempunyai kemampuan kategori tinggi dan tinggi sekali menunjukkan capaian pembelajaran sesuai dengan yang diharapkan.

Hasil pembuktian validitas instrumen penilaian kemampuan pemecahan masalah dan kemampuan mengelola diri sendiri menunjukkan hasil valid. Hal ini mengindikasikan bahwa instrumen tersebut dapat digunakan untuk mengukur kemampuan pemecahan masalah dan kemampuan mengelola diri sendiri. Hanya saja reliabilitasnya masih rendah karena kurang dari 0,70. Namun demikian, butir-butir kemampuan pemecahan masalah mempunyai reliabilitas yang lebih baik dibandingkan dengan butir-butir kemampuan mengelola diri sendiri. Model penilaian yang dikembangkan juga terfokus pada satu kompetensi dasar, dan bersifat komprehensif karena penilaian dilakukan dengan menggunakan rubrik analitis. Pada model ini, penilaian dilakukan terhadap semua indikator. Objektivitas penilaian juga menunjukkan hasil yang tinggi (sangat objektif) serta dilakukan secara jujur oleh mahasiswa. Sebanyak 89% mahasiswa menilai dirinya secara objektif dan jujur. Model *authentic self-assessment* yang dikembangkan telah memenuhi prinsip-prinsip dasar penilaian sebagaimana diungkapkan oleh Brookhart (2003), Hemispheric Project (2006) dan Jönsson (2008), yakni prinsip validitas, prinsip reliabilitas, terfokus pada kompetensi, komprehensif, prinsip objektivitas, dan prinsip mendidik. Model ini tentu memberikan proses pembelajaran kepada peserta didik melalui evaluasi diri secara terus menerus terhadap capaian kemampuan dan kompetensi. Hasil evaluasi ini menjadi dasar dalam melakukan perbaikan guna meningkatkan kinerja di masa mendatang.

Simpulan dan Saran

Hasil pengembangan perangkat penilaian guna mengevaluasi capaian *employability skills* yang diintegrasikan dalam pembelajaran aplikasi basis data terdiri atas model penilaian kemampuan *employability skills* maha-

siswa, dan model penilaian capaian pembelajaran praktik aplikasi basis data. Model penilaian kemampuan *employability skills* mahasiswa yang dikembangkan meliputi model penilaian kemampuan pemecahan masalah dan model penilaian kemampuan mengelola diri sendiri. Sementara model penilaian pembelajaran praktik aplikasi basis data, pada penelitian ini mengambil contoh kompetensi dasar "membuat rancangan aplikasi basis data secara konseptual dan logikal".

Model *authentic self-assessment* yang dikembangkan dapat digunakan untuk menilai aspek kognitif, afektif, dan psikomotorik. Pemecahan masalah dan mengelola diri merupakan bagian dari struktur kognitif dan afektif proses pembelajaran. Sementara itu, struktur psikomotorik ditunjukkan oleh kemampuan membuat rancangan aplikasi basis data secara konseptual dan logikal. Model ini telah memenuhi prinsip-prinsip dasar penilaian, yakni prinsip validitas, prinsip reliabilitas, terfokus pada kompetensi, komprehensif, prinsip objektivitas, dan prinsip mendidik. Berdasarkan hasil-hasil uji coba dapat disimpulkan bahwa kemampuan pemecahan masalah dan kemampuan mengelola diri mahasiswa dalam kategori baik, serta capaian kompetensi merancang aplikasi basis data secara konseptual dan logikal termasuk kategori tinggi.

Penelitian ini perlu dilanjutkan dengan melihat implikasi penerapan *authentic self-assessment* pada semua kompetensi dasar yang sudah dirancang dalam pembelajaran aplikasi basis data. Penelitian lanjutan juga diperlukan dengan menggunakan kelas kontrol sehingga bisa dibedakan hasilnya.

Daftar Pustaka

- Allen, J., & Velden, R. van der. (2005). The role of self-assessment in measuring skills. *REFLEX Working paper 2*. The Netherlands: Research Centre for Education and the Labour Market, Maastricht University.
- Brookhart, S.M. (2003). Developing measurement theory for classroom assessment: purposes and uses. *Educational Measurement: Issues and Practices*, Winter 2003. Diakses pada 1 Agustus 2014 dari http://datause.cse.ucla.edu/docs/smb_dev_2003.
- Cassidy, S. (2006). Developing employability skills: peer assessment in higher education. *Education + Training*, 48 (7): 508-517.
- CBI. 2012. *Learning to grow: what employers need from education and skills. Education and skills survey 2012*. Diakses pada 7 Juli 2014 dari <http://www.cbi.org.uk>.
- Cotton, K. (2001). *Developing Employability Skills*. Northwest Regional Educational Research Laboratory, Portland, OR. Diakses 1 April 2008 dari <http://www.nwrel.org/scpd/sirs/8/c015.html>.
- Crebert, G., Patrick, C.-J., Cragolini, V., Smith, C., Worsfold, K., & Webb, F. (2011). *Problem solving skills toolkit*. 2nd Edition: Diakses 1 Juli 2014 dari <http://www.griffith.edu.au/gihe/resources-support/graduate-attributes>.
- Cumming, J.J., & Maxwell, G.S. (1999). Contextualising authentic assessment. *Assessment in Education: Principles, Policies and Practices*, 6(2): 177-194.
- Earl, L. & Katz, S. (2006). *Rethinking classroom assessment with purpose in mind: assessment for learning, assessment as learning, assessment of learning*. Manitoba Education, Citizenship and Youth. Diakses pada 13 Agustus 2014 dari <http://www.edu.gov.mb.ca/ks4/assess/index.html>.
- Eseryel, D. (2002). Approaches to evaluation of training: theory & practice. *Educational Technology & Society*, 5(2): 93-98.
- Gulikers, J.T.M., Bastiaens, T.J., & Kirschner, P.A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research & Development*, 52(3): 67-86.

- Hager, P. & Holland, S. (2006). *Graduate attributes, learning, and employability*. The Netherlands: Springer.
- Hains-Wesson, R. (2013). *An example of common foundational steps in order to problem solve*, Deakin University, Melbourne, Vic.
- Hemispheric Project. (2006). *Assessment in Competency-Based Education*. Diakses 1 Juni 2014 dari http://www.moe.gov.tt/Docs/ICIU/CBET_Assessment.pdf.
- Honey, P. (2008). *Strengthen your strengths: a guide to enhancing your self-management skills*. England: Peter Honey Publication Ltd.
- Jönsson, A. (2008). Educative assessment for/of teacher competency. *Studies in science and technology education* No. 18. Diakses 1 Juni 2014 dari <http://www.mah.se/muep>.
- Lowden, K., Hall, S., Elliot, D., & Lewin, J. (2011). *Employers' perceptions of the Employability Skills of New Graduates*. The SCRE Centre University of Glasgow. Diakses 10 September 2013 dari <http://www.gla.ac.uk/faculties/education/scre/>.
- Mavin, S., Lee, L., & Robson, F. (2010). *The evaluation of learning and development in the workplace: A review of the literature*. Diakses tanggal 13 Agustus 2014 dari http://www.northumbria.ac.uk/static/5007/hrpdf/hefce/hefce_litreview.pdf.
- McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational HORIZONS*, 87: 40-49.
- Mc Sweeney, K. (2012). *Assessment for learning: from theory to practice*. Diakses 13 Agustus 2014 dari <http://stangelas.nuigalway.ie>.
- Mintah, J.K. (2003). Authentic assessment in physical education: prevalence of use and perceived impact on students' self-concept, motivation, and skill achievement. *Measurement in Physical Education and Exercise Science*, 7(3): 161-174.
- Mourtos, N. J., Okamoto, N. DeJong, & Rhee, J. (2004). *Defining, teaching, and assessing problem solving skills*. Paper presented in 7th UICEE Annual Conference on Engineering Education Mumbai, India, 9-13 February 2004.
- Mueller, J. (2014). *Authentic assessment toolbox*. North Central College, Naperville, IL. Diakses 27 Juli 2014 dari <http://jfmuller.faculty.noctrl.edu/toolbox/index.htm>.
- Newton, Paul E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education*, 14(2): 149-170.
- Pellegrino, J.W., & Hilton, M.L. (2012). *Education for life and work: developing transferable knowledge and skills in the 21st century*. Washington, D.C.: National Academy of Sciences. Diakses 7 Juli 2014 dari <http://www.nap.edu/catalog.php>.
- Rolheiser, C. & Ross, J. A. (2014) *Student self-evaluation: what research says and what practice shows*. Diakses 14 Agustus 2014 dari <http://blogs.orchardview.org/perreault/files/2013/11/Student-Self-Evaluation-article>.
- Sluijsmans, D. M. A., Straetmans, G., & Van Merriënboer, J. (2008). Integrating authentic assessment with competency based learning: the protocol portfolio Scoring. *Journal of Vocational Education and Training*, 60(2): 157-172.
- Suarta, I M., Sentosa Hardika, I N., Sanjaya, IGN., Basi Arjana, I W. (2013). *Pengembangan model pembelajaran employa-*

bility skills dalam meningkatkan daya saing lulus-an pendidikan tinggi vokasi. Laporan Penelitian Strategis Nasional Tahun 1 (2013). Badung: Politeknik Negeri Bali.

Syed Hassan, S.A.H., Yusof, K.M., Mohammad, S., Abu, M. S., & Tasir, Z. (2012). Methods to study enhancement of problem solving

skills in engineering students through cooperative problem -based learning. *Procedia - Social and Behavioral Sciences*, 56: 737 – 746.

Tome, E. (2007). Employability, skills and training in Portugal (1988-2000): Evidence from official data. *Journal of European Industrial Training*, 31(5): 336-357.

EVALUASI KOMPETENSI PENULISAN TESIS MAHASISWA S2 PROGRAM STUDI PENDIDIKAN TEKNOLOGI DAN KEJURAN

¹⁾Putu Sudira, ²⁾Soenarto, ³⁾Pardjono

^{1) 2) 3)} Universitas Negeri Yogyakarta

¹⁾putupanji@uny.ac.id, ²⁾soenarto@uny.ac.id, ³⁾jpardjono@yahoo.com

Abstrak

Penelitian ini bertujuan untuk mengevaluasi: (1) kemajuan penulisan tesis; (2) kompetensi penulisan tesis; (3) faktor-faktor yang mempengaruhi penulisan tesis; (4) faktor-faktor dominan yang mempengaruhi penulisan tesis; (5) tingkat kekreatifan capaian hasil tesis mahasiswa S2 Program Studi Pendidikan Teknologi dan Kejuruan (PTK). Penelitian ini adalah penelitian evaluasi formatif bersifat internal dilaksanakan pada bulan Mei hingga Oktober 2014. Responden penelitian mahasiswa S2 PTK angkatan 2009, 2010, 2011, 2012 sebanyak 59 orang. Pengumpulan data menggunakan angket skala Likert dan wawancara. Teknik analisis data menggunakan teknik analisis deskriptif kuantitatif. Hasil penelitian menunjukkan kemajuan penulisan tesis termasuk kategori sangat baik. Kompetensi mahasiswa dalam penulisan tesis termasuk kategori sangat baik. Penulisan tesis dipengaruhi oleh faktor dalam dan luar kampus. Faktor-faktor dominan yang mempengaruhi penulisan tesis adalah ketersediaan waktu diskusi dengan dosen pembimbing, kemampuan menerapkan metode penelitian dan analisis data, beban tugas di tempat kerja, kemampuan memahami literatur asing, dukungan dosen pembimbing, proses administrasi Pascasarjana, motivasi cepat lulus, dukungan keluarga, dan beasiswa. Tingkat kekreatifan mahasiswa dalam penulisan tesis termasuk kategori sangat baik.

Kata kunci: *evaluasi, penulisan tesis, kompetensi mahasiswa, Pendidikan Teknologi dan Kejuruan*

EVALUATION OF THESIS WRITING COMPETENCE OF MASTER'S DEGREE STUDENTS ON VOCATIONAL TECHNOLOGY EDUCATION STUDY PROGRAM

¹⁾Putu Sudira, ²⁾Soenarto, ³⁾Pardjono

^{1) 2) 3)} Universitas Negeri Yogyakarta

¹⁾putupanji@uny.ac.id, ²⁾soenarto@uny.ac.id, ³⁾jpardjono@yahoo.com

Abstract

This study aimed to evaluate: (1) the progress of writing a thesis; (2) the competence of writing a thesis; (3) factors that influence the writing of the thesis; (4) the dominant factors that influence the writing of the thesis; (5) creativity of students in writing a thesis. This study was an internal formative evaluation conducted in May and October 2014. Respondents of this study were master degree students on Vocational and Technology Education Study Programs class of 2009, 2010, 2011, 2012 as many as 59 people. The data were collected using questionnaires and interviews. The data were analyzed using quantitative descriptive analysis techniques. The results showed that the progress of the thesis was classified in the excellent category. The competence of students in writing thesis was classified in very well category. Thesis writing is influenced by internal and external campus factors. The dominant factors affecting the thesis were the availability of a discussion with the supervisor, the ability to apply research methods and data analysis, workload in the workplace, the ability to understand foreign literatur, supervisor support, administrative processes, motivation to pass quickly, family support, and scholarship. The level of creativity of students was clasified in the excellent category.

Keywords: *evaluation, thesis, student competence, Vocational and Technology Education*

Pendahuluan

Pendidikan Teknologi dan Kejuruan (PTK) atau *Vocational and Technology Education* (VTE) diselenggarakan dalam rangka: (1) persiapan lulusan memasuki kehidupan kerja (Hansen, 2009, p.13; Billet, 2009, pp. 4-5); (2) memilih pekerjaan bagi yang belum bekerja dan mengembangkan kapasitas diri pada pekerjaan-pekerjaan yang telah dipilih bagi yang sudah bekerja (Rojewski, 2009, p.19, 25; Pavlova, 2009, p.2, 9); (3) terus-menerus mengembangkan kemampuan memecahkan permasalahan melalui kehidupan kerjanya (Hollander & Mar, 2009, p.42; Emmerik, 2009); (4) perbekalan pengalaman pendidikan untuk mendukung berbagai kemungkinan transisi dari satu pekerjaan ke pekerjaan lainnya (Billet, 2009, pp.4-5; Heinz, 2009, p.163); (5) menciptakan sendiri lapangan pekerjaan sebagai wirausaha baru (Hollander & Mar, 2009, p.43; Billet, 2009, p.5); (6) mengembangkan ilmu pengetahuan, teknologi di bidang VTE melalui riset, hingga menghasilkan karya inovatif dan teruji; (7) memecahkan permasalahan dalam bidang VTE melalui pendekatan inter dan multidisipliner; (8) mengelola riset dan pengembangan bidang VTE yang bermanfaat bagi masyarakat dan keilmuan VTE; dan (9) mempublikasikan hasil-hasil penelitian dan pengembangan dalam jurnal nasional terakreditasi dan jurnal internasional baik sebagai dosen, guru, instruktur, pengembang, pengelola, penjamin mutu pendidikan (tim pengembang Prodi PTK).

Tujuan penyelenggaraan pendidikan pada S2 Program Studi PTK Program Pascasarjana Universitas Negeri Yogyakarta (PTK PPs UNY) adalah: (1) menghasilkan lulusan Magister Pendidikan (S2) PTK yang mampu meningkatkan proses pembelajaran di Lembaga Pendidikan dan Tenaga Kependidikan (LPTK) penyelenggara program S1 PTK, lembaga-lembaga Diklat, Lembaga Penjamin Mutu Pendidikan (LPMP), dan pusat-pusat Pengembangan Pelatihan Guru; mempunyai kemampuan riset dan pengembangan bidang PTK; mempunyai kemampuan mengembangkan potensinya sebagai perencana dan pengelola PTK; dan memi-

liki kemampuan dasar untuk mengikuti program pendidikan S3; (2) menghasilkan dan menyebarluaskan karya ilmiah di bidang PTK kepada berbagai pihak yang berkepentingan, baik para pembuat kebijakan, praktisi lapangan, para pemerhati, serta masyarakat pada umumnya; (3) mengembangkan paradigma baru dalam pengelolaan PTK.

Kompetensi riset dan penulisan tesis merupakan salah satu kompetensi yang diprogramkan dalam pembelajaran pada Prodi S2 PTK. Kompetensi riset tesis yang unggul menjadi tuntutan pendidikan sarjana dan pascasarjana. Kepemilikan kompetensi riset dan penulisan tesis yang unggul menyebabkan lulusan S2 PTK PPs UNY dapat: (1) mengembangkan ilmu pengetahuan dan teknologi di bidang PTK, menghasilkan karya-karya inovatif dan teruji; (2) memecahkan permasalahan di bidang PTK melalui pendekatan inter dan multidisipliner; (3) mengelola riset dan pengembangan bidang PTK yang bermanfaat bagi masyarakat dan keilmuan PTK dengan pengakuan nasional dan internasional; (4) mempublikasikan hasil-hasil karya riset dan pengembangan dalam jurnal nasional terakreditasi dan jurnal internasional.

Penguasaan kompetensi riset dan penulisan tesis lulusan S2 Prodi PTK PPs UNY dikembangkan melalui kegiatan: (1) penyelenggaraan pendidikan dan pengajaran berkualitas dan terpadu; (2) peningkatan efektifitas kehadiran dosen-dosen dalam perkuliahan agar keefektifan pembelajaran dapat dicapai; (3) peningkatan skill belajar mahasiswa; (4) pengembangan budaya mencari, memilah, dan memilih informasi bahan ajar melalui penyediaan bahan bacaan yang kontemporer dan *up to date* di perpustakaan dan internet; (5) melaksanakan perkuliahan secara kreatif, efektif, menyenangkan, dan inovatif; (6) fasilitasi mahasiswa mengikuti berbagai pertemuan ilmiah, baik di dalam kampus, nasional, maupun internasional; (7) studi banding di dalam dan luar negeri untuk membangun wawasan dan perbandingan perkembangan riset; (8) pelibatan mahasiswa dalam berbagai kegiatan peneliti-

an payung; (9) pengembangan tema-tema penelitian payung; (10) pelaksanaan seminar nasional dan internasional secara berkala agar dapat mengikuti perkembangan ilmu; (11) pembimbingan tesis secara intensif; (12) penerbitan Jurnal Pendidikan Vokasi kerja sama dengan Asosiasi Dosen dan Guru Vokasi Indonesia (ADGVI); (13) peningkatan kerjasama dengan institusi pendidikan dan non pendidikan dalam dan luar negeri.

Kompetensi riset dan penulisan tesis lulusan S2 Prodi PTK PPs UNY diharapkan berkembang sejalan dengan peningkatan: (1) penguasaan landasan filosofi, teori, metodologi penelitian, dan evaluasi PTK; (2) kemampuan melakukan analisis masalah melalui beberapa pendekatan sistem, ekonomi, sosial, budaya, hukum, politik; (3) kemampuan merintis, mengembangkan, mengelola jejaring kerjasama PTK dengan lembaga terkait; (4) kemampuan mengelola penerbitan ilmiah nasional.

Kompetensi riset dan penulisan tesis mencakup kemampuan: (1) merumuskan permasalahan, tujuan, dan manfaat penelitian; (2) merumuskan kerangka pikir setelah melakukan deskripsi, analisis, sintesis teori-teori dan hasil penelitian yang relevan; (3) mengkaji hasil-hasil penelitian yang relevan untuk memposisikan penelitian yang telah dilakukan; (4) menerapkan metode penelitian secara valid; (5) menyajikan hasil penelitian; (6) menyusun telaah kritis hasil penelitian menggunakan perspektif berbagai teori yang relevan; (7) menyimpulkan dan membuat saran lebih lanjut; (8) menerapkan format dan tata tulis sesuai pedoman Tesis & Disertasi yang berlaku di PPs UNY.

Mahasiswa S2 Prodi PTK PPs UNY dalam penulisan tesis perlu memiliki kemampuan memecahkan masalah PTK secara kreatif. Belajar memecahkan masalah secara kreatif, menurut Piirto (2011, p.38) membutuhkan tiga hal pokok yaitu: (1) berpikir secara kreatif (*think creatively*); (2) bekerja secara kreatif dengan orang lain (*work creatively with others*); (3) menerapkan inovasi (*implement innovation*).

Berpikir kreatif (*think creatively*) meliputi kemampuan menggunakan ide-ide dan teknik-teknik kreatif yang luas tidak terbatas; menciptakan ide-ide baru yang bermanfaat; menguraikan, mencocokkan kembali, menganalisis, dan mengevaluasi ide-ide yang sudah ada dalam rangka mengembangkan dan memaksimalkan upaya-upaya kreatif. Dalam proses berpikir kreatif diperlukan sikap keterbukaan, berani mengambil resiko, toleran terhadap perbedaan, disiplin diri, kepercayaan pada kelompok (Gabrill & Gibbs, 2009; Pope, 2005, p.xvi; Weisberg, 2006, p.60).

Bekerja kreatif dengan orang lain (*work creatively with others*) dilakukan dengan cara mengembangkan, menerapkan, dan mengkomunikasikan ide-ide baru kepada orang lain secara efektif; menjadi orang yang selalu terbuka dan responsif pada perspektif baru dan berbeda, menggunakan masukan dan *feedback* kedalam pekerjaan; menunjukkan orisinalitas dalam penciptaan dan pekerjaan serta memahami betul kenyataan dan batas-batas dalam mengadopsi ide-ide baru; melihat kegagalan sebagai kesempatan untuk belajar kembali, memahami bahwa kreativitas dan inovasi adalah proses jangka panjang, siklus keberhasilan mulai dari hal-hal kecil dan bahkan sering terjadi kesalahan (Piirto, 2011, p.38; Levitt, p.2002).

Penelitian ini merupakan penelitian evaluasi program penulisan tesis mahasiswa S2 PTK PPs UNY. Sejalan dengan pendapat Suchman (1973:30) tujuan penelitian ini yaitu: (1) mendeskripsikan kemajuan mahasiswa S2 Program Studi PTK dalam penulisan tesis; (2) mendeskripsikan kompetensi mahasiswa S2 Program Studi PTK dalam mengembangkan rasionalisasi problematika PTK menjadi tesis melalui pola penyelenggaraan penulisan tesis Program Studi PTK; (3) mengetahui faktor-faktor yang mempengaruhi mahasiswa S2 Program Studi PTK dalam penulisan tesis; (4) mengklasifikasikan faktor-faktor dominan yang mempengaruhi mahasiswa S2 Program Studi PTK dalam penulisan tesis; (5) mendes-

kripsikan tingkat kekreatifan capaian hasil tesis lulusan S2 Program Studi PTK.

Penelitian evaluasi atau evaluasi program dilaksanakan dalam rangka: (1) pengukuran efektivitas program yang akan dievaluasi; (2) pengumpulan informasi untuk perancangan kembali program kegiatan; (3) interpretasi hasil temuan evaluasi untuk pengambilan keputusan. Ada lima pertanyaan pokok yang harus diperhatikan dalam evaluasi program: (1) jenis/bentuk perubahan apa yang diinginkan? (2) dengan cara apa perubahan dapat dilaksanakan? (3) indikator apa yang menunjukkan bahwa perubahan tersebut sebagai dampak dari program kegiatan yang diterapkan? (4) apakah ada perubahan yang terjadi dalam program/ lembaga? Jika ada perubahan bermfaat bagi siapa saja? (5) apakah ada dampak sampingan yang tidak diharapkan (*unintended side effect baik yang positif maupun yang negatif*).

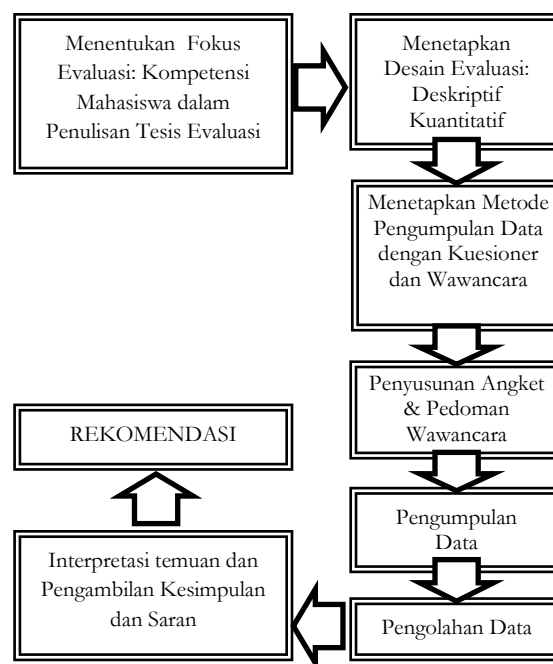
Menurut Brooks (1972, p.53) dalam evaluasi program ada 4 pertanyaan evaluasi yang harus dijawab melalui pengumpulan data dan analisis, yaitu: (1) apa tujuan dari program atau proyek? apa makna atau pentingnya arti dari berbagai tujuan program? (2) dengan sumber daya dan keterbatasan situasi yang ada, alternatif tindakan apa yang realistis untuk dipertimbangkan dalam pelaksanaannya? (3) konsekuensi atau dampak apa saja yang mungkin terjadi dari tindakan atau kegiatan yang dapat diperkirakan timbul? (4) dengan mempertimbangkan tujuan proyek atau program, ada/tidaknya alternatif-alternatif dan dampak yang dapat diperkirakan, kegiatan atau metode yang dapat diadopsi atau dikembangkan lebih lanjut?

Untuk menjawab pertanyaan evaluasi, evaluator/peneliti perlu menentukan langkah-langkah yang harus ditempuh dalam pelaksanaan evaluasi program. Suchman (1967, p.28) menyebutkan 6 langkah dalam penelitian evaluasi: (1) penentuan tujuan dan perumusan tujuan; (2) menentukan kriteria yang digunakan untuk mengukur keberhasilan (3) menentukan dan menjelaskan tingkat keberhasilan; (4) langkah-langkah evaluasi; dan (5) rekomendasi untuk program selanjutnya. Hal yang perlu difahami

dalam evaluasi program, bahwa dalam kenyataan di lapangan terjadi dampak yang diinginkan (direncanakan/diharapkan) disebut sebagai "*intended effects*" dan dampak sampingan yang tidak diharapkan "*unintended side effects*" tetapi terjadi secara bersamaan. *Unintended side effect* belum tentu jelek, justru kadang-kadang sebaliknya memberi efek yang sangat positif.

Metode Penelitian

Penelitian ini adalah penelitian evaluasi formatif bersifat internal tentang evaluasi kompetensi penulisan tesis mahasiswa S2 PTK. Penelitian ini dilaksanakan untuk peningkatan kinerja Prodi PTK dalam pengembangan program penulisan tesis. Penelitian dilaksanakan di Program Studi PTK PPs UNY pada bulan Mei sampai dengan Oktober 2014. Tahapan evaluasi program penulisan tesis mahasiswa S2 PTK PPs UNY disajikan pada Gambar 1.



Gambar 1. Tahapan Evaluasi Program Penulisan Tesis Mahasiswa S2 PTK

Kompetensi penulisan tesis adalah kemampuan mahasiswa dalam *problem solving* dimulai dari merasionalisasi problematika PTK menjadi sebuah penelitian sampai dengan menghasilkan tesis yang inovatif.

Kemampuan *problem solving* adalah proses berpikir tingkat tinggi berpikir kreatif, bekerja kreatif, dan menerapkan inovasi riset melalui proses mengenali cara-cara mencapai tujuan, memformulasikan sejumlah langkah penulisan dari permasalahan yang telah dikenali untuk ditemukan cara penulisannya. Kemampuan *problem solving* penulisan tesis mencakup kemampuan: (1) menemukan problematika-problematika riset PTK; (2) menetapkan judul tesis; (3) menyusun latar belakang penelitian; (4) menyusun identifikasi/diagnosis masalah, batasan/fokus masalah, rumusan masalah, tujuan penelitian, manfaat penelitian; (5) membangun landasan teori melalui kajian teori dan hasil penelitian yang relevan; (6) mengembangkan kerangka pikir; (7) menetapkan disain penelitian, tempat dan waktu penelitian, populasi dan sampel penelitian, variabel penelitian; (8) menetapkan teknik pengumpulan data; (9) menyusun instrumen penelitian yang valid dan reliabel; (10) melaksanakan penelitian; (11) mengumpulkan data; (12) mendeskripsikan data; (13) melakukan analisis data; (14) melakukan pem-

bahasan; (15) merumuskan simpulan, implikasi, saran; (16) mempresentasikan hasil penelitian; (17) memperhatahkan temuan penelitian; (18) melaporkan hasil penelitian akhir. Capaian pembelajaran penulisan tesis mencakup tesis sebagai hasil penelitian yang nyata, kreatif, memiliki *novelty*, kontekstual, memberi usulan solusi permasalahan secara riil. Responden penelitian adalah mahasiswa S2 PTK angkatan 2009, 2010, 2011, 2012 yang sedang menyelesaikan penulisan tesis sebanyak 59 orang. Instrumen penelitian menggunakan angket skala Likert dan lembar wawancara yang sudah divalidasi konten dan konstruksinya melalui *expert judgement* dan uji empiris. Analisis data yang digunakan dalam penelitian ini adalah dengan statistik deskriptif. Hasil analisis disajikan dalam bentuk tabel-tabel yang menampilkan jumlah responden, nilai rerata, standar deviasi, nilai minimum dan nilai maksimum. Perhitungan deskriptif ini menggunakan program SPSS 16.0. Tujuan dan indikator dari masing-masing tahap penulisan tesis seperti disajikan pada Tabel 1.

Tabel 1. Tujuan dan Indikator keberhasilan Penyelenggaraan Tesis S2 Prodi PTK

No	Tahap Penulisan Tesis	Tujuan	Indikator Keberhasilan	Waktu/Semester
1	Pengajuan Judul	- Mahasiswa mampu mengidentifikasi isu-isu, problematika, kondisi kontekstual PTK. - Mahasiswa mampu mensintesis ke dalam sebuah judul penelitian.	Mahasiswa mampu mengajukan 3 judul penelitian dilengkapi dengan permasalahan dan alasan rasional pemilihan judul, jenis penelitian, dan usulan pembimbing tesis.	II
2	Kuliah Proyek Penulisan Proposal Tesis	- Mahasiswa mampu merumuskan permasalahan berdasarkan referensi kondisi kontekstual serta data empiris - Mahasiswa mampu menghasilkan Proposal Tesis sebagai bentuk karya ilmiah sesuai dengan kaidah metode penelitian ilmiah	Mahasiswa menghasilkan proposal yang me-muat 3 bab yaitu Bab I memuat Pendahuluan, Bab II memuat Landasan Teori, Bab III Metode Penelitian.	Khusus
3	Kuliah Seminar Proposal Tesis	- Mahasiswa menguasai permasalahan-permasalahan yang diangkat dalam tesis, memilih metode penelitian, dan menerapkannya dalam bentuk proposal tesis. - Mahasiswa mampu membuat ringkasan proposal tesis dan mempresentasikannya sesuai kaidah presentasi.	- Mahasiswa menghasilkan proposal yang telah direvisi sesuai masukan saat seminar, termasuk instrumen penelitian yang akan digunakan. - Mahasiswa mampu mempresentasikan ringkasan proposal secara ringkas, jelas, dan tepat.	III
4	Penulisan Tesis Terbimbing	- Mahasiswa mendapatkan bimbingan mengenai pelaksanaan penelitian, analisis data dan pembahasan hasil penelitian, penulisan laporan tesis. - Mahasiswa dengan kesadaran dan motivasi untuk membaca, menulis, dan berkonsultasi dengan pembimbing.	Mahasiswa mampu memahami berbagai aspek penelitian tesis, dan mampu menyusun serta menyelesaikan laporan penelitian tesis.	III - IV
5	Ujian Tesis	- Mahasiswa mengklarifikasi proses dan hasil penelitian tesis yang sudah selesai ditulis. - Mahasiswa memperhatikan hasil penelitian di hadapan dewan penguji	Mahasiswa mampu memperhatikan hasil penelitian di hadapan dewan penguji	IV

Hasil dan Pembahasan

Kemajuan Penulisan Tesis

Kemajuan mahasiswa S2 PTK dalam penulisan tesis dievaluasi melalui kegiatan monitoring. Semua mahasiswa yang sedang menjalani proses penulisan tesis diundang dan menyampaikan laporan secara tertulis kemajuan dan pencapaian penulisan tesisnya masing-masing. Dalam acara monitoring mahasiswa dapat berdialog dengan ketua

dan sekretaris program studi PTK. Jumlah mahasiswa S2 PTK yang mengikuti monitoring kemajuan penulisan tesis sebanyak 51 orang (80%) dari 64 orang yang seharusnya hadir. Monitoring kemajuan penulisan tesis dilakukan untuk mengetahui capaian dan kemajuan setiap mahasiswa. Data kemajuan penulisan tesis mahasiswa S2 PTK disajikan pada Tabel 2.

Tabel 2. Kondisi Kemajuan Penulisan Tesis Mahasiswa S2 PTK

No.	Kemajuan Penulisan Tesis	Jumlah	Persentase
1	Penyusunan Proposal (BAB I, BAB II, BAB III)	23	45%
2	Pelaksanaan Uji Coba Instrumen	3	6%
3	Pengumpulan Data Penelitian	9	18%
4	Analisis Data dan Pengujian Hipotesis	10	20%
5	Penulisan Laporan	1	2%
6	Mendaftar Ujian	5	10%
Total		51	100%

Berdasarkan Tabel 2 akumulasi persentase kemajuan mahasiswa dalam penulisan tesis masih terbanyak pada penyusunan proposal. Hal ini wajar karena proses pematangan permasalahan penelitian dan perumusan teori untuk penyusunan paradigma penelitian membutuhkan kajian teori yang mendalam. Paradigma penelitian yang baik dihasilkan dari kajian teori yang baik pula. Mahasiswa membutuhkan waktu yang cukup dalam mengkaji dan mensintesis teori pendidikan teknologi dan vokasional.

Kompetensi Mahasiswa S2 Prodi PTK dalam Penulisan Tesis

Data kompetensi mahasiswa S2 PTK dalam penulisan tesis dibagi dalam empat evaluasi program yaitu: (1) Pengajuan Judul Tesis; (2) Proyek Proposal Tesis; (3) Seminar Proposal Tesis; (4) Penulisan Tesis. Tabel 3 menunjukkan data hasil evaluasi kompetensi mahasiswa S2 PTK PPs UNY dalam program penulisan penulisan tesis.

Tabel 3. Data Evaluasi Kompetensi Penulisan Tesis Mahasiswa S2 PTK

Nama Program	Jumlah Responden	Jumlah Item	Skor Teoritik		Skor Empirik					Distribusi Frekuensi					Capaian Kualitas		
			Tertinggi	Terdendah	Min	Max	Sum	Mean	Median	Modus	Std. Deviasi	S B (%)	B (%)	TB (%)	STB (%)	Persentase	Kategori
Pengajuan Judul Tesis	59	14	14	56	31	54	2724	46,17	47	45	4,36	56	41	3	0	82,45	SB
Proyek Proposal Tesis	59	35	35	140	74	136	6811	115,4	116	123	10,9	58	41	2	0	82,46	SB
Seminar Proposal Tesis	59	26	26	104	58	104	5041	85,44	86	74	9,0	54	44	2	0	82,15	SB
Penulisan Tesis	59	15	15	60	35	59	2989	50,66	52	55	5,57	66	32	2	0	81,86	SB

Keterangan:

SB = Sangat Baik; B = Baik; TB = Tidak Baik; STB= Sangat Tidak Baik

Data evaluasi program ini diperoleh melalui penyebaran kuesioner tertutup dengan pilihan jawaban skala 1, 2, 3, 4. Penyebaran skor evaluasi program penulisan tesis berdasarkan Tabel 3 menunjukkan bahwa keempat program penulisan tesis yaitu: (1) pengajuan judul tesis; (2) proyek proposal tesis; (3) seminar proposal tesis; (4) penulisan tesis semuanya dalam kategori sangat baik dengan rata-rata skor 82,23%.

Kompetensi Mahasiswa S2 PTK dalam Pengajuan Judul Tesis

Kompetensi pengajuan judul tesis mahasiswa S2 PTK termasuk kategori sangat baik. Sebanyak 56% mahasiswa merasa memiliki kompetensi yang sangat baik dan 41% merasa memiliki kompetensi baik dalam mengajukan judul tesis. Sebanyak 3% merasa kompetensinya masih tidak baik dan tidak ada mahasiswa yang merasa tidak memiliki kompetensi dalam mengajukan judul tesis. Data ini menunjukkan 96% mahasiswa sudah merasa mampu mengembangkan problematika penelitian berdasarkan isu-isu PTK secara terukur dan visibel untuk diteliti. Mahasiswa memiliki keyakinan bahwa topik penelitian yang diajukan memberi sumbangan dan manfaat untuk pengembangan keilmuan dan praksis PTK. Mahasiswa juga memiliki pemahaman yang sangat baik bahwa judul tesis yang diajukan didukung ketersediaan literatur dan dosen pembimbing yang memadai.

Tingginya kompetensi mahasiswa S2 PTK dalam pengajuan judul tesis disebabkan oleh beberapa faktor yaitu: (1) penyelenggaraan pendidikan dan pengajaran pada semester 1 dan 2 berkualitas tinggi dan terpadu; (2) dosen-dosen PTK disiplin dalam perkuliahan sehingga pembelajaran pada semester 1 dan 2 sangat efektif dan mendukung pencapaian penemuan permasalahan tesis; (3) tugas-tugas perkuliahan meningkatkan *skill* belajar mahasiswa dalam mencari dan memecahkan masalah; (4) budaya mencari, memilah, dan memilih informasi bahan ajar, sumber kajian literatur berkembang

dengan baik karena tersedia internet dan bahan bacaan yang kontemporer dan *up to date* di perpustakaan; (5) perkuliahan pada semester 1 dan 2 terlaksana secara kreatif, efektif, menyenangkan, dan inovatif; (6) mahasiswa terfasilitasi mengikuti berbagai pertemuan ilmiah, baik di dalam kampus maupun di luar kampus.

Kompetensi Mahasiswa S2 PTK dalam Proyek Proposal Tesis

Mahasiswa S2 PTK yang sedang menyelesaikan penulisan tesis sebanyak 58% merasakan memiliki kompetensi yang sangat baik dan 41% merasakan memiliki kompetensi baik dalam proyek penulisan proposal tesis. Hanya 2% merasa belum kompeten dalam menyelesaikan proyek proposal tesis. Mahasiswa S2 PTK sudah kompeten dalam merumuskan problematika penelitian, pentingnya penelitian, mengidentifikasi dan merumuskan masalah serta tujuan penelitian. Mahasiswa merasa kompeten dalam membangun kerangka pikir dan pertanyaan penelitian melalui sintesis berbagai teori yang relevan. Mahasiswa juga kompeten dalam merumuskan hipotesis penelitian, memilih pendekatan penelitian yang relevan, lokasi dan waktu penelitian, pemilihan populasi dan sampel penelitian, penetapan teknik pengambilan data dan penyusunan instrumen pengumpulan data yang valid dan reliabel, memilih teknik analisis data.

Pencapaian kompetensi proyek proposal tesis terwujud karena diselenggarakannya kuliah semester khusus yang didampingi oleh dua orang dosen pembimbing. Kegiatan semester khusus proyek proposal tesis diselenggarakan secara intensif mulai dari perumusan masalah sampai dengan pemilihan metode penelitian. Setiap mahasiswa di tengah semester harus mulai mempresentasikan *draft* proposalnya dan diakhir semester menyajikan kembali proposal yang sudah disempurnakan dari draf yang sudah disusun. Semua mahasiswa mengumpulkan proposal di akhir perkuliahan dan dinilai dengan bobot 1 sks.

Kompetensi Mahasiswa S2 PTK dalam Seminar Proposal Tesis

Seminar proposal tesis diselenggarakan pada semester 3 dengan tujuan menyempurnakan praproposal hasil proyek proposal tesis. Kuliah seminar proposal tesis dipandu oleh dua orang dosen secara tim. Dalam seminar mahasiswa harus menyajikan permasalahan penelitian berikut metode penelitian yang digunakan untuk menyelesaikan risetnya. Dalam seminar semua mahasiswa harus presentasi dua kali. Presentasi kedua harus menyajikan instrumen penelitian dan prosedur pengambilan data penelitian. Proposal penelitian dalam seminar harus sudah memenuhi format dan standar penulisan tesis sesuai panduan yang berlaku.

Kompetensi mahasiswa S2 PTK dalam seminar proposal penelitian termasuk kategori sangat bagus dengan data 54% merasa sangat baik dan 44% merasa termasuk baik kompetensinya dalam seminar proposal tesis. Mahasiswa sangat kompeten dalam mempresentasikan permasalahan, tujuan, dan manfaat penelitian berikut prosedur dan langkah-langkah melakukan penelitian secara terukur. Mahasiswa kompeten dalam mempresentasikan atau mengkomunikasikan proposal penelitian secara ringkas, jelas, dan tepat.

Dalam masa seminar mahasiswa sudah mulai terbimbing secara intensif sehingga mendapat masukan baik dari pembimbing maupun peserta seminar dan dosen pengampu mata kuliah seminar. Dalam proses seminar beberapa mahasiswa sudah selesai menyusun instrumen dan mulai melaksanakan pengumpulan data. Melalui seminar proposal tesis mahasiswa dapat menyempurnakan rancangan proposal penelitiannya dan lebih siap dalam melakukan penelitian.

Kompetensi Mahasiswa S2 PTK dalam Penulisan Tesis

Penulisan tesis terbimbing sudah dimulai pada proses proyek penulisan proposal tesis, berlanjut pada masa seminar proposal tesis dengan ditetapkannya seorang dosen pembimbing yang mendampingi

penulisan tesis. Penetapan dosen pembimbing tesis berdasarkan SK Direktur PPs UNY. Dengan demikian, semua pembimbing memiliki kewajiban yang mengikat untuk mendampingi mahasiswa dalam penulisan tesis. Penetapan dosen pembimbing tesis dilakukan melalui rapat koordinasi tim pengembang program studi. Dosen pembimbing ditetapkan berdasarkan kesesuaian keahlian pembimbing dengan topik penelitian tesis.

Sebanyak 66% mahasiswa merasa sangat kompeten dalam menulis tesis dan 44% memiliki kompetensi dalam kategori baik dalam menulis tesis. Sedikit mahasiswa yaitu 2% merasa kurang kemampuan menulis tesis. Mahasiswa merasa memiliki kompetensi yang baik dalam menulis tesis sesuai rambu atau kriteria penulisan karya ilmiah karena sudah disediakan buku panduan penulisan tesis yang sangat baik yang memuat struktur isi, format, dan cara penulisan tesis yang standar. Penulisan tesis harus memenuhi kriteria ejaan yang disempurnakan (EYD) dan bebas dari plagiarisme.

Faktor-Faktor Pendukung dan Penghambat Penulisan Tesis

Evaluasi keberhasilan penulisan tesis mahasiswa S2 PTK sangat perlu menemukan adanya faktor-faktor pendukung dan penghambat yang dirasakan dan dialami. Berdasarkan hasil wawancara dengan mahasiswa yang sedang menyelesaikan tesis diperoleh beberapa faktor pendukung dan faktor penghambat dalam penulisan tesis. Faktor-faktor pendukung dan penghambat dapat diklasifikasi menjadi dua yaitu dari dalam kampus dan luar kampus. Tabel 4 berikut menunjukkan klasifikasi dan urutan dominasi faktor-faktor penghambat dan pendukung yang dirasakan oleh mahasiswa S2 PTK tingkat akhir dalam proses penulisan tesis. Data ini sangat bermanfaat untuk rekomendasi perbaikan program penulisan tesis di Prodi S2 PTK.

Berdasarkan Tabel 4 kesulitan dalam mencari waktu diskusi dengan pembimbing dan beban tugas di tempat kerja merupakan faktor penghambat yang paling utama. Wa-

laupun hasil analisis evaluasi kompetensi penyelesaian tesis mahasiswa S2 PTK termasuk kategori sangat baik, ternyata masih ada beberapa mahasiswa (2%) yang mengalami hambatan. Mahasiswa S2 PTK yang sedang dan masih aktif sebagai guru di sekolah atau bekerja di instansi pemerintah dan swasta mengeluhkan beban kerja mengganggu kelancaran penulisan tesisnya. Beban kerja yang tinggi mengakibatkan pengaturan waktu bimbingan pada jam kerja menjadi sulit. Ini adalah penyebab utama yang menghambat penulisan tesis. Tidak

terlaksananya proses bimbingan antara mahasiswa dan pembimbing menyebabkan masalah-masalah lainnya seperti kesulitan memahami dan menerapkan metode penelitian, menganalisis data, mencari referensi sumber acuan jurnal asing bidang teknik menjadi tidak teratasi. Jarak rumah tinggal yang jauh dari kampus UNY juga membuat mahasiswa mengalami kesulitan mengatur waktu konsultasi dengan dosen pembimbingnya. Mahasiswa juga mengalami hambatan biaya pengembangan tesis R&D.

Tabel 4. Faktor Penghambat dan Pendukung Penulisan Tesis Mahasiswa S2 PTK

Faktor-Faktor	Penghambat	Pendukung
Dalam Kampus	• Kesulitan mencari waktu diskusi dengan pembimbing • Kesulitan memahami dan menerapkan metode penelitian • Kesulitan menganalisis data • Berkas terlalu lama dikoreksi pembimbing • Kesulitan mencari referensi sumber acuan jurnal asing bidang teknik • Ujian tertunda karena nilai mata kuliah teori belum keluar	• Mendapat dukungan dosen pembimbing • Proses administrasi PPs mudah • Motivasi cepat lulus • Mendapat dukungan rekan satu angkatan • Kolaborasi antar mahasiswa harmonis • Mendapat beasiswa • Mendapat bantuan akses literatur e-journal asing dari universitas lain
Luar Kampus	• Beban tugas di tempat kerja • Kurang mampu memahami literatur asing • Kurang mampu membaca pustaka rujukan kajian teori • Jarak kampus jauh dari rumah/pondokan • Urusan keluarga • Kesehatan menurun • Semangat belajar menurun • Kesulitan proses pengambilan data (lokasi penelitian jauh) • Biaya pengembangan tesis (R&D)	• Mendapat dukungan keluarga • Mendapat beasiswa

Faktor pendukung yang paling dominan dalam penulisan tesis adalah dukungan dosen dan dukungan keluarga. Ketersediaan dosen pembimbing yang kompeten, layanan administrasi akademik yang baik, atmosfir belajar penuh motivasi, hubungan antarmahasiswa yang harmonis, dukungan keluarga, tersedianya beasiswa merupakan faktor positif pendukung kelancaran penulisan tesis. Berikut solusi yang dilakukan oleh mahasiswa tingkat akhir yang telah lulus dan yang masih dalam proses penulisan tesis untuk mengatasi beberapa

hambatan yang dihadapi selama proses penulisan tesis: (1) membuat kesepakatan waktu bimbingan dengan dosen pembimbing minimal seminggu sekali; (2) disiplin waktu antara tesis dan tugas kerja, di rumah dan tanggung jawab menyelesaikan tesis; (3) membuat forum diskusi teman-teman seangkatan untuk saling bertukar pikiran dan berbagi kendala-kendala yang dihadapi; (4) mengikuti acara monitoring penulisan tesis setiap semester; (5) memanfaatkan internet melalui e-mail dalam bimbingan.

Tingkat Kreativitas Mahasiswa S2 PTK dalam Penulisan Tesis

Tingkat kreativitas mahasiswa S2 PTK dalam penulisan tesis di evaluasi dari laporan hasil penelitian tesis yang terdokumentasi di PPs UNY. Kriteria penilaian didasarkan pada keaslian, keunikan, dan kebaharuan gagasan penelitian. Di samping itu, juga dilihat ketepatan solusi dalam menyajikan data dan menguraikan analisis data dan pembahasan, keunggulan produk yang memberi manfaat pengembangan PTK. Sebaran tingkat kreativitas mahasiswa S2 PTK dalam penulisan tesis diajikan dalam Tabel 5 berikut ini.

Tabel 5. Distribusi Frekuensi Tingkat Kreativitas Capaian Penulisan Tesis Tahun 2014

No	Interval	Kategori	F	%
1	$X > 45,5$	Sangat Baik	19	54
2	$45,5 \geq X \geq 35$	Baik	16	46
3	$35 \geq X \geq 24,5$	Tidak Baik	0	0
4	$X < 24,5$	Sangat Tidak Baik	0	0
Jumlah			35	100

Penyebaran skor berdasarkan Tabel 5 menunjukkan bahwa tidak ada responden dalam kategori sangat tidak baik (0%); tidak ada responden dalam kategori tidak baik (0%); 16 responden dalam kategori baik (46%) dan 19 responden dalam kategori sangat baik (54%). Data tersebut menunjukkan mahasiswa S2 PTK sangat kreatif dalam menulis tesis. Gagasan penulisan tesis termasuk kategori cukup unik, asli, baru, dan produk penelitian memberi manfaat tinggi dalam pengembangan PTK.

Simpulan

Berdasarkan hasil analisis data dapat disimpulkan sebagai berikut:

1. Kemajuan mahasiswa S2 Program Studi PTK dalam penulisan tesis mulai dari pengajuan judul, proyek proposal, seminar, dan penulisan tesis termasuk kategori sangat baik.

2. Kompetensi mahasiswa S2 Program Studi PTK dalam penulisan penulisan tesis termasuk kategori sangat baik.
3. Faktor-faktor penghambat dan pendukung yang mempengaruhi mahasiswa S2 Program Studi PTK dalam penulisan tesis berasal dari dalam kampus dan luar kampus.
4. Tiga faktor penghambat dari dalam kampus yang paling dominan dirasakan oleh mahasiswa S2 PTK dalam penulisan tesis adalah kesulitan mencari waktu diskusi dengan dosen pembimbing, kesulitan memahami dan menerapkan metode penelitian, dan kesulitan menganalisis data.
5. Tiga faktor penghambat dari luar kampus yang paling dominan dirasakan oleh mahasiswa S2 PTK dalam penulisan tesis adalah beban tugas di tempat kerja, kurang mampu memahami literatur asing, kurang mampu membaca rujukan pustaka untuk kajian teori.
6. Tiga faktor pendukung dari dalam kampus yang paling dominan dirasakan oleh mahasiswa S2 PTK dalam penulisan tesis adalah mendapat dukungan dosen pembimbing, proses administrasi PPs mudah, motivasi cepat lulus.
7. Dua faktor pendukung dari luar kampus yang paling dominan dirasakan oleh mahasiswa S2 PTK dalam penulisan tesis adalah mendapat dukungan keluarga, dan beasiswa dari instansi asal.
8. Kreativitas mahasiswa S2 Program Studi PTK dalam penulisan tesis termasuk kategori sangat baik.

Rekomendasi

Berdasarkan hasil analisis data dapat direkomendasikan bahwa penulisan tesis melalui program pengajuan judul, kuliah proyek proposal tesis, seminar proposal tesis, penulisan tesis terbimbing perlu dipertahankan dan dilaksanakan lebih baik lagi. Faktor penghambat dan pendukung baik dari dalam kampus maupun dari luar kampus perlu diperhatikan dengan seksama. Kesulitan mahasiswa dalam mencari waktu diskusi dengan pembimbing karena beban kerja mahasiswa di tempat kerja masih ting-

gi dapat diatasi dengan diskusi melalui internet (e-mail dan media sosial). Perlu ada program penguatan kemampuan mahasiswa dalam membaca literatur asing khususnya penguatan bahasa Inggris. Beasiswa perlu dipertahankan bahkan diupayakan penambahan. Kuliah metodologi penelitian dan statistik lebih dikuatkan pada aspek penerapan melalui model perkuliahan *learn to solve problems*.

Daftar Pustaka

- Billet, S. (2011). *Vocational education: purposes, traditions, and prospects*. London: Springer Science+Business Media B.V.
- Brooks, W.F. (1972). *A self-evaluation supplement universities associated for research and education pathology*. New York: Russell Sage Foundation.
- Emmerik, I.J. H. V., Bakker, A.B, Euwema, M.C.. (2009). Explaining employees' evaluations of organizational change with the job-demands resources model; *Career Development International Journal*, 14 (6), 594-613.
- Gabril, E. & Gibbs, L. (2009). *Critical thinking for helping professionals*. New York: Oxford University Press
- Hansen, R.(2009). The pedagogical roots of technical learning and thinking. In R. Maclean, D. Wilson, & C. Chinien (Eds.), *International Handbook of Education for the Changing World of Work, Bridging Academic and Vocational Learning* (pp. 5-18). Bon: Springer.
- Heinz .W.R (2009). *Redefining the Status of Occupations*; in Rupert Maclean, David Wilson, Chris Chinien; *International Handbook of Education for the Changing World of Work, Bridging Academic and Vocational Learning*. Bon: Springer Science+Business Media.
- Hollander, A. & Mar, N.Y. (2009). Towards achieving TVET for all: the role of the unesco-unevoc international centre for technical and vocational education and training. In R. Maclean, D. Wilson, & C. Chinien (Eds.), *International Handbook of Education for the Changing World of Work, Bridging Academic and Vocational Learning* (pp. 41-58). Bon: Springer.
- Levitt, T. (2002). *Creativity is not enough*. diakses 16 April 2014 dari <http://hbr.org/2002/08/creativity-is-not-enough/ar/1>.
- Pavlova, M. (2009). *Technology and vocational education for sustainable development empowering individuals for the future*. Queensland: Springer Science Business Media B.V.
- Piirto, J. (2011). *Creativity for 21st century skills how to embed creativity into the curriculum*. Rotterdam: Sense Publishers
- Pop, R. (2005). *Creativity, history, theory, practice*. New York: Routledge.
- Rojewski, J.W (2009). A conceptual framework for Technical and Vocational Education and Training. In R. Maclean, D. Wilson, & C. Chinien (Eds.), *International Handbook of Education for the Changing World of Work, Bridging Academic and Vocational Learning* (pp. 19-40). Bon: Springer.
- Suchman, E. (1967). *Evaluative research: principles and practice in public service and social action programs*. New York: Russell Sage Foundation.
- Weisberg, R.W. (2006). *Creativity understanding innovation in problem solving, science, invention, and the arts*. New Jersey: John Wiley & Son.

PENGEMBANGAN INSTRUMEN DIAGNOSIS KESULITAN BELAJAR PADA PEMBELAJARAN KIMIA DI SMA

¹⁾Sri Yamtinah, ²⁾Budiyono

¹⁾Program Studi Pendidikan Kimia FKIP UNS, ²⁾PPs Universitas Sebelas Maret

¹⁾jengtina_sp@yahoo.com, ²⁾bud@uns.ac.id

Abstrak

Tujuan penelitian ini untuk: (1) mengembangkan instrumen *ordered multiple-choice* untuk diagnosis kesulitan belajar Stoikiometri kelas X, (2) menetapkan karakteristik dari instrumen, dan (3) membuat profil diagnostik peserta didik sebagai laporan yang informatif. Penelitian pengembangan ini menggunakan model Borg & Gall. *Preliminary study* menghasilkan pokok bahasan tersulit Stoikiometri. Penetapan atribut dan hierarki atribut dengan *focus group discussion* dilanjutkan dua putaran teknik Delphi. Indikator dan rumusan butir soal ditelaah *expert* untuk membuktikan validasi isi dengan rumus *Aiken*. Ujicoba terbatas dan ujicoba menengah pada sekolah kategori tinggi, sedang dan rendah. Ujicoba kelayakan dilakukan pada sekolah-sekolah di wilayah Surakarta, Karanganyar, Boyolali, dan Sragen. Kesimpulan: (1) penelitian berhasil mengembangkan tiga paket soal *Ordered Multiple Choice* (OMC) untuk mendeteksi kesulitan belajar peserta didik pada materi Stoikiometri kelas X; (2) butir-butir soal OMC pada paket A, B, dan C memiliki validitas konstruk yang baik dengan *Goodness of Fit* (GoF) berdasar Smart PLS > 0,36 yaitu 0,437; 0,466 dan 0,433. Instrumen memiliki reliabilitas tinggi berturut-turut 0,79; 0,81 dan 0,75; (3) profil peserta didik berupa laporan diagnostik tentang atribut yang sudah dikuasai dan belum dikuasai peserta didik.

Kata kunci: *Attribute Hierarchy Method, Ordered Multiple Choice, Graded Response Model*

DEVELOPING A DIAGNOSTIC INSTRUMENT FOR LEARNING DIFFICULTIES IN CHEMISTRY IN SECONDARY HIGH SCHOOL

¹⁾Sri Yamtinah, ²⁾Budiyono

¹⁾Program Studi Pendidikan Kimia FKIP UNS, ²⁾PPs Universitas Sebelas Maret

¹⁾jengtina_sp@yahoo.com, ²⁾bud@uns.ac.id

Abstract

The purposes of this study are to: (1) develop Ordered Multiple Choice (OMC) instruments with a model of Attribute Hierarchy Method (AHM) for the diagnosis of stoichiometry learning difficulties in chemistry learning in secondary high school, (2) determine the characteristic of instruments which have been developed based on the Graded Response Model (GRM), (3) create a diagnostic profile of learners as an informative report. This development research used the Borg & Gall model. The development of the instrument is done using the AHM with the form of OMC. The determination of attributes and attribute hierarchy was done by the focus group discussion (FGD) by three experts, six teachers, and two measurement experts and continued with the Delphi technique by three experts with two rounds. A limited try out was conducted in the high, medium, and low category schools. A feasibility try out was conducted to the high, medium, and low category schools in the regions of Surakarta, Karanganyar, Boyolali, and Sragen. The results are as follows. 1). This research has developed three packages of OMC to detect the learning difficulties of students in the subjects of Chemistry, especially in Stoichiometry of class X. 2). The OMC test items on packages A, B, and C have good construct validity with the Goodness of Fit (GoF) greater than 0.36 namely 0.437, 0.466, and 0.433. 3). The learners' profiles are created in the form of diagnostic report about the attributes which have been mastered and have not been mastered by the learners.

Keywords: *Attribute Hierarchy Method, Ordered Multiple Choice, Graded Response Model*

Pendahuluan

Kimia merupakan ilmu yang termasuk ke dalam rumpun IPA, ilmu yang mempelajari mengenai komposisi, struktur, dan sifat zat atau materi dari skala atom hingga molekul serta perubahan atau transformasi serta interaksi mereka untuk membentuk materi yang ditemukan sehari-hari. Pada awalnya kimia merupakan ilmu yang diperoleh dan dikembangkan melalui percobaan (induktif) akan tetapi perkembangan selanjutnya ilmu kimia juga diperoleh dan dikembangkan berdasarkan teori (deduktif). Sebagian aspek kimia bersifat kasat mata (*visible*), artinya dapat dibuat fakta kongkretnya dan sebagian aspek yang lain bersifat abstrak atau tidak kasat mata (*invisible*), artinya tidak dapat dibuat fakta kongkretnya. Namun demikian, aspek kimia yang tidak dapat dibuat fakta kongkretnya harus bersifat kasat logika, artinya kebenarannya dapat dibuktikan dengan logika matematika sehingga rasionalitasnya dapat dirumuskan (Depdiknas, 2003, p.2).

Mata pelajaran kimia di SMA bertujuan mempelajari segala sesuatu tentang komposisi, sifat dan struktur, transformasi, dinamika, dan energetika zat serta penerapannya untuk menyelesaikan masalah sehari-hari. Ilmu kimia merupakan produk (pengetahuan kimia yang berupa fakta, konsep, prinsip, hukum dan teori) temuan saintis dan proses (kerja ilmiah). Oleh sebab itu, dalam proses pembelajaran dan penilaian hasil belajar kimia harus memperhatikan karakteristik ilmu kimia sebagai produk, proses, nilai dan sikap ilmiah, serta aplikasi kimia dalam kehidupan sehari-hari.

Penelitian tentang konsepsi peserta didik dalam kimia didasarkan pada pendekatan konstruktivis untuk belajar, yaitu peserta didik membangun struktur kognitif mereka sendiri. Menurut pendekatan ini, peserta didik menghasilkan makna mereka sendiri berdasarkan latar belakang mereka, sikap, kemampuan, pengalaman sebelum, selama dan setelah proses pembelajaran. Oleh karena peserta didik membangun konsep mereka sendiri, maka bangunan konsepsi mereka seringkali berbeda dengan yang

dibawa oleh guru (Barke, Al Hazari & Yitbarek, 2009, p.2). Konsepsi adalah pemahaman terhadap suatu konsep. Pandangan konstruktivis tentang belajar adalah memperhitungkan ide murid baik dari pengamatan mereka sendiri atau dari bahasa sehari-hari, mengakui bahwa belajar akan melibatkan mengembangkan, memodifikasi dan bahkan menolak ide-ide yang ada, dan menerima pemahaman bahwa pembelajar membangun sesuatu untuk diri mereka sendiri (Bennet, 2007, p.28).

Terdapat dua pendekatan utama dalam konstruktivisme, yaitu konstruktivisme kognitif oleh Piaget dan konstruktivisme sosial oleh Vigotsky. Pada konstruktivisme kognitif, Piaget mengusulkan pembentukan skema atau struktur kognitif, yang merupakan makna, dan pemahaman tentang dunia individu. *Schemata* melayani beberapa fungsi dalam belajar yang mengkategorikan, mengingat, memahami dan memecahkan masalah. *Schemata* atau pengetahuan awal membuat hubungan untuk mengkategorikan pengalaman menjadi lebih efisien untuk diproses. Kategorisasi informasi memfasilitasi proses mengingat (*recall*), dan pemahaman (pemahaman), yang membuat pemecahan masalah lebih produktif. *Schemata* memfasilitasi pemilihan informasi berdasarkan kepentingan kita. Selanjutnya, setelah dipilih, *schemata* memungkinkan material yang dipilih diorganisir secara abstrak dan membantu individu dalam proses menafsirkan dan mengintegrasikan materi baru, berdasarkan apa yang sudah diketahui (Pritchard & Woolard, 2010, p.9). Selama peserta didik membangun atau membangun konsep-konsep mereka sendiri, dimungkinkan sebagian bangunan konsep mereka berbeda dari konsep para ahli.

Burton menyatakan bahwa peserta didik yang diduga mengalami kesulitan belajar, ditunjukkan oleh adanya kegagalan peserta didik dalam mencapai tujuan-tujuan belajar. Peserta didik dikatakan gagal dalam belajar apabila: (1) dalam batas waktu tertentu yang bersangkutan tidak mencapai ukuran tingkat keberhasilan atau tingkat penguasaan materi (*mastery level*) minimal

dalam pelajaran tertentu yang telah ditetapkan oleh guru (*criterion reference*), pada keadaan ini peserta didik disebut *lower group* (2) tidak dapat mengerjakan atau mencapai prestasi sebagaimana semestinya, dilihat berdasarkan ukuran tingkat kemampuan, bakat, atau kecerdasan yang dimilikinya. Peserta didik ini pada kondisi ini dapat digolongkan ke dalam *under achiever*, (3) tidak dapat mewujudkan tugas-tugas perkembangan, termasuk di dalamnya penyesuaian sosial sesuai dengan pola pada fase perkembangan tertentu pada kelompok usianya, kondisi seperti dikatakan siswa sebagai *immature*; (4) tidak berhasil tingkat penguasaan materi yang diperlukan sebagai prasyarat bagi kelanjutan tingkat peserta didik berikutnya, kondisi seperti ini disebut siswa *slow learners* (Burton, 1952, p.135).

Asesmen diagnostik adalah suatu pendekatan yang sangat menguntungkan dalam pengukuran. Penilaian diagnostik memberikan peluang melakukan suatu proses pengujian untuk melayani satu tindakan tambahan selain tujuan-tujuan yang telah ditetapkan, dan dapat digunakan untuk mengintegrasikan proses pembelajaran dan penilaian. Lebih lanjut pengujian diagnostik dapat menemukan bahan pelajaran yang bermakna terhadap setiap peserta didik (Mc Glohen, 2004, p.33).

Gorin dalam Leighton & Gierl (2007: 177) menyatakan bahwa tes diagnostik yang baik adalah yang dapat melakukan estimasi secara keseluruhan atas level kemampuan seseorang. Informasi yang spesifik harus dapat diperoleh dari respon peserta didik yang menunjukkan kelemahan atau inkonsistensi pengetahuan dan menunjukkan potret kemampuan peserta didik.

Model kognitif dalam pengukuran pendidikan mengacu pada gambaran sederhana bagaimana manusia menyelesaikan tugas-tugas standar atau tingkat yang rinci dalam memberikan penjelasan dan memberikan informasi tentang kekuatan dan kelemahan peserta didik (Nichols *et al*, 1995, p.4).

Model kognitif untuk *Cognitive Diagnostic Assessment* (CDA) memiliki setidaknya

empat karakteristik yang menentukan. *Pertama*, model ini berisi keterampilan tertentu pada ukuran butir yang tepat karena keterampilan ini harus memperbesar kinerja proses tes kognitif yang mendasari. *Kedua*, keterampilan harus dapat diukur. *Ketiga*, keterampilan harus relevan dengan pembelajaran dan bermakna bagi kelompok luas pemangku kepentingan pendidikan (*stakeholder*), termasuk peserta didik, orang tua, dan guru. *Keempat*, model kognitif sering mencerminkan hirarki keterampilan yang berjenjang dalam domain karena dependensi berbagi proses kognitif dan fungsi dalam jaringan proses, kompetensi, dan keterampilan lebih besar yang saling terkait. Penilaian yang didasarkan pada model kognitif dapat dikembangkan sehingga butir tes secara langsung dapat mengukur keterampilan kognitif tertentu yang meningkat kompleksitasnya, sehingga memungkinkan item yang menguji kinerja peserta didik dapat dihubungkan dengan informasi mengenai kekuatan dan kelemahan kognitif peserta didik dengan menggunakan model psikometrik dan kognitif diagnostik seperti *Attribute Hierarchy Method* (Leighton, *et al*, 2004, p.217; Gierl, *et al*, 2000, p.37).

Attribute Hierarchy Method (AHM) adalah metode psikometri untuk mengklasifikasikan respon jawaban butir dari peserta ujian menjadi satu set pola atribut terstruktur terkait dengan komponen yang berbeda dari model kognitif kinerja tugas. AHM ini didasarkan pada asumsi bahwa kinerja tes tergantung pada satu set kompetensi hierarkis berjenjang yang disebut atribut kognitif. Atribut adalah proses kognitif atau keterampilan dasar yang dibutuhkan untuk memecahkan tes. Peserta didik yang diuji harus memiliki atribut-atribut untuk menjawab butir tes dengan benar. Hirarki memainkan peran mendasar dalam AHM karena mengarahkan pengembangan tes dan panduan analisis sehingga skor tes psikometri memiliki nilai diagnostik. AHM ini dikembangkan untuk mengatasi dua masalah spesifik yang terkait dengan pengembangan model kognitif dan pengenalan pola statistik (Gierl, 2007, p.331; Gierl, Wang & Zhou, 2008).

Attribute Hierarchy Method (AHM) didasarkan atas asumsi bahwa butir-butir tes dapat dijelaskan melalui suatu set atribut berjenjang yang hirarkis. Asumsi dari ketergantungan atribut adalah sesuai dengan penemuan dari riset kognitif (Vosniadou & Brewer, 1992) di mana skill kognitif tidak dapat beroperasi secara terpisah tetapi merupakan suatu jaringan kompetensi yang saling terkait.

Data yang diperoleh dari studi pendahuluan melalui penyebaran angket pada guru-guru kimia di wilayah Surakarta pada bulan Maret-April 2012 menunjukkan bahwa materi Stoikiometri merupakan materi paling sulit. Hal tersebut didukung dengan data yang diperoleh guru kimia SMA Negeri 4 Surakarta, bahwa sebanyak 64,53% peserta didik kelas X dan 67,16% peserta didik kelas XI IPA tidak tuntas pada ulangan harian pokok bahasan Stoikiometri dengan batas tuntas 72 dan 75. Sementara itu, guru kimia SMA Muhammadiyah 1 Surakarta menyatakan bahwa 62,2% peserta didik kelas X tidak tuntas pada pokok bahasan Stoikiometri dengan batas tuntas 67.

Hasil penyebaran angket juga mendapatkan bahwa 72,2% guru menyatakan bahwa pokok bahasan Stoikiometri merupakan pokok bahasan esensial yang sangat diperlukan untuk memahami dan menguasai pokok bahasan-pokok bahasan lain dalam kimia. Namun demikian, sebanyak 90,7% guru menyatakan belum pernah melakukan upaya khusus untuk mendiagnosis melalui tes diagnostik. Sebanyak 9,3% guru menyatakan melakukan upaya diagnosis melalui tes uraian akan tetapi merasa kesulitan dalam memeriksa hasil pekerjaan peserta didik.

Penyebab kesulitan dalam memahami stoikiometri adalah kompleksitas hitungan yang memerlukan pemahaman tentang konsep mol, penyetaraan reaksi, kemampuan aljabar dan menginterpretasikan bahasa menjadi langkah-langkah prosedural untuk memperoleh jawaban yang benar (Okanlawon, 2010, p.108).

Salah satu alasan kesulitan peserta didik dalam memahami kimia adalah berkaitan dengan berbagai tingkat representasi

yang digunakan dalam pembelajaran kimia untuk menggambarkan dan menjelaskan fenomena kimia. Tiga representasi yang relevan untuk memahami konsep-konsep kimia adalah: (1) representasi sifat makroskopik yang menggambarkan sebagian besar fenomena nyata dan terlihat dalam pengalaman sehari-hari peserta didik ketika mengamati perubahan dalam sifat materi (misalnya perubahan warna, pH larutan mengandung air, dan pembentukan gas dan presipitat dalam reaksi kimia), (2) submikroskopis (atau molekul) representasi yang memberikan penjelasan pada tingkat partikulat yang dalam hal ini digambarkan sebagai terdiri dari atom, molekul dan ion, dan (3) simbolik (atau ikon) representasi yang melibatkan penggunaan simbol-simbol kimia, rumus dan persamaan, serta gambar struktur molekul, diagram, model dan animasi komputer untuk melambangkan hal (Chandrasegaran, Treagust & Mocerino, 2007, p.297). Dalam mempelajari kimia, kemampuan peserta didik untuk memahami peran setiap tingkat representasi dan mentransfer dari satu tingkat menjadi tingkat lain merupakan aspek penting untuk menghasilkan penjelasan yang dapat dimengerti (Rahayu, 2009, p.669)

Di dalam kelas, guru dapat menggunakan respon terbuka peserta didik untuk mendapatkan informasi tentang pemahaman peserta didik mengenai isi materi pelajaran. Namun, tuntutan mengajar banyak peserta didik (pada sekolah menengah) sering menghambat analisis mendalam tentang tanggapan peserta didik untuk butir *open-ended*. Guru yang telah memiliki sertifikat profesi memiliki kewajiban untuk mengajar di depan kelas 24 jam pelajaran per minggu, apabila alokasi mata pelajaran kimia adalah 3 jam per minggu maka guru tersebut harus mengajar sebanyak 8 kelas. Hal ini tentu akan menyulitkan guru untuk dapat melakukan tes diagnostik dengan butir soal uraian.

Penggunaan butir tes *multiple choice* yang memiliki kelebihan efisien dalam pemeriksaan hasil kerja peserta didik dianggap tidak memberikan informasi diagnostik ten-

tang pemahaman lebih cermat. Dengan demikian, sangat penting untuk dibuat sebuah instrumen yang dapat mengadopsi kelebihan butir tes jenis *open-ended* namun dapat juga memanfaatkan kepraktisan butir tes *multiple choice*.

Briggs, *et all* (2006) mengembangkan *Ordered Multiple Choice* (OMC) sebagai instrumen asesmen untuk fungsi diagnostik. OMC dikembangkan dengan mengkombinasikan efisiensi pilihan ganda tradisional dengan kelebihan kualitatif dari respon pertanyaan *open-ended*. Kunci utama pada pengembangan butir OMC adalah pada *construct map*. Istilah lain dari *construct map* adalah *learning progression*. *Construct map* adalah bentuk pendekatan perkembangan untuk mengukur pemahaman peserta didik. *Construct map* digunakan untuk menyatakan uni dimensionalitas pada masing-masing level. Setiap level merefleksikan tingkatan hirarkis dari pemahaman materi (Briggs, *et all*, 2006: 56).

Asumsi dari ketergantungan atribut adalah sesuai dengan penemuan dari riset kognitif (Vosniadou & Brewer, 1992) yang *skill* kognitif tidak dapat beroperasi secara terpisah tetapi merupakan suatu jaringan kompetensi yang saling terkait. Sampai sekarang, AHM lebih banyak hanya diperkenalkan didasarkan pada model IRT dikotomi. Ada asumsi untuk model dikotomis yang butir tes dapat dijawab dengan benar jika dan hanya jika semua atribut yang terlibat dalam butir tes dapat dijawab dengan benar. Ini berarti bahwa kehilangan satu atribut setara dengan kehilangan semua atribut yang diperlukan. Model Politomus dapat memberikan informasi lebih lanjut untuk skor dari dikotomi. Diagnosis yang akurat membutuhkan deskripsi yang kaya hasil ujian tes (Bolt, *et all*, 2004, p. 57).

Pada penelitian ini AHM dipadukan dengan bentuk tes *Ordered Multiple Choice* (OMC) yang menggunakan skor politomus. Tujuan penelitian ini adalah (1) mengembangkan instrumen *ordered multiple-choice* (OMC) dengan model *Attribute Hierarchy Method* (AHM) untuk diagnosis kesulitan belajar Stoikiometri kelas X, (2) menetapkan

karakteristik atau parameter butir dari instrumen yang telah dikembangkan berdasarkan pada *Graded Response Model* (GRM), dan (3) membuat profil diagnostik peserta didik sebagai laporan yang informatif.

Metode Penelitian

Penelitian ini merupakan jenis penelitian pengembangan yang bertujuan menghasilkan perangkat instrumen yang digunakan untuk mendiagnosis kesulitan belajar stoikiometri pada kelas yang besar. Model pengembangan menggunakan modifikasi model R & D dari Borg & Gall (1983, p.771-796).

Penggunaan model ini dengan pertimbangan bahwa model Borg & Gall khusus didesain untuk meningkatkan hasil pendidikan. Hal tersebut sejalan tujuan utama dari penelitian ini. Model penelitian dan pengembangan adalah proses yang digunakan untuk mengembangkan dan memvalidasi produk dalam dunia pendidikan.

Pengembangan diawali dengan *preliminary study* melalui telaah literatur dan penelitian empiris. Hasil dari telaah literatur menunjukkan bahwa materi stoikiometri merupakan konsep yang sulit sehingga di banyak negara pun, materi stoikiometri juga banyak menimbulkan kesulitan pada peserta didik. Hasil telaah literatur tersebut sejalan dengan hasil temuan empiris yang dilakukan di sekolah-sekolah di wilayah eks karesidenan Surakarta. Dari penyebaran angket diperoleh bahwa yang dianggap sebagai materi (pokok bahasan) tersulit adalah stoikiometri.

Setelah menetapkan pokok bahasan Stoikiometri, selanjutnya melaksanakan *focus group discussion* (FGD) dengan enam orang guru kimia SMA, tiga orang dosen kimia dan pendidikan kimia serta dua orang ahli pengukuran. FGD menghasilkan atribut dan hirarki atribut dilanjutkan dengan teknik Delphi yang melibatkan tiga orang pakar kimia dan pendidikan kimia. Teknik Delphi dilaksanakan dua putaran untuk memperoleh kesepakatan tentang atribut dan hirarki atribut dari Stoikiometri.

Penetapan atribut dan hirarki atribut digunakan untuk menyusun matriks A, matriks R dan matriks Qr sebagai dasar untuk menyusun kisi-kisi butir soal. Kisi-kisi butir soal yang disusun direview oleh enam orang guru yang sejak awal terlibat FGD. Setelah terpilih 20 butir soal maka dikembangkan soal-soal dengan bentuk OMC. Butir-butir soal ini direview dan divalidasi oleh enam orang guru kimia dan tiga orang pakar kimia dan pendidikan kimia. Selanjutnya dilakukan ujicoba kelompok terbatas pada 45 peserta didik, masing-masing 15 orang dari SMA kategori tinggi, sedang dan kurang. Setelah melalui revisi berdasarkan masukan peserta didik, selanjutnya dilakukan ujicoba menengah untuk karakterisasi parameter butir soal. Terakhir dilakukan uji feasibility pada 12 sekolah di wilayah Surakarta, Boyolali, Karanganyar dan Sragen.

Subjek uji coba dalam penelitian ini terbagi menjadi beberapa kategori:

- a. Pada uji coba terbatas yang mengungkap aspek keterbacaan dan ketercukupan waktu dengan menggunakan 15 siswa masing-masing dari sekolah dengan kategori baik, sedang dan kurang, yaitu SMA Negeri 4 Surakarta, SMA Negeri 7 Surakarta dan SMA Batik 2 Surakarta.
- b. Pada uji coba skala luas menggunakan 1 sekolah kategori baik, 1 sekolah kategori sedang dan 1 sekolah kategori kurang pada 4 wilayah Surakarta yaitu masing-masing 5 kelas dari SMA Negeri 4 Surakarta, SMA Negeri 7 Surakarta dan SMA Batik 2 Surakarta.
- c. Pada uji kelayakan menggunakan data yang berasal dari uji coba skala luas yaitu masing-masing 1 sekolah kategori baik, 1 sekolah kategori sedang dan 1 sekolah kategori kurang yang berasal dari Kota Surakarta, Kabupaten Karanganyar, Kabupaten Sragen dan Kabupaten Boyolali, yaitu, SMA Negeri 1 Sragen, SMA Negeri Gondang Sragen, SMA Negeri Sukodono Sragen, SMA Negeri 1 Boyolali, SMA Negeri 1 Simo Boyolali, SMA Negeri Nogosari Boyolali, SMA Negeri 2 Karanganyar dan SMA Negeri Gondang Rejo Karanganyar.

Pada studi pendahuluan data yang diperoleh dari angket terbuka dan wawancara pada guru dianalisis secara deskriptif untuk menemukan pokok bahasan yang banyak menimbulkan kesulitan peserta didik pada proses pembelajaran. Data pada uji coba terbatas dianalisis secara deskriptif untuk mencari kelemahan soal dan kemudian merevisinya.

Analisis data pada uji coba skala luas dilakukan untuk mendapatkan informasi validitas konstruk, reliabilitas instrumen, tingkat kesulitan butir, daya pembeda butir, kecocokan model, dan fungsi informasi. Analisis data secara luas menggunakan SmartPLS untuk menguji validitas konstruk dan reliabilitas. Untuk pengujian daya pembeda, tingkat kesukaran dan kecocokan model digunakan Parscale dan untuk menguji efektivitas digunakan Anava.

Hasil Penelitian dan Pembahasan

Langkah awal pada pengembangan instrumen ini adalah menetapkan atribut dan hierarki atribut melalui *focus group discussion* (FGD) yang melibatkan tiga orang dosen, enam orang guru kimia SMA dan dua orang pakar pengukuran. Hasil penetapan atribut dan hierarki atribut melalui FGD tersebut disempurnakan melalui teknik Delphi yang melibatkan tiga orang ahli. Teknik Delphi dilakukan sebanyak dua putaran untuk memperoleh kesepakatan tentang atribut dan hierarki atribut.

Hierarki disusun didasarkan pada *learning progression*. Apakah *learning progression* menjadi dasar dari asesmen diagnostik formal atau non formal pada level sempit (kelas) atau level besar (negara), pilihan tersebut berhubungan dengan model psikometri, kesalahan dalam memilih model akan berakibat fatal (Briggs, & Alonzo, 2009). Pengembangan butir-butir *Cognitive Diagnostic Assessment* (CDA) dipengaruhi oleh empat hal berdasar penelitian Gierl & Cui (2008, p.207), yaitu ukuran butir, dapat terukur, relevansi dengan proses pembelajaran, dan keterampilan yang berjenjang.

Tabel 1. Atribut, Nama Atribut dan Hirarki Atribut Stoikiometri

Atribut	Nama Atribut	Hirarki Atribut
A1	Simbol Atom	<pre> graph TD A1((A1)) --> A2((A2)) A1 --> A3((A3)) A1 --> A4((A4)) A2 --> A5((A5)) A5 --> A6((A6)) A5 --> A7((A7)) A7 --> A8((A8)) </pre>
A2	Massa atom relatif (Ar) dan massa molekul relatif / massa rumus relatif (Mr)	
A3	Reaksi kimia dan penyetaraannya	
A4	Hukum-hukum dasar Kimia I: Lavoisier, Dalton dan Proust	
A5	Konsep mol	
A6	Rumus empiris (RE) dan rumus molekul (RM)	
A7	Hukum dasar kimia II: Avogadro	
A8	Pereaksi pembatas	

Untuk menentukan hubungan antara atribut dengan hirarki yang menggunakan AHM, matriks *adjacency* dan *reachability* didefinisikan. Hubungan langsung antara atribut yang ditentukan oleh matriks adjacency biner (A) dari order (k, k), di mana k adalah jumlah atribut, seperti bahwa elemen ijth mewakili ketiadaan (yaitu 0) atau kehadiran (yaitu 1) dari sebuah hubungan langsung antara atribut Ai dengan Aj (Tatsuoka, 2009, p.37).

$$\begin{pmatrix}
 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
 \end{pmatrix}$$

Gambar 1. Matriks A

Setiap baris dan kolom pada matriks A menggambarkan sebuah atribut. Pada baris pertama menggambarkan atribut A1, terlihat bahwa pada kolom 2, 3, dan 4 bertanda angka 1 yang artinya bahwa atribut A1 memiliki hubungan langsung dengan atribut A2, A3, dan A4. Pada baris kedua

yang menggambarkan atribut A2, terlihat bahwa hanya memiliki hubungan langsung dengan atribut A5. Demikian seterusnya sampai dengan baris terakhir yaitu baris ke delapan yang menggambarkan atribut A8, tidak memiliki hubungan langsung dengan atribut manapun. Hubungan langsung (*direct connection*) ini bermakna atribut tersebut menjadi prasyarat bagi atribut yang lain.

$$\begin{pmatrix}
 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\
 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\
 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
 \end{pmatrix}$$

Gambar 2. Matriks R

Matriks R adalah gambaran hubungan langsung (*direct connection*) dan hubungan tak langsung (*indirect connection*). Seperti halnya matriks A setiap baris dan kolom pada matriks R juga menggambarkan sebuah atribut. Baris pertama menggambarkan atribut A1 dan baris ke delapan menggambarkan atribut A8. Terlihat pada baris pertama, bahwa semua kolom terisi angka 1, yang artinya

adalah bahwa atribut A1 menjadi memiliki hubungan langsung dan tak langsung dengan dirinya sendiri A1 dan dengan atribut A2, A3, A4, A5, A6, A7 dan A8. Sementara itu, pada baris kedua, pada kolom 2 terisi angka 1 dan kolom 5, 6, 7, dan 8 terisi angka 1 bermakna bahwa atribut 2 memiliki hubungan langsung (*direct connection*) dan tak langsung (*indirect connection*) dengan dirinya sendiri dan atribut no 5, 6, 7, 8.

Item pool direpresentasikan sebagai *incidence matrix* (Q) dengan ordo (k,p) di mana k adalah jumlah atribut dan p adalah jumlah kemungkinan butir. *Item pool* merupakan semua kombinasi dari atribut jika atribut-atribut independen satu sama lain. Dengan demikian, menyebabkan beberapa kemungkinan butir soal tidak sesuai dengan hirarki yang telah disusun sebagai model kognitif, untuk itu perlu dilakukan reduksi terhadap kemungkinan butir soal, sehingga terbentuk *reduced incidence matrix* (Qr) dengan menerapkan batasan hierarki atribut seperti yang didefinisikan oleh matriks R.

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Gambar 3. Matriks Qr

Matriks Qr adalah *blue print* dari butir-butir soal yang hendak dikembangkan dan hanya terbatas untuk menyatakan hubungan antara butir soal dengan atribut-atribut yang telah tersusun.

Melalui matriks Qr tersebut disusun kisi-kisi instrumen berupa indikator-indikator soal. Berdasarkan review guru diperoleh sebanyak 44 indikator soal, namun karena sebagian di antaranya saling tumpang tindih sehingga bias diwakili oleh indikator soal

yang lain, maka dipilih 20 butir indikator untuk ditulis soalnya dalam bentuk OMC.

Salah satu contoh butir soal OMC adalah sebagai berikut.

Gas hidrogen dapat dibuat dengan mereaksikan antara logam Aluminium ($A_r=27$) dengan larutan asam sulfat. Untuk membuat 6,72 liter gas hidrogen (0°C , 1 atm), maka di antara pernyataan berikut yang paling tepat adalah....

- Perbandingan mol Al : H_2SO_4 :
 $\text{Al}_2(\text{SO}_4)_3 : \text{H}_2 = 2 : 3 : 1 : 3$
- Persamaan reaksi : $2\text{Al}(s) + 3\text{H}_2\text{SO}_4(aq) \rightarrow \text{Al}_2(\text{SO}_4)_3(aq) + 3\text{H}_2(g)$
- Massa Al yang dibutuhkan pada reaksi tersebut = 5,4 gram
- Jumlah mol H_2 yang dihasilkan pada reaksi tersebut = 0,3 mol

Pada butir OMC tersebut, setiap pilihan adalah jawaban benar yang berasal dari urutan langkah mengerjakan. Langkah pertama untuk dapat menyelesaikan soal tersebut adalah apabila peserta didik dapat menyelesaikan persamaan reaksi (opsi B), selanjutnya adalah mengubah volume gas hidrogen menjadi mol (opsi D). Kemudian dengan membuat perbandingan mol (opsi A) dan akhirnya dapat menghitung massa aluminium (opsi C). Tingkat kesulitan pada langkah-langkah tersebut tidak berurutan, artinya option C yang menjadi level tertinggi pada penyelesaian soal tersebut justru adalah langkah paling mudah, yaitu hanya dengan mengkonversikan mol yang diperoleh pada level 3 (opsi A) ke dalam massa aluminium dengan mengalikan mol dan massa atom relative (A_r Al). Diagnosis dilakukan dengan memperhatikan pilihan jawaban peserta didik. Pilihan jawaban tersebut menunjukkan kemampuan atau level yang dimiliki oleh peserta didik.

Untuk membuktikan validitas isi, maka sebanyak enam orang guru dan tiga orang ahli memberikan penilaian dan masukan terhadap butir-butir soal yang telah disusun. Dengan jumlah rater sebanyak 9 orang maka sebuah item dianggap valid secara isi manakala memiliki V sebesar 0.78 atau lebih (Aiken, 1980, p.134). Seluruh

butir soal dikatakan memiliki validitas isi karena memiliki harga V lebih dari 0,78. Butir-butir soal yang telah dibuktikan validitas isinya kemudian diuji keterbacaan melalui uji coba terbatas oleh 45 peserta didik, yang terdiri dari 15 peserta didik sekolah kategori tinggi, sedang dan rendah.

Berdasarkan uji *multivariate normality distribution* menggunakan Lisrel data penelitian ini tidak berdistribusi normal yaitu tidak memenuhi asumsi *multivariate normality* karena nilai p value skewness dan kurtosis < 0,05 maka digunakan SmartPLS untuk menganalisis validitas konstruk instrumen ini. Validitas kontrak dari instrumen dapat ditunjukkan melalui *Discriminant Validity*. Pada penentuan *discriminant validity* dari model pengukuran reflektif, indikator dinilai berdasarkan nilai-nilai *cross loading* dengan konstruksya. Jika korelasi konstruk dengan butir pengukuran lebih besar daripada ukuran konstruk lainnya, maka hal ini menunjukkan bahwa konstruk laten memprediksi ukuran pada blok mereka lebih baik daripada ukuran pada blok lainnya dengan demikian dikatakan memiliki *discriminant validity* yang baik (Ghozali, 2008, p.24).

Harga *Goodness of Fit (GoF) Index* instrumen pada SmartPLS ditentukan berdasarkan formula:

$$GoF = \sqrt{R^2 \times Com}$$

(Tenenhaus, M., et all, 2005:159-205; Vinzi, et al, 2010, p.58)

Berdasarkan perhitungan diperoleh bahwa nilai-nilai GoF pada masing-masing paket A= 0,437; paket B = 0,466 dan paket C = 0,433. Berdasarkan nilai-nilai GoF tersebut, maka dapat dikatakan bahwa masing-masing paket memiliki harga GoF yang tinggi yaitu > 0,36 sehingga memenuhi persyaratan sebagai instrumen yang baik (cocok model).

Di samping pembuktian validitas konstruk, perlu juga dilakukan uji reliabilitas konstruk yang dapat diukur menggunakan 2 (dua) macam cara yaitu *internal consistency* dengan menerapkan formula *composite reliabi-*

lity dari Werts, Linn dan Joreskog (Ghozali, 2008, p.25) dan *Cronbach Alpha*. Jika dibandingkan dengan *composite reliability*, *Cronbach Alpha* cenderung *lower bound estimate reliability*, sedangkan formula *composite reliability* merupakan *closer approximation* dengan asumsi estimasi parameter yang akurat. Untuk itu, penelitian ini menggunakan formula *composite reliability* untuk menentukan reliabilitas konstruk dengan menggunakan SmartPLS.

Pada paket soal A, nilai *composite reliability* untuk masing-masing faktor adalah 0,67; 0,80; 1,00; 0,67; 0,78; 0,79; 0,62; 0,77 dan secara keseluruhan adalah 0,79. Dengan demikian, maka seluruh faktor dikatakan memiliki *composite reliability* yang tinggi. Nilai *composite reliability* pada paket soal B berturut-turut adalah 0,77; 0,73; 1,00; 0,70; 0,82; 0,31; 0,80; 0,74 dan secara keseluruhan adalah 0,81. Pada paket B ini terlihat bahwa pada faktor 6 memiliki *composite reliability* yang rendah (< 0,6) hal ini disebabkan pada faktor 6 terdapat sebuah soal yaitu pada nomor 11 memiliki loading faktor sangat rendah (-0,361). Akan tetapi, secara keseluruhan, nilai *composite reliability* termasuk kategori tinggi yaitu 0,81. Paket soal C nilai *composite reliability* setiap faktor berturut-turut adalah 0,79; 0,76; 1,00; 0,32; 0,78; 0,75; 0,69; 0,75 dan secara keseluruhan adalah 0,75. Pada paket C ini terlihat bahwa pada faktor 4 memiliki *composite reliability* yang rendah (< 0,6) hal ini disebabkan pada faktor 4 terdapat dua butir soal yaitu pada nomor 12 dan 13 yang setiap soal memiliki loading faktor sangat rendah yaitu -0,006 dan -0,161. Namun secara keseluruhan, nilai *composite reliability* termasuk kategori tinggi yaitu 0,75.

Fit statistics juga berguna dalam mengevaluasi model-model untuk data yang memiliki respons sama jika data-data tersebut dikelompokkan berdasarkan parameternya (Mathilda, 2003, p.704). Untuk melihat *butir fit statistics* digunakan analisis dengan Parscale dengan ketentuan bahwa butir soal dikatakan fit jika memiliki nilai prob > 0,05. Berdasarkan analisis *butir fit statistics* dari Parscale tersebut dapat dilihat bahwa pada paket soal A terdapat butir-butir yang tidak

fit yaitu pada nomor 7, 10, 12, 17 dan 18; sedangkan untuk paket soal B terdapat butir-butir soal yang dinyatakan tidak fit yaitu pada nomor 7, 12, 13, dan 18. Adapun pada paket soal C terdapat butir-butir soal tidak fit yaitu pada nomor 7, 10, 12, 17, dan 18. Dengan demikian terdapat nomor-nomor yang pada ketiga paket A, B, dan C dinyatakan tidak fit yaitu pada nomor 7, 12, dan 18.

Adapun tingkat kesukaran pada butir-butir soal dilihat dari parameter b1, b2 dan b3. Parameter b1 menunjukkan tingkat kesulitan antara option level 1 dan level 2, parameter b2 menunjukkan tingkat kesulitan antara option level 2 dan level 3, parameter b3 menunjukkan tingkat kesulitan antara option level 3 dan level 4. Semakin tinggi level, maka tingkat kesukaran akan semakin tinggi. Parameter tingkat kesukaran pada semua paket A, B, dan C memiliki nilai $b1 < b2 < b3$. Dengan demikian, terbukti bahwa pada paket-paket soal semakin tinggi level option pemahaman semakin sulit.

Daya pembeda butir soal dilihat dari parameter a yang pada seluruh butir soal pada paket A, B, dan C memiliki nilai positif, kecuali pada 7 dan 15. Option b dan c pada soal no 7 memiliki level kemampuan yang sama, yaitu pada level 2 sehingga pada butir soal nomor 7, level yang dijelaskan adalah 1, 2, 2, 3.

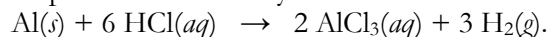
Butir soal nomor 7 yaitu pada faktor 3 dengan atribut persamaan reaksi, terdapat 2 option yang memiliki level sama option b dan c, pada butir soal berikut.

Logam aluminium direaksikan dengan larutan asam klorida encer dan membentuk larutan aluminium klorida dan gas hidrogen. Maka...

- rumus senyawa aluminium klorida adalah $AlCl_3$.
- larutan aluminium klorida dan gas hidrogen sebagai produk.

c. logam aluminium dan larutan asam klorida encer sebagai pereaksi.

d. persamaan reaksinya adalah:



Option b dan c terdapat pada butir soal tersebut. Hal tersebut dimungkinkan menjadi penyebab butir soal nomor 7 memiliki daya pembeda yang rendah yaitu 0,169. Butir soal 15 juga memiliki daya pembeda rendah, hal ini dimungkinkan karena butir soal nomor 15 tersebut juga memiliki *loading factor* yang rendah pada uji validitas konstruk.

Instrumen ini dikemas dalam bentuk *paper and pencil* mengingat kondisi di lapangan yang terbatas kepemilikan sarana komputer dalam jumlah yang memadai. Analisis hasil pekerjaan peserta didik diolah dengan menggunakan bantuan program komputer berbasis Excel. Penggunaan program analisis data dimaksudkan untuk membantu guru lebih cepat dalam menganalisis hasil pekerjaan peserta didik dan menghasilkan output berupa profil diagnostik peserta didik. Program komputer yang dipergunakan untuk analisis hasil pekerjaan peserta didik dibuat dalam bentuk Excel yang sudah cukup familiar di kalangan guru, hal ini dimaksudkan agar guru mudah dalam mengoperasikan.

Uji efektivitas instrumen menggunakan Anova dengan menggunakan SPSS menghasilkan uji beda seluruhnya baik pada paket A, B, dan C berbeda secara signifikan ($P < 0,05$). Artinya instrumen dapat membedakan secara nyata antara peserta didik yang berasal dari sekolah kategori tinggi, sedang dan rendah. Pada pengujian lanjut menggunakan *Scheffe* dan *Bonferroni* untuk paket A, B, dan C terbukti seluruhnya signifikan ($P < 0,05$). Artinya, bahwa kelompok peserta didik kemampuan tinggi > kelompok peserta didik kemampuan sedang > kelompok peserta didik kemampuan rendah.

Nama Siswa: XXX			
No. Soal	SKOR	KEMAMPUAN SUDAH DIKUASAI	KEMAMPUAN BELUM DIKUASAI
1	3	Membentuk dan memberi nama senyawa yang dibentuk dari ion positif bermuatan > 1 dan ion negatif bermuatan 1; membentuk dan memberi nama senyawa yang dibentuk dari ion positif bermuatan 1 dan ion negatif bermuatan > 1 ; membentuk dan memberi nama senyawa yang berasal dari ion-ion dengan muatan positif sama dengan muatan negatif.	Membentuk dan memberi nama senyawa yang berasal dari ion-ion dengan muatan positif dan muatan ion negatif masing-masing > 1 dan berbeda
2	3	Menentukan jumlah atom unsur-unsur yang menyusun sebuah molekul senyawa; memberikan nama suatu molekul senyawa yang diketahui rumus strukturnya; menentukan unsur-unsur penyusun suatu molekul senyawa	Menghitung jumlah atom yang menyusun sejumlah molekul senyawa

Gambar 4. Contoh *Output* Profil Diagnostik

Simpulan dan Saran

Berdasarkan hasil penelitian pengembangan ini dapat dibuat simpulan sebagai berikut.

1. Penelitian ini telah berhasil mengembangkan 3 (tiga) paket soal *Ordered Multiple Choice* (OMC) untuk mendeteksi kesulitan belajar pada pokok bahasan Stoikiometri kelas X. Setiap butir soal OMC terdiri dari pernyataan dan 4 option berisi jawaban benar dengan level pemahaman berjenjang.
2. Butir-butir soal OMC pada paket A, B, dan C yang dikembangkan seluruhnya memiliki indeks validitas isi Aiken lebih dari 0,75 dengan demikian seluruh butir soal memiliki validitas isi yang baik. Paket A, B, dan C memiliki memiliki *Goodness of Fit* (GoF) lebih besar dari 0,36 yaitu 0,437; 0,466 dan 0,433 berdasarkan Smart PLS sehingga memiliki validitas konstruk yang baik. Instrumen memiliki reliabilitas tinggi berdasarkan SmartPLS berturut-turut 0,79; 0,81 dan 0,75.
3. Profil individu peserta didik berfungsi sebagai laporan diagnostik memuat atribut yang dikuasai dan belum dikuasai oleh peserta didik.

Saran

Beberapa hal yang dapat diajukan sebagai saran berdasarkan penelitian ini adalah sebagai berikut.

1. Instrumen *Ordered Multiple Choice* (OMC) yang telah dikembangkan ini lebih tepat dipergunakan bagi peserta didik yang berasal dari sekolah dengan kategori kurang, namun memerlukan petunjuk pengerjaan yang lebih mudah dipahami oleh peserta didik.
2. Sekolah-sekolah yang telah memiliki perangkat komputer yang cukup untuk sejumlah peserta didik, maka akan lebih memudahkan jika perangkat instrumen dikembangkan dalam bentuk *Computer Based Test* (CBT).

Daftar Pustaka

- Aiken, L.R. (1985). Three Coefficient for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45
- Barke, H.D, Al Hazari & Yitbarek,S (2009). *Misconceptions in Chemistry, Addressing Perceptions in Chemical Education*. German: Springer-Verlag Berlin Heidelberg.

- Bennet, J. (2007). *Teaching and Learning Science*. London: Biddles Ltd., King's Lynn. Norfolk
- Borg, W.R & Gall, M.D.(1983). *Educational Research: An Introduction*, Fourth edition. New York: Longman, Inc
- Briggs, D.C., Alonzo, A.C., Schwab, C., & Wilson, M. (2006). Diagnostik Assessment With Ordered Multiple-Choice Items. *Educational Assessment*, 11(1), 33-63
- Briggs, D.C & Alonzo, A.C (2009). The Psychometric Modeling of Ordered Multiple-Choice Item Responses for Diagnostik Assessment with a Learning Progression. *Paper presented at The Learning Progression in Science (LeaPS) Conference, June 2009, Iowa City*.
- Burton, W.H.(1952). *The Guidance of Learning Activities*. New York: Apleton-Century-Crofts, Inc
- Chandrasegaran, A.L., Treagust, D.F., & Mocerino, M (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, 2007, 8 (3), 293-307.
- Depdiknas. 2003. *Kurikulum Berbasis Kompetensi Mata Pelajaran Kimia SMA*. Jakarta : Puskur Depdiknas
- Ghozali, I. (2008). *Structural Equation Modeling: Metode Alternatif dengan Partial Least Square (PLS)*. Badan Penerbit Universitas Diponegoro: Semarang
- Gierl, M.J (2007). Making Diagnostic Inferences About Cognitive Attributes Using the Rule-Space Model and Attribute Hierarchy Method. *Journal of Educational Measurement Winter 2007, Vol. 44, No. 4, pp. 325-340*
- Gierl, M. J., Leighton, J. P., & Hunka, S. (2000). Exploring the logic of Tatsuo-oka's rule-space model for test development and analysis. *Educational Measurement: Issues and Practice*, 19, 34-44
- Gierl, M.J., Wang, C., & Zhou, J. (2008). Using the Attribute Hierarchy Method to Make Diagnostic Inferences about Examinees' Cognitive Skills in Algebra on the SAT©. *Journal of Technology, Learning, and Assessment*, 6(6). <http://www.jtla.org>.
- Gierl, M.J & Cui, Y. (2008). Defining Characteristics of Diagnostic Classification Models and the Problem of Retrofitting in Cognitive Diagnostic Assessment. *Measurement*, 6 : 263-27.
- Gierl, M.J., Cui, Y & Zhou, J (2009). Reliability and Attribute Based Scoring in Cognitive Diagnostic Assessment. *Journal of Educational Measurement Fall 2009, Vol 46, No. 3, pp. 293-313*
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J (1991). *Fundamental of Item Response Theory*. Sage Publications
- Leighton, J.P & Gierl, M.J (2011). *The Learning Sciences in Educational Assessment: The Role of Cognitive Models*. Cambridge University Press 32 Avenue of the Americas, New York, NY 10013-2473, USA
- Leighton, J.P & Gierl, M.J (2007). *Cognitive Diagnostik Assessment: theory and applications*. Cambridge University Press
- Leighton, J.P, Gierl, M.J & Hunka, S (2004). The Attribute Hierarchy Method for Cognitive Assessment: A Variation on Tatsuo-oka's Rule-Space Approach. *Journal of Educational Measurement Fall 2004, Vol, 41, No. 3, pp. 205-237*
- Mc Glohen, M.K. (2004). The application of a cognitive diagnosis model via an analysis of a large-scale assessment and a computerized adaptive testing administration. *Disertasi. The university*

- of Texas at Austin*. Diakses tanggal 10 Juli 2010, dari scholar.google.com
- Nichols, P.D., Chipman, S.F., & Brennan, R.L. (1995). *Cognitively Diagnostic Assessment*. Lawrence Erlbaum Associate Publishers. Hillsdale. New Jersey
- Okanlawon, A. E. (2010) Constructing a Framework for Teaching Reaction Stoichiometry Using Pedagogical Content Knowledge. *Chemistry*, 19(2), 27-44
- Pritchard, A & Woolard, J (2010). *Psychology for the Classroom: Constructivism and Social Learning*. Routledge 2 Park Square, Milton Park, Abingdon, Oxon.
- Rahayu, S., & Kita, M., (2009). An Analysis Of Indonesian And Japanese Students' Understandings Of Macroscopic And Submicroscopic Levels Of Representing Matter And Its Changes. *International Journal of Science and Mathematics Education* (2010) 8: 667- 688
- Tatsuoka, K.K., (2009). *Cognitive Assessment: An Introduction to the Rule Space Method*. Routledge. Taylor & Francis Group. New York
- Tenenhaus, M., et al (2005). PLS Path Modelling. *Computational Statistics & Data Analysis* 48, 159-205
- Van der Linden, W.J & Hambleton, R.K (1997). Item response theory: Brief history, common models, and extensions. Dalam W.J. Van der Linden & R.K. Hambleton (Eds). *Handbook of Modern Item Response Theory* (pp. 1-28). New York: Springer-Verlag Inc
- Vinci, V.E., et al. (2010). *Handbook of Partial Least Square: Concept, Methods and Applications*. Springer-Verlag Berlin Heidelberg: Germany
- Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24, 535-585.

PENGEMBANGAN MODEL EVALUASI PROSES PEMBELAJARAN MATEMATIKA DI SMP BERDASARKAN KURIKULUM 2013

¹⁾Sugiyanto, ²⁾Badrun Kartowagiran, ³⁾Jailani

¹⁾LPMP Provinsi Kalimantan Tengah, ^{2, 3)}Universitas Negeri Yogyakarta

¹⁾sugi_kalteng58@yahoo.com, ²⁾badrunkw@yahoo.com, ³⁾jailani_uny@yahoo.com

Abstrak

Penelitian ini bertujuan mengembangkan model evaluasi proses pembelajaran matematika di SMP berdasarkan Kurikulum 2013, dan menilai keefektifan model yang dikembangkan. Penelitian ini menggunakan metode penelitian dan pengembangan. Subjek coba adalah guru, kepala sekolah, dan pengawas. Instrumen pengumpul data adalah kuesioner dan lembar observasi. Proses pengembangan dilakukan dalam 4 tahap yaitu: (1) investigasi awal; (2) perencanaan pengembangan, validasi; (3) uji coba, dan (4) finalisasi produk. Uji coba dilakukan tiga tahap dengan jumlah subjek meningkat setiap tahapnya. Instrumen evaluasi terdiri instrumen penilaian perencanaan pembelajaran, pelaksanaan pembelajaran, dan pelaksanaan penilaian. Validitas isi dilakukan melalui *focus group discussion* (FGD). Validitas konstruk dilakukan dengan menggunakan analisis faktor, dan estimasi reliabilitasnya menggunakan *Intraclass Correlation Coefficient* (ICC). Hasil penelitian berupa model evaluasi proses pembelajaran matematika di SMP, yaitu prosedur evaluasi, panduan evaluasi, dan instrumen evaluasi. Berdasarkan hasil uji coba, semua instrumen memiliki kecocokan model yang baik, karena validitas konstruk, dan reliabilitas memenuhi persyaratan akademik. Penilaian para praktisi dan pengguna, model evaluasi proses pembelajaran matematika sangat efektif untuk diterapkan.

Kata kunci: *model evaluasi, proses pembelajaran, matematika SMP*

DEVELOPING A MODEL FOR AN EVALUATION OF MATHEMATICS LEARNING PROCESSES AT JUNIOR HIGH SCHOOLS BASED ON CURRICULUM 2013

¹⁾Sugiyanto, ²⁾Badrun Kartowagiran, ³⁾Jailani

¹⁾LPMP Provinsi Kalimantan Tengah, ^{2, 3)}Universitas Negeri Yogyakarta

¹⁾sugi_kalteng58@yahoo.com, ²⁾badrunkw@yahoo.com, ³⁾jailani_uny@yahoo.com

Abstract

This study aims to develop a model for an evaluation of mathematics learning processes at junior high schools (JHSs) based on Curriculum 2013 and evaluate the effectiveness of the model developed. This was a research and development study. The subject of the research was the teachers, principals, and supervisors. The data collecting instruments were questionnaires and observational sheet. The development process was carried out in 4 stages namely: (1) preliminary investigation; (2) planning development, validation; (3) tryout; (4) product finalization. The try out was performed in 3 stages with the increasing number of participants for each stage. The learning process evaluation model was supported by instrument to assess lesson plans, learning implementation, and the implementation of learning outcome assessment. The content validity was assessed through focus group discussion (FGD). The construct validity was assessed by the factor analysis, and the reliabilities were estimated by the *Intraclass Correlation Coefficient* (ICC). The result of the research was a model of evaluation of mathematics learning processes at junior high schools (JHSs) based on Curriculum 2013, consisting of procedures, evaluation guideline and the well designed instruments. Based on the try out, all of the instruments have had good suitability, because construct validity and reliability meet all of the academic requirements. According to the evaluation from the practitioners and the model users, the model of the evaluation of mathematics learning processes at junior high schools (JHSs) based on Curriculum 2013 was very effective to be implemented.

Keywords: *development model, evaluation of learning processes, SMP mathematics*

Pendahuluan

Pembelajaran matematika berdasarkan Kurikulum 2013 mulai dilaksanakan pada tahun pelajaran 2013/2014 di sekolah, dari jenjang pendidikan dasar sampai dengan jenjang pendidikan menengah di seluruh Indonesia. Adapun kompetensi lulusannya menekankan pada ranah sikap, pengetahuan, dan keterampilan yang pelaksanaan dan kriterianya telah diatur dalam Peraturan Menteri Pendidikan dan Kebudayaan (Permendikbud) RI No.54 Tahun 2013 tentang Standar Kompetensi Lulusan dan Permendikbud No. 65 Tahun 2013 tentang Standar Proses. Untuk dapat mewujudkan ketercapaian ketiga ranah tersebut, setiap sekolah pada semua jenjang pendidikan dalam pelaksanaan kegiatan pembelajarannya harus mengacu pada standar yang telah ditetapkan, baik pembelajaran yang dilakukan di kelas maupun pembelajaran yang dilakukan di luar kelas dalam lingkungan sekolah.

Pengertian pembelajaran menurut Nitko & Brookhart (2007, p.18), merupakan proses yang digunakan untuk mengarahkan peserta didik dengan kondisi membantu mereka mencapai target belajar. Hal senada disampaikan oleh Suherman (2003, p.7) bahwa pembelajaran merupakan upaya penataan lingkungan yang memberi nuansa agar program belajar tumbuh dan berkembang secara optimal. Martiyono (2012, p.12) menambahkan bahwa pembelajaran merupakan kegiatan melaksanakan rencana pembelajaran untuk mencapai tujuan, sehingga proses pembelajaran berjalan secara baik dan mencapai hasil yang optimal. Dalam pembelajaran, guru memiliki peran penting dalam mengkondisikan lingkungan agar menunjang terjadinya perubahan perilaku bagi peserta didik (Darmadi, 2010, pp.177-178). Kegiatan pembelajaran yang dilakukan di sekolah, menurut Permendikbud No.65 tentang Standar Proses, diawali dengan kegiatan penyusunan perencanaan pembelajaran, pelaksanaan proses pembelajaran, dan selanjutnya dilakukan penilaian hasil belajar.

Perencanaan pembelajaran dirancang dalam bentuk Silabus dan Rencana Pelaksanaan Pembelajaran (RPP). Silabus meru-

pakan acuan penyusunan RPP. Rencana pelaksanaan pembelajaran dikembangkan dari silabus untuk mengarahkan kegiatan pembelajaran peserta didik dalam upaya mencapai Kompetensi Dasar (KD). Perencanaan pembelajaran menurut Martiyono (2012, p.22) adalah kegiatan merencanakan semua komponen pembelajaran dan langkah-langkahnya yang akan dilaksanakan untuk mencapai tujuan, sehingga proses pembelajaran berjalan secara baik dan mencapai hasil yang optimal. Hal ini juga dinyatakan oleh Uno (2008, p.2) yang menyebutkan bahwa, perencanaan adalah cara membuat kegiatan berjalan dengan baik, disertai berbagai langkah yang antisipatif guna memperkecil kesenjangan yang terjadi sehingga kegiatan dapat mencapai tujuan yang telah ditetapkan. Perencanaan pembelajaran dirancang dalam bentuk silabus RPP, yang dilengkapi dengan pemanfaatan media dan bahan ajar. Menurut Majid (2008, p.38) silabus adalah rancangan pembelajaran yang berisi rencana bahan ajar mata pelajaran tertentu. Dijelaskan pula oleh Darling, (2010, p.11), bahwa silabus adalah titik penting dari interaksi antara guru dengan peserta didik baik di dalam ataupun di luar kelas. Silabus perlu disusun agar hasil pembelajaran sesuai dengan keluaran yang diharapkan. Daley (2010, p.166), mengatakan, silabus perlu diorganisasi sehingga setiap hasil belajar memiliki hasil secara pasti. Silabus adalah penjabaran SK dan KD ke dalam materi pokok, kegiatan pembelajaran, indikator, penilaian, alokasi waktu, sumber belajar. Hal senada disampaikan oleh Rusman (2012, p.6) bahwa silabus adalah rencana pembelajaran pada suatu kelompok mata pelajaran tertentu yang mencakup SK, KD, materi pokok, pembelajaran, indikator, penilaian, alokasi waktu, dan belajar.

Proses pembelajaran merupakan implementasi dari rencana pembelajaran yang meliputi kegiatan pendahuluan, inti dan penutup. Kegiatan pendahuluan meliputi penyiapan peserta didik oleh guru secara psikis dan fisik untuk mengikuti pembelajaran, memberi motivasi, mengajukan pertanyaan berkaitan pengetahuan sebelumnya dengan

materi yang akan dipelajari, menjelaskan tujuan, dan menyampaikan cakupan materi. Pada kegiatan inti, guru menggunakan model dan metode, media, sumber belajar sesuai dengan karakteristik peserta didik dan mata pelajaran. Pada kegiatan penutup, dilakukan refleksi untuk mengevaluasi seluruh rangkaian aktivitas pembelajaran dan hasil-hasil yang diperoleh dan menemukan manfaat dari hasil pembelajaran. Untuk mengetahui perolehan hasil belajar, guru melakukan penilaian hasil belajar peserta didik.

Penilaian hasil belajar peserta didik diatur melalui Kemendikbud Nomor 66 Tahun 2013 yaitu mencakup kompetensi sikap, pengetahuan dan keterampilan. Penilaian sikap dilakukan melalui observasi, penilaian diri, penilaian teman sejawat dan jurnal. Instrumen yang digunakan untuk observasi, penilaian diri, dan penilaian antar-peserta didik adalah daftar cek atau skala penilaian yang disertai rubrik, sedangkan penilaian melalui jurnal berupa catatan dari pendidik. Penilaian kompetensi pengetahuan dilakukan pendidik melalui tes tulis, tes lisan, dan penugasan. Penilaian kompetensi keterampilan, dilakukan pendidik melalui penilaian kinerja, yaitu penilaian yang menuntut peserta didik mendemonstrasikan suatu kompetensi tertentu dengan menggunakan tes praktik, proyek, dan penilaian portofolio.

Dalam Permendikbud No. 68 Tahun 2013 disebutkan bahwa matematika merupakan salah satu mata pelajaran yang wajib diberikan di SMP. Proses pembelajaran matematika di SMP berdasarkan Kurikulum 2013 menggunakan pendekatan saintifik sebagaimana tertuang pada Lampiran IV Permendikbud Nomor 81A, yaitu melalui kegiatan: (1) mengamati; (2) menanya; (3) mengumpulkan informasi; (4) mengasosiasi; dan (5) mengkomunikasikan. Kompetensi yang dikembangkan adalah melatih kesungguhan, ketelitian, dan mencari informasi. Kegiatan penilaian mengacu pada Permendikbud No. 66 Tahun 2013 tentang Standar Penilaian, yang mencakup penilaian sikap, penilaian pengetahuan dan penilaian keterampilan. Untuk dapat mengetahui efektivitas

pelaksanaan proses pembelajaran matematika di setiap satuan pendidikan secara komprehensif, perlu dilakukan kegiatan evaluasi.

Seiring dengan pelaksanaan Kurikulum 2013 di sekolah, peneliti melakukan observasi ke sekolah SMP di Provinsi Kalimantan Tengah yang telah melaksanakan Kurikulum 2013. Observasi dilakukan di 4 sekolah SMP yang berada di Kota Palangkaraya, yaitu SMP Negeri 1 Palangkaraya, SMP Negeri 2 Palangkaraya, SMP Negeri 3 Palangkaraya dan SMP Nahdatul Ulama Palangkaraya. Tujuannya adalah untuk memperoleh informasi berkaitan dengan pelaksanaan evaluasi pembelajaran matematika di SMP berdasarkan Kurikulum 2013. Observasi dilakukan pada pelaksanaan pembelajaran yang meliputi kegiatan perencanaan pembelajaran, kegiatan proses pembelajaran, dan kegiatan pelaksanaan penilaian hasil belajar.

Kegiatan observasi perencanaan pembelajaran dilakukan melalui studi dokumentasi terkait dengan keberadaan silabus, RPP, dan bahan ajar. Kegiatan observasi pada pelaksanaan proses pembelajaran dilakukan dengan mengamati secara langsung pelaksanaan pembelajaran yang dilakukan guru di kelas. Pengamatan difokuskan pada penataan tempat duduk peserta didik yang memungkinkan peserta didik untuk melakukan diskusi kelas, kegiatan pendahuluan, kegiatan inti, dan kegiatan penutup di kelas. Kegiatan observasi pada pelaksanaan pengolahan dan pelaporan hasil belajar peserta didik dilakukan melalui wawancara dan studi dokumentasi. Kegiatan difokuskan pada adanya dokumen perencanaan penilaian hasil belajar dan buku laporan hasil penilaian yang dibuat oleh guru. Pelaksanaan penilaian dilakukan melalui pengamatan pada akhir pembelajaran setelah satu bahasan selesai dilaksanakan pembelajaran.

Berdasarkan hasil observasi diperoleh data sebagai berikut. (1) Sekolah belum melakukan evaluasi proses pembelajaran matematika sesuai Kurikulum 2013, disebabkan karena sekolah belum memiliki instrumen dan panduan pelaksanaan evaluasi proses

pembelajaran. (2) Silabus yang seyogyanya telah dipersiapkan oleh pemerintah, tetapi pada saat dilakukan pengamatan, silabus masih dalam bentuk draf yang tingkat keterbacaannya masih sulit dipahami oleh guru. (3) Rencana pelaksanaan pembelajaran yang dipersiapkan guru masih menggunakan contoh yang diperoleh pada waktu guru mengikuti pelatihan implementasi kurikulum. (4) Bahan ajar belum disusun oleh guru karena guru masih mengalami kesulitan dalam melakukan telaah buku paket matematika yang telah disediakan oleh pemerintah. (5) Penataan ruang kelas belum mengarah pada pembelajaran dengan pendekatan saintifik. (6) Guru belum memiliki perangkat perencanaan penilaian berdasarkan Kurikulum 2013, hal ini disebabkan karena guru belum memahami komponen-komponen yang harus dipersiapkan dalam perencanaan penilaian hasil belajar peserta didik, belum memahami cara melakukan penilaian serta belum memahami cara pengolahan dan pelaporan hasil belajar peserta didik. Hal ini senada dengan penelitian Kartowagiran, dkk (2013), bahwa kepala SMP di DIY siap melaksanakan implementasi Kurikulum 2013, namun guru belum siap melaksanakan implementasi Kurikulum 2013.

Selanjutnya, melalui diskusi dengan pengawas sekolah, diperoleh informasi bahwa: (1) aktivitas pembelajaran belum teramati secara optimal; (2) fungsi kepengawasan masih menekankan pada kelengkapan administrasi yang dimiliki guru; (3) sekolah belum memiliki instrumen evaluasi pelaksanaan pembelajaran berdasarkan Kurikulum 2013; (4) sekolah belum memiliki model laporan evaluasi pembelajaran menurut Kurikulum 2013.

Berkaitan dengan kegiatan evaluasi, Kaufman & Thomas (1980, p.4) menyatakan, evaluasi adalah suatu proses untuk menilai kualitas sesuatu yang sedang berlangsung. Menurut Gronlund & Linn (1985, p.5), evaluasi adalah proses yang sistematis yang bertujuan mengumpulkan informasi, diinterpretasi guna mengetahui tingkat keberhasilan sasaran. Fitzpatrick, Sander & Worthen (2011, p.7) menyatakan Salah satu

kegiatan evaluasi yang terpenting adalah membantu *stakeholder* mengambil keputusan yang telah dikaji kekurangan dan kelebihan-nya dalam mengambil keputusan. Proses pelaksanaan evaluasi membutuhkan dukungan dan kerja sama dari semua pihak pendidikan (Hood & Hopson (2008, p.416). Stufflebeam, Madaus, & Kellaghan (2002, pp.19-20) menyatakan model adalah suatu rancangan yang dibuat untuk melatih atau membentuk seseorang untuk memiliki suatu cara atau gaya hidup tertentu. Model evaluasi menurut Stufflebeam, Madaus, & Kellaghan (2000, pp.119-120) adalah model yang menggambarkan pandangan pembuat model konsep dan struktur dalam melakukan evaluasi, sekaligus menyediakan pedoman evaluasi menggunakan model yang dirancang untuk mendeskripsikan, memberi penilaian dan rekomendasi. Sehubungan hal tersebut, penelitian ini mengembangkan sebuah model evaluasi pembelajaran matematika di SMP berdasarkan Kurikulum 2013.

Model evaluasi yang dikembangkan meliputi prosedur pelaksanaan evaluasi, instrumen evaluasi pembelajaran, dan panduan pelaksanaan evaluasi. Model evaluasi proses pembelajaran ini diharapkan dapat memberikan informasi secara komprehensif dan representatif tentang pelaksanaan pembelajaran matematika yang dilakukan guru di sekolah. Model evaluasi yang dikembangkan mengacu pada Permendikbud No.65 tentang Standar Proses, Permendikbud No. 66 tentang Standar Penilaian, dan Permendikbud No.81A tentang Implementasi Kurikulum.

Metode Penelitian

Penelitian ini merupakan penelitian pengembangan (*research and development*) bertujuan untuk menghasilkan suatu produk berupa sebuah model evaluasi proses pembelajaran matematika di SMP berdasarkan Kurikulum 2013. Model penelitian yang digunakan mengacu pada desain model yang dikembangkan oleh Borg & Gall (1983, p.775) dengan 10 langkah, yang disederhanakan peneliti menjadi 4 langkah/tahapan yaitu: (1) tahap studi pendahuluan, (2) tahap

perencanaan pengembangan, (3) tahap uji coba, evaluasi, revisi, dan (4) tahap implementasi.

Pada tahap awal dilakukan studi pendahuluan, yaitu untuk memperoleh data berkaitan dengan pelaksanaan evaluasi pembelajaran di sekolah. Selanjutnya, dilakukan kajian teori berkaitan pembelajaran matematika di SMP, mengkaji model-model evaluasi, dan mengkaji hasil penelitian terdahulu yang relevan. Data dari hasil studi pendahuluan dijadikan acuan untuk menyusun perencanaan pengembangan. Pada tahap perencanaan, pengembangan disusun desain model evaluasi proses pembelajaran yang terdiri dari prosedur evaluasi, komponen evaluasi, instrumen, panduan, dan desain uji cobanya. Selanjutnya, semua hasil perencanaan dikaji keterbacaan oleh pakar dan praktisi melalui *focus group discussion* (FGD). Pelaksanaan FGD dilakukan oleh 7 orang pakar (dosen) dan 20 orang widyaiswara LPMP. Uji keterbacaan dilakukan oleh 12 orang guru, 5 orang kepala sekolah SMP, 20 orang pengawas. Pada tahap uji coba, evaluasi, dan revisi dilakukan uji coba model, untuk mengetahui sejauhmana instrumen dan panduan dapat mengevaluasi pelaksanaan pembelajaran. Data hasil uji coba dianalisis apakah model sudah baik atau masih perlu diperbaiki. Setelah Instrumen dan panduan sudah baik selanjutnya diimplementasikan ke sekolah. Implementasi dilakukan pada bulan Maret-April 2014. Implementasi dilakukan melalui pengamatan pembelajaran di sekolah, meliputi pengamatan perencanaan, pelaksanaan, dan penilaian hasil belajar matematika di kelas. Hasil pengamatan kemudian diinformasikan kepada guru matematika yang telah diamati. Hasil dari semua tahapan implementasi adalah model evaluasi proses pembelajaran matematika di SMP berdasarkan Kurikulum 2013.

Penelitian pengembangan ini dilakukan pada bulan Juli sampai dengan Oktober 2013. Penelitian dilakukan di sekolah SMP yang pada tahun ajaran 2013/2014 melaksanakan pembelajaran berdasarkan Kurikulum 2013 di wilayah Provinsi Kalimantan Tengah.

Subjek penelitian adalah guru mata pelajaran matematika di SMP yang mulai tahun ajaran 2013/2014 melaksanakan pembelajaran matematika berdasarkan Kurikulum 2013, kepala sekolah SMP yang tergabung pada kelompok kerja kepala sekolah (KKKS), dan pengawas sebagai pembina di sekolah SMP di Provinsi Kalimantan Tengah.

Prosedur pengembangan yang ditempuh dalam penelitian ini adalah dimulai dengan penyusunan desain model evaluasi proses pembelajaran matematika berdasarkan Kurikulum 2013 dan perangkat model evaluasi, dilanjutkan dengan validasi pakar, uji coba, analisis data, dan implementasi hasil.

Penyusunan desain model evaluasi diawali dengan perencanaan model evaluasi yaitu evaluasi perencanaan pembelajaran, evaluasi pelaksanaan pembelajaran, dan evaluasi penilaian hasil belajar peserta didik. Evaluasi perencanaan pembelajaran meliputi kualitas silabus, kualitas RPP, dan kualitas bahan ajar. Evaluasi pelaksanaan pembelajaran meliputi kualitas kegiatan pendahuluan, kegiatan inti, dan penutup. Evaluasi penilaian hasil belajar meliputi kualitas perencanaan penilaian, pelaksanaan penilaian, pengolahan dan pelaporan hasil penilaian. Berdasarkan komponen-komponen tersebut disusun instrumen, prosedur, dan panduannya. Semua instrumen menggunakan skala Likert dengan 4 pilihan. Model evaluasi ini merupakan modifikasi model CIPP dan Kirkpatrick, yang menekankan pada komponen proses pembelajaran. Desain prosedur, instrumen, dan panduan penggunaan model evaluasi proses pembelajaran matematika yang disusun merupakan draf awal dari model yang dikembangkan.

Setelah prosedur evaluasi, instrumen dan panduan pelaksanaan evaluasi pembelajaran disusun, selanjutnya divalidasi dan dilakukan uji keterbacaan oleh para ahli (*expert judgement*) yaitu dari akademisi atau dosen, dan praktisi yaitu guru, kepala sekolah, pengawas, dan widyaiswara. Proses validasi dilakukan melalui FGD (*focus group discussion*), bertujuan mengetahui validitas in-

strumen, kejelasan prosedur, kekomprehensifan komponen, dan kejelasan panduan. Instrumen dinilai berdasarkan kejelasan petunjuk, cakupan isi, bahasa baku, dan tata tulis. Panduan evaluasi dinilai berdasarkan kejelasan dan aspek bahasa. Berdasar hasil penilaian melalui FGD dan uji keterbacaan, selanjutnya disusun desain model evaluasi.

Instrumen dan panduan evaluasi proses pembelajaran yang telah direvisi berdasarkan penilaian dan masukan yang diperoleh dari FGD, selanjutnya diujicobakan di beberapa sekolah. Subjek coba penelitian adalah terdiri dari guru matematika SMP, kepala sekolah SMP, dan pengawas di Provinsi Kalimantan Tengah. Uji coba dilakukan untuk mengetahui kesesuaian model evaluasi, ketepatan model evaluasi, kemudahan penggunaan model evaluasi, dan efisiensi pelaksanaan model evaluasi. Panduan model evaluasi proses pembelajaran matematika divalidasi dari segi kepraktisan penggunaan, kejelasan instrumen evaluasi, dan kesesuaian dengan Bahasa Indonesia yang baik dan benar. Uji coba instrumen dilakukan sebanyak tiga tahap dengan jumlah subjek berturut-turut 90 orang, 150 orang, 240 orang terdiri guru matematika, kepala sekolah SMP, dan pengawas. Data dari hasil uji coba dianalisis untuk membuktikan validitas dan mengestimasi reliabilitasnya. Data tentang kejelasan prosedur evaluasi, kekomprehensifan komponen model, kejelasan instrumen, dan kejelasan panduan evaluasi dianalisis secara deskriptif. Data pelaksanaan proses yang diperoleh melalui pengamatan pembelajaran dianalisis menggunakan teknik *inter-rater* untuk mengetahui reliabilitasnya. Pengamatan proses pembelajaran matematika di kelas dilakukan oleh dua orang *rater* yang terdiri dari kepala sekolah/wakil kepala sekolah, pengawas satuan pendidikan, atau guru senior.

Hasil akhir model evaluasi proses pembelajaran matematika yang telah dianalisis, merupakan *prototype* model evaluasi proses pembelajaran matematika di SMP yang baik. Selanjutnya, model evaluasi proses pembelajaran matematika dikemas dan dilakukan implementasi model ke beberapa

sekolah SMP di Provinsi Kalimantan Tengah yang telah melaksanakan kegiatan proses pembelajaran matematika dengan menggunakan Kurikulum 2013. *Prototype* dalam hal ini adalah instrumen evaluasi perencanaan pembelajaran matematika, instrumen evaluasi pelaksanaan pembelajaran matematika, instrumen evaluasi pelaksanaan penilaian hasil belajar matematika peserta didik, dan panduan penggunaan model evaluasi pembelajaran matematika yang telah diujicobakan dan divalidasi oleh para pakar, praktisi, dan pengguna model evaluasi proses pembelajaran matematika di SMP.

Produk model evaluasi proses pembelajaran matematika SMP yang diuji coba adalah draf model evaluasi yang terdiri dari instrumen evaluasi pelaksanaan proses pembelajaran matematika, dan panduan pelaksanaan evaluasi. Uji coba dimaksudkan untuk memperoleh data secara lengkap sebagai bahan perbaikan produk.

Instrumen yang diujicobakan yaitu instrumen evaluasi perencanaan pembelajaran matematika, instrumen evaluasi pelaksanaan pembelajaran, dan instrumen evaluasi pelaksanaan penilaian hasil belajar peserta didik. Uji coba dilakukan untuk membuktikan validitas, mengestimasi reliabilitas, dan mengetahui tingkat kesepakatan antar-rater. Kejelasan instrumen dilihat dari: petunjuk instrument evaluasi, cakupan kualitas pembelajaran yang dievaluasi, bahasa baku yang digunakan, dan tata tulis. Cakupan kualitas proses pembelajaran adalah: perencanaan pembelajaran, pelaksanaan proses pembelajaran, dan penilaian hasil belajar peserta didik. Penilaian aspek bahasa diarahkan pada: (1) pernyataan yang komunikatif, dan (2) penggunaan kata dan kalimat yang mudah dipahami. Penilaian tata tulis instrumen diarahkan pada bentuk huruf, ukuran huruf dan format atau *layout* instrumen evaluasi.

Uji kejelasan instrumen evaluasi proses pembelajaran matematika difokuskan pada penilaian terhadap: (1) kejelasan petunjuk instrumen hasil pembelajaran matematika; (2) indikator penilaian sikap, penilaian pengetahuan, dan penilaian keterampilan matematika peserta didik, (3) pernyataan

yang komunikatif, (4) penggunaan kalimat yang mudah untuk dipahami, dan (5) penilaian tata tulis pada instrumen evaluasi hasil pembelajaran matematika di SMP berdasarkan Kurikulum 2013.

Penilaian panduan penggunaan model evaluasi difokuskan pada aspek isi panduan dan bahasa yang digunakan. Penilaian isi panduan meliputi: (1) kejelasan petunjuk pelaksanaan evaluasi; (2) langkah-langkah kegiatan evaluasi; (3) kejelasan arah dan tujuan rekomendasi hasil evaluasi; dan (4) kejelasan waktu pelaksanaan evaluasi. Penilaian aspek bahasa diarahkan pada: (1) perumusan pada pernyataan yang komunikatif; dan (2) penggunaan kalimat yang mudah dipahami oleh pengguna model evaluasi. Langkah-langkah yang dilakukan pada uji coba adalah: (1) penjelasan tentang konsep model evaluasi pembelajaran matematika, dan naskah konsep model yang akan dikembangkan kepada pengguna model; (2) kejelasan pengguna model evaluasi dalam mempelajari kelengkapan konsep, dan memberikan saran dan masukan atas kekurangan pada model yang akan dikembangkan.

Subjek coba dalam penelitian ini adalah guru matematika SMP, kepala sekolah, dan pengawas yang melaksanakan kurikulum 2013. Adapun data yang diambil adalah data kuantitatif yang diubah menjadi data kualitatif. Data meliputi: (a) data model evaluasi yang meliputi: objektivitas model evaluasi pembelajaran, kepraktisan model evaluasi pembelajaran dan efisiensi model evaluasi pembelajaran; (b) data panduan evaluasi pembelajaran meliputi: kelengkapan panduan evaluasi pembelajaran dan kejelasan panduan evaluasi pembelajaran, (c) data kualitas pelaksanaan pembelajaran SMP meliputi: kualitas silabus, bahan ajar, RPP, kualitas kegiatan awal, kegiatan inti, dan kualitas kegiatan akhir, kualitas perencanaan penilaian, kualitas pelaksanaan penilaian, dan kualitas pelaksanaan pengolahan dan pelaporan hasil penilaian.

Dalam penelitian ini, instrumen pengumpul data terdiri atas: (a) panduan wawancara (b) angket/kuesioner, (c) lembar observasi, dan (d) dokumentasi. Wawancara

dilakukan pada tahap studi pendahuluan, untuk memperoleh informasi mengenai pelaksanaan evaluasi proses pembelajaran matematika di SMP. Wawancara dilakukan pada guru matematika SMP, kepala sekolah, dan pengawas sekolah. Selain itu, wawancara juga digunakan untuk mendapatkan informasi yang mendalam pada forum diskusi dengan pakar, praktisi, dan guru mengenai model evaluasi beserta perangkatnya. Observasi dilakukan untuk melengkapi data hasil wawancara. berkaitan pelaksanaan evaluasi pembelajaran. Kegiatan observasi dilakukan pada pelaksanaan proses pembelajaran yang dilakukan oleh guru matematika di kelas. Selain itu, observasi juga dilakukan pada keterlaksanaan model yang dikembangkan mulai dari: (1) persiapan pembelajaran; (2) pelaksanaan pembelajaran; dan (3) penarikan kesimpulan hasil. Observasi dilaksanakan di kelas pembelajaran yang dilakukan guru, diamati oleh teman sejawat, kepala sekolah atau pengawas. Dalam penelitian ini, angket digunakan untuk memperoleh data tentang kejelasan instrumen, dan kejelasan panduan penggunaan model evaluasi pembelajaran yang dikembangkan. Responden adalah guru-guru matematika SMP, kepala sekolah, dan pengawas. Instrumen pengumpul data terdiri dari: (1) instrumen untuk menggali data tentang kejelasan prosedur evaluasi, digunakan kuesioner; (2) instrumen tentang keefektifan model digunakan kuesioner; (3) instrumen tentang kejelasan dan penggunaan panduan evaluasi digunakan kuesioner, (4) instrumen tentang proses pembelajaran di kelas, digunakan observasi, (5) instrumen kejelasan petunjuk, cakupan indikator, penggunaan bahasa, dan tata tulis digunakan kuesioner.

Validitas instrumen meliputi validitas isi dan validitas konstruk. Pembuktian validitas isi dilakukan dengan melalui *rational judgement* dan *expert judgement*. Pembuktian validitas konstruk dilakukan dengan menggunakan analisis faktor. Estimasi reliabilitas instrumen evaluasi proses pembelajaran matematika dilakukan melalui pengamatan terhadap perencanaan pembelajaran, pelaksanaan pembelajaran, dan pelaksanaan

penilaian hasil belajar peserta didik. Teknik yang digunakan untuk menentukan reliabilitas instrumen adalah melalui *inter-rater reliability*, yaitu kesepakatan (*agreement*) yang dibuat antar-rater dalam menilai suatu objek (Gwet, 2012, p. 4). Untuk mengukur tingkat kesepakatan antarpenilai (*Inter-rater-reliability*) dilakukan dengan mengamati koefisien reliabilitas Alpha. Batas bawah koefisien reliabilitas yang digunakan untuk suatu tes yang baik yaitu sebesar 0,70 (Linn, 1989, p.106)

Teknik analisis data menggunakan analisis kuantitatif dan deskriptif kualitatif. Analisis kuantitatif digunakan mengetahui validitas instrumen evaluasi, reliabilitas instrumen evaluasi, kecocokan model instrumen, dan keefektifan model. Analisis deskriptif kualitatif digunakan untuk menganalisis prosedur evaluasi, instrumen evaluasi, dan panduan evaluasi. Data dianalisis oleh pakar/ahli, praktisi, dan pengguna model evaluasi.

Analisis data kuantitatif digunakan untuk membuktikan validitas dan mengestimasi reliabilitas instrumen. Pemeriksaan validitas dilakukan dengan *exploratory factor analysis*, reliabilitas dihitung dengan teknik *internal consistency*, yaitu melihat koefisien Alpha, dengan bantuan *SPSS 17.00 for Windows*. Pengukuran reliabilitas instrumen pelaksanaan pembelajaran dilakukan dengan teknik *interrater reliability*. Teknik untuk mengestimasi reliabilitas menggunakan rumus Cohen's Kappa sebagai berikut.

$$Kc = \frac{pa-pe}{1-pe} \text{ (Gwet KL, 2012, p.12)}$$

$$Pa = \frac{n11+n12}{n} \text{ dan}$$

$$pe = [(n1+/n)(n+1/n)][n2+/n)(n+2/n)]$$

Perhitungan reliabilitas instrumen dilakukan dengan bantuan Program *SPSS 17.00 for Windows*. Pedoman yang digunakan menginterpretasi koefisien Kappa mengacu pada pedoman yang diberikan Fleiss, Levin & Paik, (2003, p. 604), sebagai berikut.

$K > 0,75$: *exelent agreement*

$0,4 \leq K \leq 0,75$: *fair to good agreement*

$K < 0,4$: *poor agreement*

Analisis data secara deskriptif kualitatif digunakan dalam menganalisis data hasil validasi dari para ahli. Berdasarkan skor yang diberikan, dicari skor rerata, dikonversi ke skala dan dideskripsikan. Konversi data kuantitatif ke kualitatif merupakan modifikasi dari Sudijono (2011, p.329).

Tabel 1. Konversi Data Kuantitatif ke Data Kualitatif

Rumus	Rerata	Klasifikasi
$X > \bar{X}_t + 1,8 \text{ Sb}_i$	$X > 4,2$	Sangat Baik
$\bar{X}_t + 0,6\text{Sb}_i < X \leq \bar{X}_t + 1,8 \text{ Sb}_i$	$3,4 < X \leq 4,2$	Baik
$\bar{X}_t - 0,6\text{Sb}_i < X \leq \bar{X}_t + 0,6\text{Sb}_i$	$2,6 < X \leq 3,4$	Cukup
$\bar{X}_t - 1,8\text{Sb}_i < X \leq \bar{X}_t - 0,6\text{Sb}_i$	$1,8 < X \leq 2,6$	Kurang
$X \leq \bar{X}_t - 1,8 \text{ Sb}_i$	$X \leq 1,8$	Tidak Baik

Berdasar rerata skor dan klasifikasi hasil pengembangan tersebut, diadakan penilaian terhadap instrumen, dan panduan model evaluasi sebagai hasil pengembangan dengan dengan standar penilaian disajikan pada Tabel 2.

Tabel 2. Kriteria Penilaian Instrumen, dan Panduan

Rerata Skor (X)	Klasifikasi
$X > 4,2$	Sangat Baik
$3,4 < X \leq 4,2$	Baik
$2,6 < X \leq 3,4$	Cukup
$1,8 < X \leq 2,6$	Kurang Baik
$X \leq 1,8$	Tidak Baik

Hasil Penelitian dan Pembahasan

Tahap Pendahuluan

Kegiatan yang dilakukan meliputi: (1) observasi awal di SMP yang melaksanakan Kurikulum 2013 di Provinsi Kalimantan Tengah, (2) identifikasi masalah di lapangan, (3) kajian pustaka/kajian peraturan pemerintah, dan (4) kajian hasil penelitian yang relevan.

Observasi awal bertujuan untuk mengetahui sejauh mana sekolah telah melaksanakan pembelajaran matematika berdasar-

kan Kurikulum 2013, pelaksanaan evaluasi pembelajaran. menemukan model evaluasi, dan menetapkan komponen model evaluasi yang dikembangkan.

Identifikasi masalah lapangan diperoleh dari data hasil observasi awal yang dilakukan di beberapa SMP di Provinsi Kalimantan Tengah. Identifikasi masalah yang diperoleh berkaitan dengan pembelajaran matematika adalah: (1) Sekolah belum melakukan evaluasi pembelajaran matematika sesuai Kurikulum 2013; (2) Penataan ruang kelas belum mengarah pada pembelajaran matematika dengan pendekatan saintifik; (3) Aktivitas pembelajaran matematika yang dilakukan guru di kelas belum teramati secara optimal oleh kepala sekolah ataupun pengawas satuan pendidikan; (4) fungsi kepengawasan masih menekankan kelengkapan administrasi pembelajaran yang dimiliki oleh guru; (5) Evaluasi terhadap kegiatan pembelajaran matematika di kelas belum melibatkan teman sejawat; (6) Sekolah belum memiliki instrumen evaluasi proses pembelajaran matematika; (7) Sekolah belum memiliki model laporan evaluasi proses pembelajaran matematika menurut Kurikulum 2013.

Tahap Perencanaan Pengembangan

Kajian pustaka dan kajian Peraturan Pemerintah melalui Permendikbud RI Nomor 65 Tahun 2013 tentang Standar Proses, dan Permendikbud Nomor 81A tentang Implementasi kurikulum digunakan untuk menyusun sebuah model evaluasi proses pembelajaran matematika di sekolah SMP berdasarkan Kurikulum 2013. Model evaluasi yang disusun meliputi prosedur evaluasi, komponen, indikator dan instrumennya, serta panduan pelaksanaan evaluasi.

Kajian penelitian terdahulu yang relevan digunakan untuk mendukung kelengkapan model evaluasi yang telah disusun. Selanjutnya model evaluasi tersebut divalidasi oleh para pakar melalui *Focus Group Discussion (FGD)*. Hasil diskusi yang dilaksanakan melalui FGD disepakati bahwa ada tiga komponen yang layak untuk dikembangkan dalam evaluasi proses pembelajar-

an matematika di SMP yaitu: evaluasi perencanaan pembelajaran, evaluasi pelaksanaan proses pembelajaran, dan evaluasi pelaksanaan penilaian hasil belajar. Setiap komponen disusun subkomponen berdasarkan indikatornya. Subkomponen evaluasi perencanaan meliputi silabus, RPP, dan bahan ajar, subkomponen evaluasi pelaksanaan pembelajaran meliputi kegiatan awal, inti, dan penutup, dan subkomponen evaluasi pelaksanaan penilaian meliputi evaluasi perencanaan penilaian, evaluasi pelaksanaan penilaian, evaluasi pelaksanaan pengolahan dan pelaporan hasil belajar. Setiap subkomponen diuraikan menjadi butir pernyataan yang selanjutnya disusun dalam instrumen-instrumen sesuai dengan bunyi indikatornya.

Sesuai dengan komponen yang telah ditetapkan dalam desain model evaluasi, prosedur evaluasi pembelajaran yang harus diikuti adalah sebagai berikut. Pertama, dilakukan evaluasi perencanaan pembelajaran matematika yang meliputi kualitas silabus, RPP, dan bahan ajar yang telah dipersiapkan guru. Kedua, dilakukan evaluasi terhadap pelaksanaan proses pembelajaran, yang meliputi kegiatan pendahuluan, inti, dan penutup selama pelaksanaan pembelajaran matematika di kelas. Ketiga, dilakukan evaluasi terhadap pelaksanaan penilaian hasil belajar yang meliputi kualitas perencanaan, pelaksanaan penilaian hasil belajar, dan kualitas pengolahan serta pelaporan hasil belajar.

Tahap Uji Coba

Hasil Uji Keterbacaan

Uji keterbacaan atau kejelasan perangkat model evaluasi proses pembelajaran matematika didasarkan pada hasil validasi (penilaian) dari para ahli evaluasi (dosen 7 orang), pemakai model evaluasi (kepala sekolah 11 orang, pengawas 20 orang, guru matematika SMP 20 orang), serta praktisi (widyaiswara LPMP Kalimantan Tengah 20 orang), yaitu yang memberikan masukan-masukan dalam rangka perbaikan instrumen. Kejelasan perangkat model evaluasi proses pembelajaran matematika meliputi

prosedur pelaksanaan evaluasi pembelajaran, instrumen evaluasi pembelajaran, panduan pelaksanaan evaluasi pembelajaran, dan efektivitas model evaluasi proses pembelajaran yang dikembangkan.

Kejelasan Prosedur Evaluasi

Hasil uji keterbacaan prosedur evaluasi proses pembelajaran disajikan pada Tabel 3. Uji keterbacaan prosedur evaluasi pembelajaran diarahkan pada aspek: (1) kejelasan prosedur evaluasi, (2) keefektifan kalimat yang digunakan, (3) ketepatan ejaan dan tanda baca, (4) kelengkapan komponen evaluasi, (5) kepraktisan prosedur pelaksanaan evaluasi, (6) kepraktisan dan kesederhanaan prosedur penyajiannya, (7) efisien waktu, (8) efisien biaya, (9) efisien tenaga. Penilaian keterbacaan prosedur evaluasi pelaksanaan pembelajaran menggunakan skala 5.

Tabel 3. Rekapitulasi Hasil Uji Keterbacaan Prosedur Evaluasi

No	Rerata Skor Penilai					Rata Skor	Skor Total
	DS	WI	PS	KS	GR		
1	4,2	4,1	4,0	4,0	4,0	4,06	
2	4,3	4,0	4,0	4,1	4,0	4,06	
3	4,0	4,1	4,1	4,1	4,0	4,06	
4	4,1	4,1	4,0	4,1	4,2	4,10	
5	4,2	4,0	4,1	4,0	4,2	4,10	4,07
6	4,0	4,0	4,2	4,1	4,1	4,08	
7	4,0	4,2	4,0	4,1	4,1	4,08	
8	4,1	4,2	4,0	4,0	4,0	4,06	
9	4,0	4,1	4,0	4,2	4,0	4,06	

Tabel 3 menunjukkan bahwa rerata skor total adalah 4,07. Sesuai pedoman konversi yang ditetapkan, rerata skor termasuk pada klasifikasi baik dapat digunakan melakukan evaluasi tanpa perbaikan. Hal ini memberikan indikasi bahwa tingkat keterbacaan pada prosedur evaluasi proses pembelajaran yang telah dikembangkan dapat diklasifikasikan layak atau baik.

Kejelasan Instrumen Evaluasi

Hasil penilaian uji keterbacaan instrumen evaluasi disajikan pada Tabel 4. Uji

keterbacaan diarahkan pada aspek-aspek yaitu: (1) kejelasan petunjuk pengisian instrumen, (2) kelengkapan instrumen evaluasi perencanaan, (3) kelengkapan instrumen evaluasi pelaksanaan, (4) kelengkapan instrumen evaluasi penilaian, (5) kesesuaian butir instrumen dengan indikator, (6) kejelasan maksud kalimat/ Pernyataan instrumen, (7) penggunaan kata/istilah yang mudah dipahami oleh responden, (8) ketepatan ejaan dan tanda baca, (9) bentuk dan ukuran huruf yang digunakan pada instrumen.

Tabel 4. Rekapitulasi Hasil Uji Kejelasan Instrumen

No	Rerata Skor Penilai					Rerata Skor Total
	DS	WI	PS	KS	GR	
1	4,2	4,1	4,0	4,0	4,0	
2	4,3	4,0	4,0	4,1	4,0	
3	4,0	4,1	4,1	4,1	4,0	
4	4,1	4,1	4,0	4,1	4,2	
5	4,2	4,0	4,1	4,0	4,2	4,07
6	4,0	4,0	4,2	4,1	4,1	
7	4,0	4,2	4,0	4,1	4,1	
8	4,1	4,2	4,0	4,0	4,0	
9	4,0	4,1	4,0	4,2	4,0	

Penilaian keterbacaan instrumen menggunakan skala 5. Tabel 4 menunjukkan rerata skor total sebesar 4,05. Sesuai pedoman, rerata skor termasuk klasifikasi baik, dapat digunakan tanpa perbaikan.

Kejelasan Panduan

Uji keterbacaan panduan evaluasi diarahkan: (1) kejelasan petunjuk umum; (2) kejelasan langkah evaluasi; (3) kejelasan rekomendasi; (4) kejelasan waktu pelaksanaan evaluasi; (5) kejelasan cara pengolahan skor penilaian dan pelaporan hasil penilaian; (6) kejelasan isi panduan, (7) keefektifan kalimat/kalimat baku panduan evaluasi, (8) ketepatan penggunaan ejaan dan tanda baca, (9) penggunaan istilah yang mudah dipahami oleh responden, (10) bentuk dan ukuran huruf. Adapun hasil penilaian uji keterbacaan instrumen evaluasi disajikan pada Tabel 5

Tabel 5. Rekapitulasi Hasil Uji Keterbacaan Panduan Evaluasi

No	Rerata Skor Penilai					Rerata Skor Total
	DS	WI	PS	KS	GR	
1	4,1	4,1	4,0	4,0	4,0	4,08
2	4,1	4,1	4,1	4,1	4,0	
3	4,2	4,1	4,3	4,1	4,0	
4	4,1	4,1	4,2	4,0	4,0	
5	4,1	4,2	4,2	4,0	4,0	
6	4,2	4,0	4,1	4,0	4,1	
7	4,0	4,0	4,1	4,0	4,0	
8	4,0	4,0	4,2	4,1	4,1	
9	4,0	4,1	4,1	4,1	4,1	
10	4,0	4,1	4,0	4,0	4,1	

Penilaian keterbacaan panduan menggunakan skala 5. Tabel 5 menunjukkan rerata skor total besarnya 4,08. Sesuai dengan pedoman konversi rerata skor tersebut termasuk pada klasifikasi baik atau dapat digunakan untuk melakukan evaluasi tanpa perbaikan.

Validitas dan Reliabilitas Instrumen

Pelaksanaan validasi instrumen model evaluasi proses pembelajaran matematika di SMP melalui uji coba di lapangan sebanyak tiga kali dengan jumlah responden yang semakin meningkat. Hasil validasi Ahli dan Praktisi sebagai berikut.

Validitas dan Reliabilitas Instrumen

Hasil validasi oleh pakar (ahli) dan praktisi (Guru, kepala sekolah, pengawas, widyaiswara) melalui FGD (*Focus group discussion*) disajikan pada Tabel 6.

Pelaksanaan validasi instrumen model evaluasi proses pembelajaran matematika di SMP melalui uji coba di lapangan sebanyak tiga kali dengan jumlah responden yang semakin meningkat, seperti pada Tabel 7.

Tabel 6. Rekapitulasi Hasil Validasi Pakar

No	Kompetensi yang dievaluasi	Skor Rerata		Rerata	Kualifikasi
		Ahli	Prakt		
1	Prosedur	4,10	4,18	4,14	4,08
2	Instrumen	4,16	4,20	4,18	
3	Panduan	4,22	4,11	4,17	

Tabel 7. Rekapitulasi Hasil Validasi Pakar

Instrumen	KMO	Loading Factor		Total Var	Koef Kapa
		Min	Mk		
Perencanaan	1	0,73	0,70	0,96	78%
	2	0,84	0,76	0,97	87%
	3	0,82	0,75	0,97	89%
Pelaksanaan	1	0,72	0,65	0,92	70%
	2	0,76	0,72	0,96	85%
	3	0,79	0,74	0,96	86%
Penilaian.	1	0,71	0,63	0,96	80%
	2	0,72	0,62	0,97	84%
	3	0,81	0,72	0,97	86%

Semua Uji *Bartlett,s test of Sphericity* signifikan ($\text{sig}<0,05$)

Berdasarkan tabel rekapitulasi hasil analisis instrumen evaluasi pembelajaran matematika, semua harga KMO lebih besar dari 0,5, dan semua hasil uji *Bartlett,s test of Sphericity* signifikan ($\text{sig}<0,05$). Muatan faktor untuk semua instrumen ketiga tahap uji coba semuanya lebih besar dari 0,5 sehingga instrumen dinyatakan valid. Semua instrumen evaluasi proses pembelajaran memiliki koefisien reliabilitas di atas 0,7, untuk ketiga tahap uji coba. Dengan demikian, dapat dikatakan bahwa instrumen evaluasi perencanaan pembelajaran, instrumen evaluasi pelaksanaan proses pembelajaran, dan instrumen evaluasi pelaksanaan penilaian hasil belajar peserta didik telah memiliki validitas dan reliabilitas yang baik. Dengan demikian, telah dapat digunakan untuk mengumpulkan data.

Tahap Implementasi

Untuk mengetahui efektivitas model evaluasi proses pembelajaran matematika di SMP berdasarkan Kurikulum 2013, maka model di sebarakan kepada 20 orang guru matematika dan 15 orang kepala SMP untuk dinilai efektivitasnya. Suatu model evaluasi dikatakan efektif atau baik jika memenuhi syarat: komprehensif, praktis, ekonomis, dan didukung oleh instrumen pengumpul data yang valid dan reliabel, serta memiliki kejelasan bahasa. Untuk itu, penilaian model evaluasi proses pembelajaran matematika ini difokuskan pada kejelasan, komunikatif, komprehensif, kepraktisan, dan efisien.

Penilaian terhadap efektivitas model evaluasi diarahkan pada: (1) kejelasan prosedur (2) kekomunikatifan bahasa yang digunakan, (3) kelengkapan komponen model evaluasi, (4) kelengkapan indikator instrumen, (5) kepraktisan panduan model evaluasi, (6) kepraktisan instrumen model evaluasi pembelajaran, (7) efisien dari segi waktu, (10) efisiensi dari segi biaya yang diperlukan, (11) efisiensi dari segi tenaga yang digunakan. Penilaian efektivitas model evaluasi proses pembelajaran menggunakan skala 5, yaitu, 5=sangat efektif, 4=efektif, 3=cukup, 2=kurang efektif, dan 1=tidak efektif. Berdasarkan hasil penilaian t, dihitung skor total. Rerata skor hasil penilaian terhadap model evaluasi proses pembelajaran matematika di SMP berdasarkan Kurikulum 2013 disajikan dalam Tabel 8.

Tabel 8. Rekapitulasi Hasil Penilaian Efektivitas Model Evaluasi

No Aspek	Rerata Skor Penilai					Rata Skor
	Ds	WI	PS	KS	Guru	
1	4,1	4,0	4,0	4,0	4,2	4,06
2	4,0	4,1	4,1	4,1	4,1	4,08
3	4,0	4,0	4,1	4,0	4,1	4,04
4	4,0	4,1	4,1	4,1	4,1	4,08
5	4,0	4,1	4,1	4,0	4,0	4,04
6	4,1	4,2	4,0	4,0	4,1	4,08
7	4,0	4,1	4,0	4,1	4,1	4,06
8	4,1	4,1	4,1	4,0	4,0	4,06
9	4,1	4,2	4,0	4,0	4,1	4,08
Rerata Skor Total = 4,06						

Berdasarkan rerata skor total 4,06 dan dikonsultasikan dengan pedoman konversi skor, masuk pada klasifikasi efektif, memberikan indikasi bahwa tingkat keterbacaan instrumen evaluasi pembelajaran yang dikembangkan dapat diklasifikasikan baik. Untuk memperoleh informasi yang akurat terhadap pelaksanaan proses pembelajaran, diperlukan instrumen pengumpul data yang valid dan reliabel. Berdasarkan hasil uji coba lapangan, ternyata instrumen yang dikembangkan dalam penelitian ini telah memiliki validitas, reliabilitas, maupun kecocokan model yang baik. Oleh karena semua per-

syaratan sebagai model tergolong efektif dan telah terpenuhi, maka dapat dikatakan bahwa model evaluasi proses pembelajaran matematika sangat efektif untuk diterapkan dalam mengevaluasi proses pembelajaran matematika di SMP berdasarkan Kurikulum 2013.

Simpulan dan Saran

Simpulan

Berdasarkan hasil penelitian diperoleh simpulan sebagai berikut.

Pertama, evaluasi proses pembelajaran terdiri tiga komponen yaitu: evaluasi perencanaan pembelajaran, evaluasi pelaksanaan pembelajaran, dan evaluasi pelaksanaan penilaian hasil belajar. Komponen evaluasi perencanaan pembelajaran meliputi silabus, RPP, dan bahan ajar. Komponen evaluasi pelaksanaan pembelajaran meliputi kegiatan pendahuluan, kegiatan inti, dan kegiatan penutup. Komponen evaluasi pelaksanaan penilaian hasil belajar meliputi evaluasi perencanaan penilaian, evaluasi pelaksanaan penilaian, dan evaluasi pelaksanaan pengelolaan dan pelaporan hasil belajar peserta didik.

Kedua, kualitas panduan evaluasi dan instrumen evaluasi yang digunakan pada kategori baik. Hal ini berdasarkan hasil validasi dan estimasi reliabilitas yang dilakukan melalui *FGD* oleh para pakar dan praktisi.

Ketiga, panduan pelaksanaan evaluasi terdiri dari: ketentuan umum, langkah-langkah pelaksanaan evaluasi pembelajaran, pedoman penyekoran, langkah-langkah penilaian, waktu pelaksanaan evaluasi, rekomendasi hasil evaluasi, dan format laporan hasil penilaian peserta didik.

Keempat, efektivitas model dinilai dari kejelasan prosedur, bahasa yang digunakan, kekomprehensifan komponen model, kejelasan instrumen, kepraktisan panduan evaluasi, dan keekonomisan dari segi biaya, tenaga, dan waktu, termasuk kategori efektif. Penilai efektivitas model evaluasi proses pembelajaran adalah guru, kepala sekolah, dan pengawas. Jawaban atas pernyataan instrumen dinyatakan dalam bentuk skor

dengan skala 1 sampai 5, dan hasilnya adalah: (a) kejelasan prosedur amat baik, (b) kejelasan bahasa amat baik, (c) kekomprehensifan komponen baik, (d) kejelasan indikator baik, (e) kepraktisan panduan sangat baik, (e) keekonomisan segi waktu, biaya dan tenaga sangat baik. Berdasarkan pedoman konversi, rerata skor seluruhnya pada kategori sangat baik. (5) Model evaluasi proses pembelajaran yang dikembangkan sudah baik dan efektif berdasarkan penilaian para pakar, praktisi dan pengguna model.

Saran

Berdasarkan simpulan tersebut disampaikan saran-saran sebagai berikut: (1) Evaluasi model yang dikembangkan dapat dijadikan model untuk mengevaluasi pelaksanaan pembelajaran matematika di SMP berdasarkan Kurikulum 2013; (2) dalam melakukan evaluasi terhadap proses pembelajaran matematika di sekolah, evaluator melakukan kegiatan pengamatan secara terprogram; (3) Sebelum evaluasi model P3 digunakan, sebaiknya terlebih dahulu dilakukan diseminasi secara luas, yang melibatkan pihak-pihak terkait, baik kepada kepala sekolah, guru, pengawas, LPMP, maupun pihak Dinas Pendidikan; (4) Model evaluasi P3 akan lebih lengkap jika dikembangkan dengan komponen produk, sebagai dasar pengambilan keputusan/kebijakan; (5) Model evaluasi ini akan lebih menarik apabila dikembangkan lebih lanjut dengan pembuatan program komputer sehingga seorang evaluator hanya mengoreksi data yang ada dan mengentrikan ke dalam program. Selanjutnya, evaluator tinggal mengambil *print-out* berupa hasil evaluasi dan rekomendasi hasil evaluasi dalam rangka perbaikan proses pembelajaran.

Daftar Pustaka

Borg, W. R., & Gall, M.D. (2003). *Educational research: An Introduction*. Fourth edition. New York: Longman.

Daley, G & Kim, L. (2010). *A teacher evaluation system that works*. National Institute for Excellent in Teaching.

Darling-Hammond, L. (2010). *Evaluating Teacher Effectiveness*. <http://www.google.co.id/url>.

Darmadi, H. (2012). *Kemampuan dasar mengajar: Landasan konsep dan implementasi*. Bandung: Alfabeta.

Depdikbud. (2013). *Permendikbud Nomor 54 tentang Standar kompetensi Lulusan*. Jakarta: Pusat Kurikulum Depdikbud

Depdikbud. (2013). *Permendikbud Nomor 65 tentang Standar Proses*. Jakarta: Pusat Kurikulum Depdikbud.

Depdikbud. (2013). *Permendikbud Nomor 66 tentang Standar Penilaian*. Jakarta: Pusat Kurikulum Depdikbud.

Depdikbud. (2013). *Permendikbud Nomor 68 tentang Kerangka Dasar dan Struktur Kurikulum SMP/MTs*. Jakarta: Pusat Kurikulum Depdikbud.

Depdikbud. (2013). *Permendikbud Nomor 81A tentang Implementasi Kurikulum 2013*. Jakarta: Pusat Kurikulum Depdikbud.

Fitzpatrick, J.L., Sander, J. R., & Worthen, B.L., (2011). *Program evaluation: alternative approaches and practical guidelines*. Boston: Pesron.

Fleiss, J.L., Levin, B., & Paik, M.C. (2003). *Statistical methods for rates and proportions (3rd)*. New Jersey: John Wiley & Sons, Inc, Publication.

Gronlund & Linn. (1985). *Measurement and evaluating teaching*: 6th Edition. New York: Macmillan Publishing Company.

Gwet, K. L. (2012). *Handbook of inter-rater reliability*. Third edition Advanced Analytics, LLC Gaithersburg, USA

Hood, S., & Hopson. R. K. (2008). Evaluation roots reconsidered: Asa Hilliard's "Nobody Knows My Name" project, and African Educational Excellence. *Review of Education Research*, 78, 410-426.

- Howell, K.W & Nolet, V. (2000). *Curriculum based evaluation: teaching and decision making*. Canada: Wadsworth Thomson Learning.
- Kartowagiran, B., Jaedun, A., & Retnawati, H. (2013). *Kesiapan SMP di Daerah Istimewa Yogyakarta untuk menngimplementasikan kurikulum 2013*. Penelitian PEP, Program Pascasarjana Universitas Negeri Yogyakarta.
- Kaufman, R., & Thomas, S. (1980). *Evaluation without fear*. New York: New Viewpoints.
- Linn, R. L. (ed) (1989) *Educational measurement (3th)*. New York: American Council on Education/Mac-millan Publishing Company.
- Majid, A. (2008). *Perencanaan Pembelajaran*. Bandung: Remaja Rosdakarya
- Martiyono. (2012). *Perencanaan Pembelajaran* Yogyakarta : Aswaja Presindo
- Nitko, A.J, & Brookhart S.M. (2007). *Educational asesment of students*. Ohio: Pearson (Merril Prentice New Jersey: Pearson Education.
- Rusman. (2012). *Manajemen kurikulum*. Jakarta: Raja Grafindo Persada.
- Stufflebeam, L.D., Madaus, G.F., & Kellaghan, T.(2002). *Evaluation models: viewpoints on educational and human services evaluation*. Dordrecht; Kluwer Academic Publisher
- Suherman, E. (2003). *Strategi Pembelajaran Kontemporer*. Bandung: Jurusan Pend Matematika Fakultas Pendidikan MIPA, Universitas Pendidikan Indonesia.
- Sudijono, A. (2011). *Pengantar evaluasi pendidikan*. Jakarta: PT Raja Grafindo Persada.
- Uno, H. B. (2008). *Perencanaan pembelajaran*. Jakarta: PT Bumi Aksara.

PENGEMBANGAN INSTRUMEN PENILAIAN KINERJA KEMAMPUAN INKUIRI PESERTA DIDIK PADA MATA PELAJARAN FISIKA SMA

¹⁾Supahar, ²⁾Zuhdan Kun Prasetyo

^{1,2)}Universitas Negeri Yogyakarta, Universitas Negeri Yogyakarta

¹⁾supahar@uny.ac.id, ²⁾zuhdan@uny.ac.id.

Abstrak

Penelitian ini bertujuan mengembangkan instrumen penilaian kinerja kemampuan inkuiri mata pelajaran fisika peserta didik SMA. Instrumen terdiri atas empat paket perangkat tes yang masing-masing terdiri atas 30 butir tes termasuk 9 butir bersama yang telah divalidasi oleh ahli pendidikan fisika, ahli fisika, ahli pengukuran, dan praktisi. Instrumen diujicobakan pada 2015 peserta didik. Model respon butir politomus empat kategori digunakan untuk menganalisis data hasil uji coba menurut *Partial Credit Model (PCM)*. Hasil penelitian menunjukkan bahwa, berdasarkan kriteria batas terendah dan tertinggi *mean INFIT MNSQ*, yaitu 0,77 dan 1,30 butir tes semuanya *fit* dengan model *PCM*. Tingkat kesulitan butir-butir tes berada pada rentang antara -0,83 sampai +0,84 yang berarti butir-butir tes dalam kategori baik karena nilainya di antara -2,0 dan +2,0. Fungsi informasi total tes relatif tinggi untuk kemampuan antara -2,5 sampai +2,5. Dengan demikian, instrumen yang dikembangkan dapat digunakan untuk mengukur kemampuan inkuiri mata pelajaran fisika siswa SMA. Hasil pengukuran menunjukkan bahwa kemampuan inkuiri peserta didik pada mata pelajaran fisika SMA di DIY berada di atas kemampuan rata-rata.

Kata kunci: *kemampuan inkuiri, mata pelajaran fisika SMA, Penilaian kinerja.*

DEVELOPING ASSESSMENT INSTRUMENT OF THE PERFORMANCE OF THE INQUIRY ABILITIES IN PHYSICS SUBJECT OF HIGH SCHOOL STUDENTS

¹⁾Supahar, ²⁾Zuhdan Kun Prasetyo

^{1,2)}Universitas Negeri Yogyakarta, Universitas Negeri Yogyakarta

¹⁾supahar@uny.ac.id, ²⁾zuhdan@uny.ac.id.

Abstract

The research was to develop an assessment instrument of performance for the inquiry ability of physic subject of senior high school students. The instrument consisted of four package sets of test and each of the sets consisted of 30 test items, including 9 items that had been validated by the expert of physic education, the expert of physic, the expert of measurement and the practitioner. The instrument was tested to 2,015 learning participants. The four-category politomous item response model was implemented for the data analysis of the experiment results according to the Partial Credit Model (PCM). The results of the research showed that based on the criteria of the lowest benchmark and the highest benchmark of mean INFIT MNSQ, was 0.77 and 1.30, all of the test items had been fit with the PCM model. The difficulty level of the test items was between -0.83 and +0,84 which implied that the test items belonged to the good category because the value of the test items was between -2.0 and +2.0. The informative function of the total tests was relatively high for ability between -2.5 to +2.5. Thereby, the instrument developed can be used for measuring the inquiry ability of the physic subject of the senior high school students. The results of measurement showed that the inquiry ability of the learning participants in physic subject within the senior high schools in DIY was above average.

Keywords: *inquiry ability, physic subject in senior high schools, performance assessment*

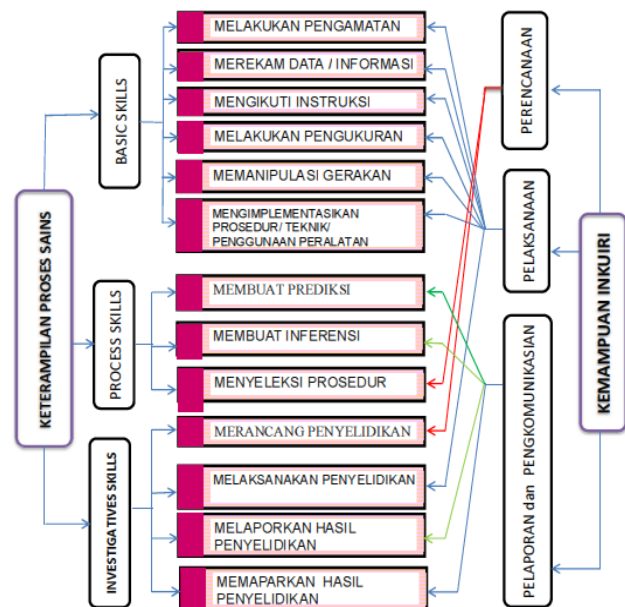
Pendahuluan

Mata pelajaran fisika di SMA menyajikan berbagai pengalaman belajar untuk memahami konsep, melatih kemampuan inkuiri, dan menggunakan ketrampilan proses sains untuk mempelajari fenomena alam. Salah satu tuntutan dalam rumusan kompetensi dasar (KD) dalam standar isi kurikulum nasional untuk jenjang SMA adalah: (1) melakukan percobaan, dan (2) memahami prinsip-prinsip pengukuran, dan (3) melakukan pengukuran besaran fisika secara langsung dan tidak langsung secara cermat, teliti, dan obyektif. Dengan demikian, untuk mencapai kompetensi dasar mata pelajaran fisika SMA diperlukan kemampuan inkuiri

National Science Education Standards dalam Wenning (2005, p.3) menyatakan bahwa, inkuiri bagi peserta didik sebagai kegiatan peserta didik untuk mengembangkan pengetahuan dan pemahaman tentang ide-ide ilmiah, serta pemahaman bagaimana cara para ilmuwan mempelajari dunia alam. Moyer, *et.al.* (2007, p.5) mendefinisikan inkuiri sebagai proses-proses sains yang mencerminkan metode saintifik. Bass, *et.al.* (2009, p.17) mendefinisikan inkuiri sebagai suatu cara ilmuwan melakukan penyelidikan. Dengan demikian, peserta didik di SMA yang terlibat dalam kegiatan inkuiri harus belajar metode saintifik yang meliputi perencanaan, pelaksanaan, dan pelaporan termasuk didalamnya pengkomunikasian hasil penyelidikan.

Chiappetta & Koballa (2010, p.130) menyebutkan bahwa, kemampuan inkuiri meliputi strategi keterampilan proses sains yang didalamnya termasuk keterampilan membuat model dalam bentuk diagram alir atau membuat model matematika serta keterampilan mengkomunikasikan hasil pengamatan atau percobaan. Proses-proses sains menurut Bryce, *et.al.* (1990, pp.1-4) terdiri atas aspek: (a) *basic skills* dan (b) *process skills*. Proses-proses sains menurut Rezba, *et.al.* (1995, p.1) terdiri atas aspek: (a) *basic science process skills*, dan (b) *integrated science process skills*. Proses-proses sains menurut Subali (2009, pp. 581-593) terdiri atas aspek: (a) keterampilan dasar, (b) keterampilan

mengolah/memproses, dan (c) keterampilan menginvestigasi. Berdasarkan pendapat para ahli seperti tersebut, maka dapat dirumuskan bahwa proses-proses sains sebagai kemampuan inkuiri (*processes science as inquiry*) terdiri atas subaspek: (1) perencanaan penyelidikan, (2) pelaksanaan penyelidikan, dan (3) pelaporan hasil penyelidikan. Indikator subaspek perencanaan penyelidikan meliputi: merancang penyelidikan dan menyeleksi prosedur. Indikator subaspek pelaksanaan penyelidikan meliputi kemampuan: melakukan pengamatan, merekam data/informasi, mengikuti instruksi, melakukan pengukuran, memanipulasi gerakan, mengimplementasikan prosedur/teknik/penggunaan peralatan, dan melaksanakan penyelidikan. Indikator subaspek pelaporan terdiri atas subsubaspek: membuat prediksi, membuat inferensi, melaporkan hasil penyelidikan, dan memaparkan hasil penyelidikan. Secara ringkas inkuiri melalui proses-proses sains disajikan pada Gambar 1.



Gambar 1. Kemampuan Inkuiri melalui strategi keterampilan proses sains.

Penilaian hasil belajar peserta didik perlu dilakukan secara berkesinambungan agar perkembangan hasil belajarnya terpan-tau. Sistem penilaian di sekolah sudah di-undang dalam bentuk Peraturan Menteri Pendidikan dan Kebudayaan Nomor 104

Tahun 2014. Format penilaian dapat berupa tes tertulis, observasi, tes praktik, dan penugasan perseorangan atau kelompok. Glencoe science (t.t.,p.3) menyatakan bahwa, penilaian kinerja memberikan jendela pada seberapa baik siswa menggunakan satu dan lebih unsur literasi. Penilaian kemampuan inkuiri mata pelajaran fisika peserta didik SMA dapat diukur menggunakan penilaian kinerja.

Penilaian kinerja idealnya dilakukan melalui metode *direct observation* yang dapat dijadikan sebagai *benchmark*. Metode alternatif lain, sebagai pengganti observasi langsung adalah melalui metode: *notebooks*, simulasi komputer, dan *paper and pencil test* (Ruiz-Primo & Shavelson: 1996, pp.1047-1050). Dalam penelitian ini penilaian kinerja dilakukan menggunakan format penilaian kinerja melalui *paper and pencil test*.

Penilaian kinerja melalui *paper and pencil test* dapat dilakukan sebelum dan sesudah pembelajaran. Tujuannya untuk memastikan bahwa peserta didik telah memiliki informasi dan pengetahuan yang cukup akurat sebelum melakukan percobaan dan sebagai strategi untuk mengungkapkan hasil keterlibatan para peserta didik dalam belajar.

Penilaian kinerja melalui *paper and pencil test* juga digunakan untuk mengungkapkan kemampuan peserta didik tentang pengetahuan dan pemahaman proses-proses sains dan prosedur-prosedur inkuiri (Bass *et. al.*, 2009, p. 173). Pengujian kinerja dilakukan pada kondisi tertentu yang merupakan tiruan dari keadaan yang sesungguhnya. Testi diminta untuk menampilkan kemampuannya dalam mengerjakan suatu tugas. Respon/jawaban peserta didik dinilai dan dibandingkan dengan kriteria yang telah ditetapkan dalam rubrik.

Hasil penilaian kinerja bidang sains dalam program *for International Student Assessment* (PISA), diketahui bahwa skor rata-rata kinerja bidang sains mendudukkan peserta didik Indonesia berada di ranking ke-50 dari 57 negara (Chiappetta & Koballa, 2010, pp. 22-24). Keadaan ini menggambarkan bahwa kemampuan kinerja bidang sains

umumnya dan fisika khususnya bagi peserta didik di Indonesia masih rendah dibandingkan negara-negara lainnya. Penilaian kinerja melalui tes kemampuan inkuiri diharapkan dapat untuk mengukur basis pengetahuan, kemampuan, dan keterampilan inkuiri pada mata pelajaran fisika SMA yang meliputi kemampuan perencanaan, pelaksanaan, dan pelaporan sehingga memberikan gambaran yang lebih lengkap tentang prestasi peserta didik.

Praktik penilaian di sekolah menggunakan metode *testing* dengan butir berbentuk pilihan ganda masih dominan dipergunakan dalam berbagai keperluan pengujian. Di lain pihak, penilaian kinerja jarang digunakan utamanya untuk pengukuran dalam skala besar. Berdasarkan survei awal pada pelajaran fisika SMA, pembelajaran secara inkuiri telah dilakukan tetapi guru belum melaksanakan penilaian kinerja kemampuan inkuiri karena belum tersedianya instrumen untuk mengukur kinerja tersebut. Oleh karena itu, adanya penelitian untuk mendukungnya perlu dilakukan. Penelitian ini akan menjadi sangat penting artinya bagi kegiatan pembelajaran fisika di SMA agar kemampuan inkuiri peserta didik dapat terukur.

Masalah utama yang akan diungkap dalam penelitian ini adalah bagaimanakah instrumen penilaian kinerja kemampuan inkuiri mata pelajaran fisika SMA pada setiap aspek yang disertai karakteristiknya yang meliputi kemampuan perencanaan, pelaksanaan, dan pelaporan?

Sejalan dengan rumusan masalah yang akan diselesaikan, maka tujuan penelitian ini adalah: (1) menghasilkan instrumen pengukur kemampuan inkuiri dalam mata pelajaran fisika di SMA, dan (2) mendapatkan karakteristik instrumen penilaian kinerja kemampuan inkuiri yang meliputi kemampuan perencanaan, pelaksanaan, dan pelaporan.

Metode Penelitian

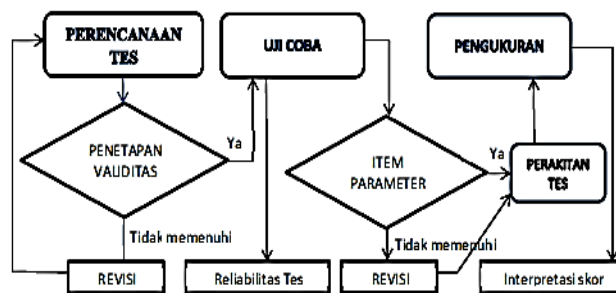
Penelitian pengembangan instrumen penilaian kinerja kemampuan inkuiri mata pelajaran fisika SMA ini menggunakan pendekatan deskriptif kuantitatif. Model

pengembangan perangkat tes kemampuan inkuiri mata pelajaran fisika SMA di DIY menggunakan modifikasi model menurut Mardapi (2012, p.110) dan Oriondo & Dalo-Antonio (1984, pp.10-11). Tahapan pengembangan perangkat tes terdiri atas tiga tahap, yakni (1) tahap perencanaan tes, (2) tahap uji coba perangkat tes, dan (3) tahap pengukuran.

Penelitian dilakukan di SMA Negeri di DIY pada bulan Januari sampai dengan Juni 2013. Perencanaan tes yang berupa penyiapan tes, validasi, dan perakitan tes dilakukan pada bulan Januari-Maret 2013. Uji coba dilakukan di sepuluh SMA Negeri di DIY pada bulan April sampai dengan Mei 2013. Pengukuran dilakukan di tiga belas SMA Negeri di DIY pada bulan Mei-Juni 2013. Pemilihan sekolah sebagai lokasi penelitian didasarkan pada ranking ujian nasional (UN) tahun 2012 yang dipilih secara *purposive sampling*. Pemilihan SMA Negeri diharapkan dapat mewakili variasi yang besar pada karakteristik peserta didik seperti latar belakang keluarga, budaya, etnik, agama, sosial, dan ekonomi.

Subjek penelitian meliputi seluruh peserta didik SMA kelas X, dan XI-IPA yang hadir di lokasi penelitian pada saat pengujian. Kelas XII-IPA tidak dilibatkan sebagai subjek penelitian karena sedang menghadapi UN tahun 2013. Uji coba perangkat penilaian melibatkan 2015 responden dan tahap pengukuran melibatkan 2383 responden.

Prosedur penelitian meliputi kegiatan: (1) Perencanaan tes meliputi penetapan tujuan tes, penyiapan *blue print*, penyusunan butir tes dan rubrik, perakitan butir tes, penetapan validitas oleh pakar, revisi, dan perakitan instrumen. (2) Tahap uji coba meliputi penetapan SMA, pelaksanaan ujian, penskoran, analisis butir, revisi butir tes yang belum memenuhi kriteria parameter butir tes yang diinginkan, dan (3) Tahap pengukuran meliputi perakitan tes berdasarkan hasil uji coba, penetapan SMA, pengujian, penskoran, dan interpretasi hasil pengukuran. Secara bagan dapat disajikan seperti Gambar 2.



Gambar 2. Skema prosedur penelitian

Data penelitian berupa skor hasil tes kemampuan inkuiri mata pelajaran fisika SMA. Skor hasil tes berupa data politomus empat kategori. Instrumen pengumpul data menggunakan perangkat tes kemampuan inkuiri mata pelajaran fisika SMA. Perangkat tes terdiri atas 93 butir soal. Mengingat dalam setiap tatap muka di kelas disediakan waktu selama 90 menit, maka tidak mungkin setiap peserta didik dapat menyelesaikan seluruh butir tes sesuai alokasi waktu yang disediakan. Atas dasar pertimbangan tersebut, rancangan tes dibagi menjadi empat perangkat tes, yaitu perangkat tes berkode A, B, C, dan D. Setiap perangkat tes disusun dengan mempertimbangkan keterwakilan setiap aspek kemampuan inkuiri yang akan diukur.

Setiap perangkat tes terdiri atas 30 butir soal termasuk 9 butir soal *anchor*. Setiap responden hanya diwajibkan mengerjakan satu perangkat tes yang disediakan di bawah pengawasan guru kelas. Keuntungan yang diperoleh dengan desain pengujian seperti ini antara lain, kecurangan dalam mengerjakan soal dapat diminimalkan.

Data respons testi dianalisis menurut *partial credit model* (PCM) yang merupakan pengembangan dari model *Item Response Theory* (IRT) *1 parameter of logistic* (1-PL). Analisis data meliputi: (1) pengujian *goodness of fit item* dengan model PCM; (2) penetapan indeks reliabilitas tes; (3) estimasi tingkat kesukaran item; (4) estimasi parameter *ability*; dan (5) melukiskan kurva fungsi informasi total dan *standard error of measurement* (SEM).

Pengujian *goodness of fit test* terhadap PCM ditentukan besaran *INFIT Mean of*

Square (Mean INFIT MNSQ) beserta simpangan bakunya. Apabila besarnya nilai rata-rata *INFIT MNSQ* mendekati 1.0 dengan simpangan baku mendekati 0.0 maka keseluruhan butir tes *fit* dengan *PCM*. Penetapan *goodness of fit* setiap butir terhadap *PCM* mengikuti kaedah yang ditetapkan oleh Adam & Khoo (1996, p. 30), bahwa suatu butir *fit* dengan model bila besarnya nilai *INFIT MNSQ* pada kisaran 0.7 sampai 1.30. Kisaran nilai *INFIT MNSQ* tersebut membatasi sebaran skor terkalibrasi dan masih berada pada kurva yang berbentuk *Leptokurtic* yang mencerminkan masih dalam kondisi *unity*.

Pengujian *difficulty index* menggunakan kriteria menurut Hambleton & Swaminathan (1985, p. 36), bahwa suatu butir tes dikategorikan sangat sulit jika *difficulty index* (*b*) $> +2$ dan dikategorikan sangat mudah jika $b < -2$. Dengan demikian, pada pengembangan tes ini diusahakan butir tes memiliki nilai $-2 \leq b \leq +2$. Penetapan kriteria reliabilitas tes didasarkan pada pendapat Sumadi Suryabrata (2000, p. 40), bahwa hasil testing untuk membuat keputusan tentang individu hendaknya digunakan tes yang koefisien reliabilitasnya sekurang-kurangnya 0,90. Semakin tinggi koefisien reliabilitas suatu tes, maka kemungkinan kesalahan yang terjadi akan semakin kecil kalau akan membuat keputusan berdasarkan skor yang diperoleh dalam tes.

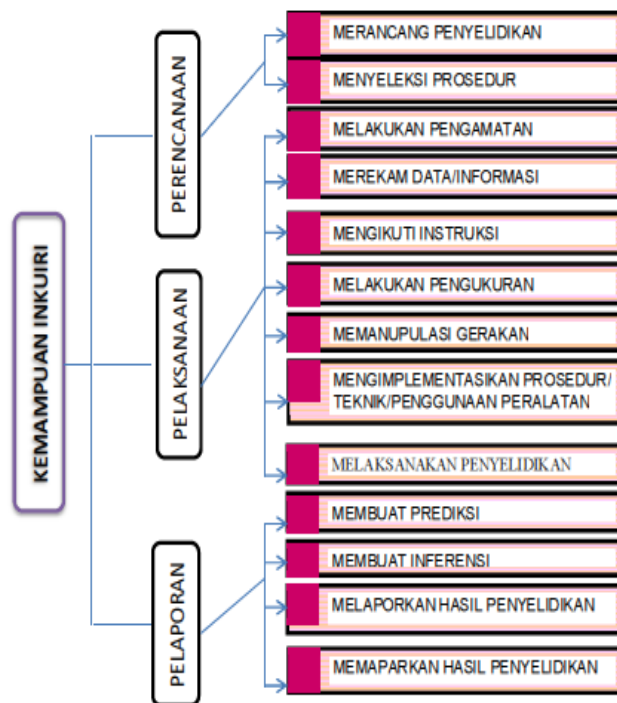
Hasil pengukuran kemampuan inkuiri disajikan dalam bentuk: (1) distribusi frekuensi *ability* dalam skala logit, dan (2) tabel persentase kategori level kemampuan.

Hasil Penelitian dan Pembahasan

Hasil Penelitian

Hasil pengembangan instrumen meliputi rumusan *blue print* tes, indikator, dan butir-butir penilaian kinerja kemampuan inkuiri. Perumusan *blue print* tes penilaian kinerja kemampuan inkuiri diharapkan dapat mencerminkan konstruk keilmuan maupun sebagai target pembelajaran fisika SMA. Indikator penilaian kinerja kemampuan inkuiri mata pelajaran fisika SMA dijadikan

dasar untuk mengembangkan butir tes. Rumusan *blue print* tes penilaian kinerja kemampuan inkuiri mata pelajaran fisika SMA dalam penelitian disajikan pada Gambar 3.



Gambar 3. Blue print tes penilaian kinerja kemampuan inkuiri mata pelajaran fisika SMA.

Indikator aspek perencanaan yang meliputi subaspek merancang penyelidikan disajikan pada Tabel 1, dan menyeleksi prosedur disajikan pada Tabel 2. Indikator aspek pelaksanaan yang meliputi subaspek melakukan pengamatan disajikan pada Tabel 3, merekam data/informasi disajikan pada Tabel 4, mengikuti instruksi disajikan pada Tabel 5, melakukan pengukuran disajikan pada tabel 6, memanipulasi gerakan disajikan pada Tabel 7, mengimplementasikan prosedur/teknik/penggunaan peralatan disajikan pada Tabel 8, dan melaksanakan penyelidikan disajikan pada Tabel 9.

Indikator pelaporan yang meliputi subaspek: membuat inferensi disajikan pada Tabel 10, membuat prediksi disajikan pada Tabel 11, melaporkan hasil penyelidikan disajikan pada Tabel 12, dan memaparkan hasil penyelidikan disajikan pada Tabel 13.

Tabel 1. Merancang Penyelidikan

No.	Indikator
1.	Menetapkan variabel bebas sebagai faktor perlakuan dalam penyelidikan.
2.	Mengidentifikasi hubungan antara variabel bebas dan variabel gayut dalam suatu penyelidikan.
3.	Mengemukakan latar belakang pentingnya diperlukan suatu penyelidikan.
4.	Merumuskan tujuan penyelidikan.
5.	Merumuskan manfaat suatu penyelidikan
6.	Merancang pengendalian variabel. pengganggu (<i>suppressed variabel</i>) menjadi variabel kendali (<i>Control variabel</i>) dalam suatu eksperimen.
7.	Mengidentifikasi suatu variabel acak/random dalam suatu penyelidikan
8.	Menetapkan hipotesis penelitian dalam suatu penyelidikan
9.	Menetapkan variabel pengukuran dalam penyelidikan
10.	Menetapkan prosedur pengumpulan data suatu penyelidikan
11.	Merancang rangkaian peralatan penyelidikan.
12.	Merancang jenis data yang harus dihimpun dalam suatu penyelidikan
13.	Merancang penyajian data hasil suatu penyelidikan.
14.	Merancang teknik analisis data suatu hasil penyelidikan.

Tabel 2. Menyeleksi Prosedur

No	Indikator
1.	Mengantisipasi resiko dan mengadopsi tindakan pencegahan dan prosedur yang tepat dalam melakukan setiap percobaan laboratorium/ penyelidikan praktis.
2.	Mengidentifikasi suatu prosedur yang sesuai (seperti: pengamatan, pencacahan, pengukuran, pengambilan sampel, pemanasan) atau memilih peralatan yang tepat sesuai .
3.	Memilih komponen peralatan yang sesuai untuk menghasilkan pengukuran yang teliti/akurat.
4.	Memilih variabel yang sesuai, mengumpulkan data yang relevan, dan memilih suatu bentuk sajian hasil yang sesuai untuk suatu prosedur penyelidikan yang telah dipilih.

Tabel 3. Melakukan Pengamatan

No.	Indikator
1.	Mengidentifikasi jenis data yang dapat dihimpun dalam melakukan pengamatan dengan atau tanpa alat ukur.
2.	Mengidentifikasi jenis data yang dapat dihimpun dalam melakukan pengamatan sesuai dengan alat indera yang digunakan.
3.	Mengidentifikasi objek berdasarkan deskripsinya.
4.	Mengidentifikasi objek untuk dicocokkan dengan rujukan tertentu (warna/bentuk/fase)
5.	Mengidentifikasi persamaan/perbedaan antarobjek.
6.	Mencocokkan suatu objek dengan beragam representasi visualnya.
7.	Mengidentifikasi situasi berpotensi penuh resiko saat bekerja dilaboratorium dengan yang ada di dalam kehidupan sehari-hari.

Tabel 4. Merekam Data

No.	Indikator
1.	Merekam informasi sederhana dengan merepresentasikannya dalam bermacam bentuk sesuai dengan klasifikasinya.
2.	Membuat suatu ringkasan
3.	Memberi label suatu bagan/ diagram.
4.	Menggambar bentuk sederhana
5.	Menyusun informasi dalam bentuk table lengkap dengan judul table.
6.	Melengkapi suatu bagan/carta, grafik, atau histogram.

Tabel 5. Mengikuti Instruksi

No.	Indikator
1.	Menyelesaikan suatu prosedur (dengan/tanpa alat) setelah guru mendemonstrasikannya.
2.	Menyelesaikan suatu prosedur (dengan atau tanpa alat) menurut perintah yang disampaikan secara lisan.
3.	Menyelesaikan suatu prosedur (dengan/tanpa alat) berdasarkan kartu petunjuk kerja yang seluruhnya/sebagian besar berisi informasi dalam bentuk gambar.
4.	Menyelesaikan suatu prosedur (dengan/tanpa alat) berdasarkan kartu petunjuk kerja yang seluruhnya/sebagian besar berisi informasi berupa tulisan.

Tabel 6. Melakukan Pengukuran

No.	Indikator
1.	Menyediakan unit/alat pengukur untuk melakukan pengukuran
2.	Melakukan pengukuran menggunakan alat ukur yang skalanya dibaca dengan posisi horizontal.
3.	Melakukan pengukuran dengan alat ukur analog.
4.	Melakukan pengukuran dengan alat yang skalanya dibaca dengan posisi vertical.
5.	Melakukan pengukuran dengan alat ukur digital.
6.	Mengestimasi besaran Fisika secara kasar
7.	Menggunakan kertas grafik untuk mengestimasi panjang lintasan/ arah lintasan/ luas penampang permukaan benda.

Tabel 7. Memanipulasi Gerakan

No.	Indikator
1.	Mendemonstrasikan kendali motorik kasar dalam berbagai cara (misalnya menyalakan alat ukur analog/digital).
2.	Mendemonstrasikan kendali motorik halus dalam berbagai cara (misalnya mengkalibrasi, merangkai, melengkapi bagan, memasang alat ukur)

Tabel 8. Mengimplementasikan
Prosedur/teknik/penggunaan
Peralatan

No.	Indikator
1.	Mengidentifikasi peralatan laboratorium yang biasa digunakan.
2.	Memilih peralatan laboratorium yang sesuai dengan spesifikasi tugas yang dikerjakan.
3.	Mengadopsi prosedur laboratorium dengan meminimalkan resiko /kesalahan.
4.	Memindahkan suatu material padat
5.	Memindahkan suatu material cair
6.	Memindahkan material padat atau cair dengan menggunakan cara tertentu
7.	Memisahkan benda menurut ukurannya.
8.	Menggunakan alat ukur untuk menghasilkan suatu volume tertentu yang ditetapkan
9.	Mencampur sejumlah kecil bahan untuk menghasilkan campuran/ramuan sesuai standar yang ditentukan
10.	Menghancurkan material padat menggunakan prosedur/metode tertentu
11.	Menjaga keselamatan kerja jika bekerja menggunakan api atau alat yang mudah terbakar
12.	Menjaga keselamatan kerja jika bekerja dengan peralatan sumber cahaya
13.	Menjaga keselamatan kerja jika bekerja dengan peralatan kaca
14.	Merangkai suatu rangkaian listrik sederhana sesuai bagan rangkaian yang ditetapkan.

Tabel 9. Melaksanakan Penyelidikan

Indikator
Secara simultan melakukan: (1) menyusun rangkaian peralatan penyelidikan/eksperimen sesuai rancangan; (2) mengatur perlakuan dalam eksperimen sesuai rancangan; (3) memanipulasi variabel pengganggu atau variabel asing sesuai rancangan;(4) melakukan pengukuran untuk menghimpun data sesuai rancangan; (5) mencatat seluruh data eksperimen sesuai rancangan; (6) menjaga kesetabilan kondisi lingkungan saat melakukan eksperimen; (7) mengamati dan mencatat data yang menunjukkan penyimpangan bila melakukan eksperimen.

Tabel 10. Membuat Inferensi

No.	Indikator
1.	Membedakan antara hasil observasi dengan rujukannya.
2.	Membuat generalisasi yang masuk akal berdasar hasil observasi.
3.	Mengkombinasikan observasi dan informasi yang diberikan untuk merumuskan hipotesis.
4.	Membuat deduksi dari hipotesis
5.	Menggunakan hasil observasi untuk mengkonfirmasi atau membuktikan kesalahan/menyangkal hipotesis yang ada.
6.	Memodifikasi hipotesis untuk mengakomodasi observasi/penelitian baru

Tabel 11. Membuat Prediksi

Indikator
1. Memprediksi perubahan kondisi objek.
2. Memprediksi perubahan bentuk objek.
3. Memprediksi perubahan objek dari segi tingkatan/ fase.

Tabel 12. Melaporkan Hasil Penyelidikan

No.	Indikator
1.	Menyajikan hasil suatu penyelidikan dalam bentuk diagram, tabel, ataupun grafik
2.	Menyajikan kesimpulan data hasil suatu penyelidikan
3.	Menyajikan pembahasan suatu hasil penyelidikan
4.	Menetapkan kebenaran hasil penyelidikan
5.	Menarik kesimpulan umum hasil penyelidikan berdasarkan hasil observasi dan generalisasi data/hasil analisis statistika.
6.	Menuangkan hasil, pembahasan, dan simpulan suatu penyelidikan dalam laporan secara tertulis

Tabel 13. Memaparkan Hasil Penyelidikan

Indikator
Memaparkan hasil penyelidikan, pembahasan dan kesimpulan hasil penyelidikan dalam suatu laporan secara lisan dalam suatu seminar/presentasi kelas.

Pengembangan butir tes berpedoman rumusan indikator. Setiap rumusan indikator dibuatkan minimal satu butir soal. Perangkat penilaian kemampuan inkuiri mata pelajaran fisika SMA seluruhnya berjumlah 93 butir tes dengan 9 butir tes digunakan sebagai *anchor items* yang dibagi 4 paket perangkat tes. Tabel 14 menyajikan distribusi butir tes pada setiap paket perangkat tes.

Tabel 14. Distribusi Butir Tes pada Setiap Perangkat Tes

PAKET	Nama Kode Butir	Jumlah butir
A	10, 11, 12, 13, 15, 16, 17, 18, 1, 2, 3, 4, 5, 6, 7, 8, 9 , 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30	30
B	31, 32, 33, 34, 35, 36, 37, 38, 39, 1, 2, 3, 4, 5, 6, 7, 8, 9 , 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51	30
C	52, 53, 54, 55, 56, 57, 58, 59, 60, 1, 2, 3, 4, 5, 6, 7, 8, 9 , 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72	30
D	73, 74, 75, 76, 77, 78, 79, 80, 81, 1, 2, 3, 4, 5, 6, 7, 8, 9 , 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93	30
Note:	No.1,2,3,4,5,6,7,8,9 sebagai <i>anchor items</i>	9

Hasil Uji Coba

Uji coba instrumen melibatkan 2015 responden yang berasal dari 10 SMA Negeri di DIY. Responden pada kegiatan uji coba meliputi peserta didik kelas X berjumlah 1100 peserta didik dan kelas XI-IPA berjumlah 915 peserta didik. Pelaksanaan ujian dirancang sedemikian rupa agar peserta didik yang duduk berdampingan mengerjakan tes dengan kode paket perangkat tes yang berbeda. Waktu yang disediakan untuk mengerjakan tes 90 menit. Pelaksanaan uji-

an melibatkan seluruh guru pengampu mata pelajaran fisika di setiap sekolah sebagai pengawas ujian.

Perangkat tes uji coba terdiri atas 4 paket soal. Selama pelaksanaan ujian keempat paket tes berkode A, B, C, dan D diujikan secara bersamaan di setiap kelas dan setiap peserta didik dalam satu kelas mendapat 1 paket tes untuk dikerjakan. Secara berturut-turut perangkat tes berkode A, B, C, dan D diujikan kepada 538, 508, 515, dan 454 responden.

Data respon peserta didik dalam penelitian ini diskor secara politomous dengan 4 kategori, yaitu kategori 4, 3, 2, dan 1. Kriteria pencapaian setiap tingkat kategori bila memenuhi persyaratan sebagai berikut.

- 4: Dapat menyebutkan 3 pilihan pernyataan jawaban yang disediakan secara benar, dan menyebutkan jawaban lainnya yang belum disediakan yang dapat dikategorikan jawaban benar
- 3: Dapat menyebutkan 3 pilihan pernyataan jawaban yang disediakan secara benar, dan tidak atau menyebutkan jawaban lainnya yang belum disediakan yang dapat dikategorikan jawaban salah.
Atau:
Dapat menyebutkan 2 pilihan pernyataan jawaban yang disediakan secara benar, dan menyebutkan jawaban lainnya yang belum disediakan yang dapat dikategorikan jawaban benar
- 2: Dapat menyebutkan 2 pilihan pernyataan jawaban yang disediakan secara benar, dan tidak atau menyebutkan jawaban lainnya yang belum disediakan yang dapat dikategorikan jawaban salah
Atau:
Dapat menyebutkan 1 pilihan pernyataan jawaban yang disediakan secara benar, dan menyebutkan jawaban lainnya yang belum disediakan yang dapat dikategorikan jawaban benar
- 1: Dapat menyebutkan 1 pilihan pernyataan jawaban yang disediakan secara benar, dan tidak atau menyebutkan jawaban lainnya yang belum disediakan yang dapat dikategorikan jawaban salah
Atau:
Tidak menyebutkan pilihan tetapi dapat menyebutkan jawaban lainnya yang belum disediakan yang dapat dikategorikan jawaban benar

Data politomous 4 kategori dianalisis menurut *Partial Credit Model (PCM)*. Hasil analisis kecocokan tes yang dilihat dari parameter *INFIT* untuk *Mean Square (MNSQ)*

menunjukkan bahwa instrumen penilaian kinerja kemampuan inkuiri mata pelajaran fisika SMA memenuhi kriteria *fit statistic* menurut *PCM* yang selengkapnya disajikan pada Tabel 5. Hasil analisis ke-93 butir tes memiliki nilai *INFIT MNSQ* antara 0,77 sampai dengan 1,3 yang artinya semua butir tes *fit* dengan *PCM*. Peta kecocokan butir tes dengan *PCM* disajikan pada Gambar 15.

Tabel 15. Parameter *Fit* Statistik Tes pada Level Peluang 0,50

No.	Parameter tes	Estimasi butir	Estimasi case
1.	<i>INFIT MNSQ</i>	1,10±0,09	1,11±0,08
2.	<i>OUTFIT MNSQ</i>	1,10±0,09	1,10±0,12
3.	Difficulty rerata	0,0±0,24	

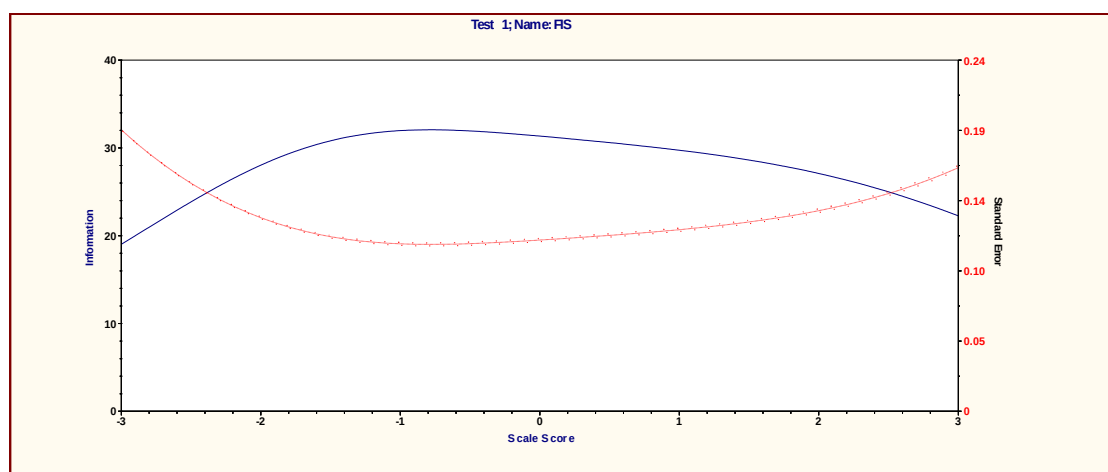
Spesifikasi butir perangkat tes memiliki *difficulty index* di antara nilai -0,83 (item4) dengan kategori mudah sampai dengan +0,84 (item2) dengan kategori sedang. *Difficulty index* rata-rata sebesar $0,0 \pm 0,24$ dengan kategori sedang. *Difficulty index* rata-rata aspek kemampuan inkuiri adalah sub-aspek perencanaan = 0,033; subaspek pelaksanaan = 0,02; dan subaspek pelaporan = -0,07 yang secara rinci disajikan pada Tabel 16.

Berdasarkan hasil analisis butir juga diperoleh kurva hubungan antara ability dengan *Total Information Function* (TIF) dan *Standard Error of Measurement* (SEM). Gambar 4 menyajikan kurva *Total Information*

Function dan *Standard Error of Measurement* instrumen penilaian kinerja kemampuan inkuiri mata pelajaran fisika SMA. Berdasarkan kurva tersebut diketahui bahwa instrumen penilaian kinerja kemampuan inkuiri mata pelajaran fisika SMA lebih tepat diujikan pada responden dengan kemampuan antara -2,5 sampai dengan +2,5. Karena instrumen ini dapat memberikan informasi yang tinggi terhadap kemampuan inkuiri peserta didik dan tingkat kesalahan pengukuran yang rendah bila diujikan pada responden yang mempunyai kemampuan (ability) antara -2,5 sampai dengan +2,5.

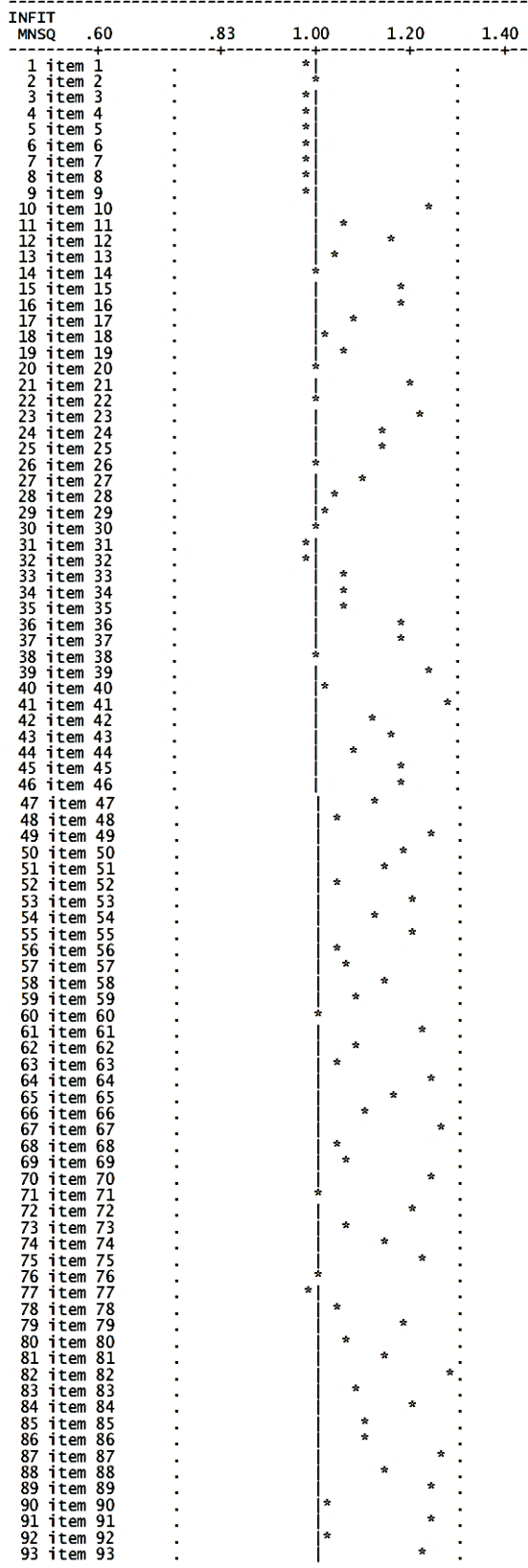
Tabel 16. Difficulty Rerata Aspek/subaspek Kemampuan Inkuiri

Aspek/Subaspek Kemampuan Inkuiri	Difficulty Rerata
PERENCANAAN	0,033
a. Merancang penyelidikan	0,09
b. Menyeleksi Prosedur	-0,15
PELAKSANAAN	0,02
a. Melakukan Pengamatan	-0,01
b. Merekam data/informasi	0,03
c. Mengikuti Intruksi	-0,18
d. Melakukan Pengukuran	0,14
e. Memanipulasi Gerakan	-0,13
f. Menimplementasikan prosedur/ teknik/penggunaan peralatan	-0,03
g. Melaksanakan Penyelidikan	0,23
PELAPORAN	-0,07
a. Membuat inferensi	-0,11
b. Membuat Prediksi	-0,24
c. Melaporkan hasil penyelidikan	0,06
d. Pengkomunikasian	0,15



Gambar 4. Kurva TIF dan SEM Instrumen Tes Kemampuan Inkuiri mata pelajaran fisika SMA.

QUEST: The Interactive Test Analysis System

Item Fit 13/ 6/13 8:20
all on all (N =2015 L = 93 Probability Level= .50)

Gambar 5. Peta Kecocokan Butir dengan Model PCM.

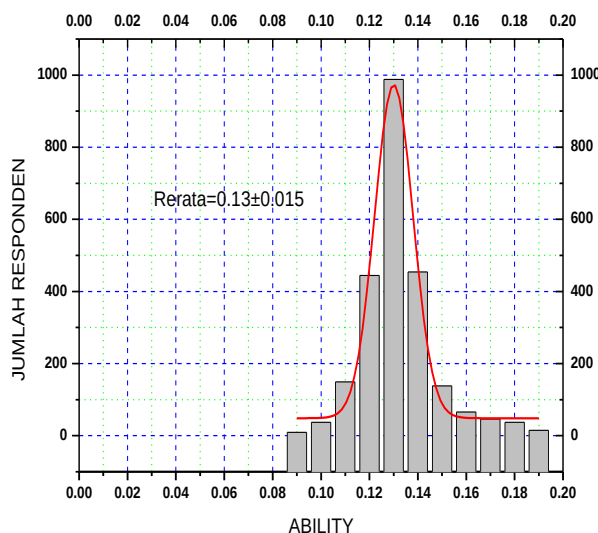
Hasil Pengukuran

Pengukuran kinerja kemampuan inkuiri mata pelajaran fisika SMA diikuti oleh 2383 responden. Responden terdiri atas peserta didik kelas X sejumlah 1269 responden dan peserta didik kelas XI-IPA sejumlah 1114 responden. Distribusi skor seluruh peserta didik di antara 0,1 sampai dengan 0,19 dalam skala logit antara -4 sampai dengan +4. Gambar 6 menyajikan distribusi skor dan rata-rata hasil pengukuran kemampuan inkuiri mata pelajaran fisika SMA di lokasi pengukuran dalam skala logit sebesar $0,13 \pm 0,015$. Artinya, peserta didik mempunyai kemampuan inkuiri di atas rata-rata.

Kemampuan inkuiri peserta didik dapat diestimasi menggunakan persentase rata-rata respon siswa menjawab benar pada setiap aspek dan subaspek yang diukur. Berdasarkan Tabel 17, diketahui bahwa respon siswa lebih dominan pada kategori 2 dan 3 dan hanya sebagian kecil persentase responden yang mampu mencapai kategori 4. Dengan kata lain, kemampuan inkuiri peserta didik masih ada yang belum memuaskan.

Tabel 17. Dominasi Respon Peserta Didik pada Setiap Aspek/Subaspek yang Diukur

ASPEK/SUBASPEK KEMAMPUAN INKUIRI	Dominasi Respon
PERENCANAAN	
a. Merancang penyelidikan	Kategori 2 3
b. Menyeleksi Prosedur	Kategori 3
PELAKSANAAN	
a. Melakukan Pengamatan	Kategori 2,3
b. Merekam data/informasi	Kategori 2,3
c. Mengikuti Intruksi	Kategori 3
d. Melakukan Pengukuran	Kategori 2 3
e. Memanipulasi Gerakan	Kategori 3
f. Mengimplementasikan prosedur/teknik/penggunaan peralatan	Kategori 2,3
g. Melaksanakan Penyelidikan	Kategori 2
PELAPORAN	
a. Membuat inferensi	Kategori 2
b. Membuat Prediksi	Kategori 3
c. Melaporkan hasil penyelidikan	Kategori 2,3
d. Pengkomunikasian hasil penyelidikan	Kategori 2,3



Gambar 6. Distribusi Skor Kemampuan Inkuiri Hasil Pengukuran.

Pembahasan

Perangkat penilaian kemampuan inkuiri mata pelajaran fisika SMA seluruhnya berjumlah 93 butir tes yang dibagi menjadi 4 paket perangkat tes, yakni perangkat tes I berkode A, tes II berkode B, tes III berkode C, dan tes IV berkode D. Setiap paket perangkat tes disusun berdasarkan pertimbangan keterwakilan setiap aspek dan sub-aspek kemampuan inkuiri yang akan diukur. Setiap paket perangkat tes terdiri atas 30 butir dengan 9 butir *anchor* termasuk di dalamnya. Jumlah butir *anchor* telah memenuhi jumlah minimal yang dipersyaratkan yaitu 20% dari total butir tes, dan masing-masing mewakili aspek yang diukur.

Instrumen penilaian kinerja kemampuan inkuiri mata pelajaran fisika SMA yang terdiri atas perangkat A, B, C, dan D kesemuanya memenuhi kriteria valid. Validitas isi instrumen diperoleh melalui *profesional judgement*, dan validitas secara empiris dibuktikan dengan *goodness of fit* menurut *patial credit model (PCM)*. Faktor-faktor pendukung sehingga instrumen yang dikembangkan memenuhi persyaratan validitas dapat diidentifikasi sebagai berikut. Pertama, butir-butir tes dikembangkan sesuai prosedur pengembangan instrumen penilaian. Kedua, butir tes dikembangkan dari indikator yang diturunkan dari setiap aspek yang diukur.

Ketiga, instrumen penilaian kinerja kemampuan inkuiri mata pelajaran fisika SMA telah divalidasi melalui *profesional judgement* dengan melibatkan sejumlah pakar pendidikan. Keempat, responden mengerjakan tes secara sungguh-sungguh di bawah pengawasan guru fisika di sekolahnya.

Fungsi informasi total tes relatif tinggi untuk kemampuan antara -2,5 sampai +2,5. Hal ini berarti bahwa instrumen yang dikembangkan juga memiliki reliabilitas yang tinggi karena tersusun dari butir-butir yang memiliki fungsi informasi yang tinggi dan butir-butir tes yang dikembangkan telah sesuai dengan kemampuan peserta didik yang diuji.

Index difficulty instrumen penilaian kinerja kemampuan inkuiri mata pelajaran fisika SMA yang dikembangkan pada penelitian ini bervariasi antara -2,0 sampai dengan +2,0, sehingga instrumen ini tergolong baik, karena telah memenuhi kriteria yang dipersyaratkan. Butir tes dengan indeks *difficulty* -2,0 dikategorikan butir sangat mudah, sedangkan indeks *difficulty* +2,0 dikategorikan butir sangat sulit. Dengan demikian, instrumen yang dikembangkan memenuhi persyaratan kategori baik.

Hasil pengukuran kemampuan inkuiri mata pelajaran fisika peserta didik SMA Negeri di DIY yang diikuti oleh 2383 responden menghasilkan distribusi skor antara 0,10 sampai dengan 0,19 dalam skala logit antara -4 sampai dengan +4. Rata-rata hasil pengukuran kemampuan inkuiri mata pelajaran fisika SMA sebesar $0,13 \pm 0,015$. Berdasarkan skor rata-rata kinerja kemampuan inkuiri mata pelajaran fisika peserta didik SMA di DIY diketahui bahwa peserta didik mempunyai kemampuan inkuiri di atas rata-rata.

Kemampuan inkuiri peserta didik yang diestimasi menggunakan persentase rata-rata respon siswa menjawab benar pada setiap aspek dan subaspek yang diukur dapat diketahui bahwa respon siswa dominan pada kategori 2 dan 3 dan hanya sebagian kecil persentase responden yang mampu mencapai kategori 4. Keadaan ini mengindikasikan bahwa peserta didik SMA sebagian

besar te-lah memiliki kemampuan inkuiri di atas rata-rata.

Simpulan dan Saran

Simpulan

Perangkat tes penilaian kinerja kemampuan inkuiri yang dikembangkan pada kemampuan perencanaan, pelaksanaan, dan pelaporan terdiri atas perangkat tes A, B, C, dan D yang setiap terdapat 30 butir soal dengan 9 *anchor items*.

Perangkat tes penilaian kinerja kemampuan inkuiri telah memenuhi validitas isi dengan *expert judgement* dan telah mendapatkan bukti empiris *fit* dengan partial credit model (PCM) berdasarkan data politomus empat kategori.

Seluruh butir soal pada perangkat tes penilaian kemampuan inkuiri dalam kriteria baik karena tingkat kesulitannya masih pada rentang antara -2,0 sampai dengan 2,0. Tingkat kesulitan tes paling sulit adalah butir-butir aspek kemampuan merencanakan, kemudian disusul aspek melaksanakan, dan pelaporan hasil penyelidikan.

Berdasarkan fungsi informasi total, perangkat tes penilaian kinerja kemampuan inkuiri mata pelajaran fisika SMA sangat tepat digunakan untuk mengukur peserta didik yang berkemampuan antara -2,5 sampai dengan 2,5. Perangkat tes penilaian kinerja kemampuan inkuiri dapat digunakan untuk mengukur kemampuan inkuiri peserta didik menurut partial credit model berdasar pada data politomus empat kategori.

Profil kemampuan inkuiri peserta didik pada mata pelajaran fisika SMA memiliki distribusi skor antara 0,10 sampai dengan 0,19 dengan rata-rata $0,13 \pm 0,148$ dari skala logit (*log odds unit*) antara +4 sampai dengan -4. Skor yang dicapai peserta didik sudah berada di atas kemampuan rata-rata, dan berada pada kriteria tinggi.

Profil kemampuan inkuiri peserta didik pada mata pelajaran fisika SMA pada aspek-aspek perencanaan, pelaksanaan, dan pelaporan dominan pada kategori-2 dan kategori-3 dari kategori-1 hingga kategori-4.

Artinya, kemampuan inkuiri mata pelajaran fisika SMA siswa berada di atas rata-rata.

Saran

Para guru fisika SMA hendaknya melakukan penilaian kemampuan inkuiri secara periodik di sekolahnya agar perkembangan kemampuan inkuiri peserta didik terukur. Bagi Dinas Pendidikan, perlu dilakukan pelatihan penyusunan instrumen penilaian kinerja kemampuan inkuiri mata pelajaran fisika bagi para guru. Peneliti lain dapat melakukan penelitian lebih lanjut agar kemampuan inkuiri peserta didik pada mata pelajaran fisika SMA dapat didiskripsikan secara komprehensif dengan responden yang meliputi SMA negeri dan Swasta.

Daftar Pustaka

- Adams, R.J. & Khoo, Siek-Toon. (1996). *Quest: The interactive test analysis system version 2.1*. Camberwell, Victoria: The Australian Council for Educational Research.
- Bryce, T.G.K, McCall, J., Mac Groger, J., Robertson, I.J. and Weston, R.A.J. (1990). *Techniques for assessing process skill in practical science: teacher's guide*. Oxford: Heinemann Educational Books.
- Chiappetta, E.L. & Koballa, T.R (2010). *Science instruction in the middle and secondary schools 7th ed*. Pearson Education, Inc.
- Depdikbud. (2014). Peraturan Menteri Pendidikan dan kebudayaan No. 104 Th 2014 Tentang Penilaian hasil belajar oleh pendidik pada pendidikan dasar dan pendidikan menengah.
- Glencoe Science. *Performance assessment in the classroom*. Glencoe McGraw-Hill, New York
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory*. Boston: KluwerNijhoff Publishing.

- Mardapi, D. (2012). *Pengukuran, penilaian, dan evaluasi pendidikan*. Yogyakarta: Nuha Litera.
- Moyer, R.H, Hackett, J.K, & Everett, S.A. (2007). *Teaching science as investigations modeling inquiry through learning cycle lessons*. New Jersey: Pearson, Merrill Prentice Hall.
- Oriundo, L.L. & Dallo-Antonio. (1998). *Evaluating educational outcomes (test, measurement, and evaluation)*, (5th ed.). Queson City: REX Printing Company, Inc.
- Rezba, R.J, Sparague, C.S., Fiel, R.L, Fuck, H.J., Okey, J.R., & Jaus, H.H. (1995). *Learning and assessing science process skill*. 3rd ed. Iowa: Kendall/Hunt Publishing Company.
- Ruiz-Primo, M.A. & Shavelson, R.J. (1996). *Rhetoric and Reality in science performance assessment: an update*. Journal of reseach and science teaching, Vol.33
- Subali, B. (2009). *Pengembangan tes pengukur keterampilan proses sains pola divergen mata pelajaran biologi SMA*. Prosiding Seminar Nasional Biologi, Lingkungan dan Pembelajaran, jurdik Biologi, FMIPA UNY, 4 Juli 2009, 581-593.
- Wenning, C.J. (2005). *Implementing inquiry-based instruction in the science classroom*. A new model for solving the improvement of practice problem. *J. Phys. Tchr. Educ. Online* 2(4), May 2005.
- Wenning, C.J. (2005). Levels of inquiry: Hierarchies of pedagogical practices and inquiry processes. *JPTEO, J. Phys. Tchr. Educ. Online* 2(3), February 2005.

**EVALUASI IMPLEMENTASI PKP DALAM PEMBELAJARAN BIOLOGI
DI SMAN KOTA SEMARANG MENGGUNAKAN PENDEKATAN MIXED-METHOD**

¹⁾Wini Isnaeni, ²⁾Kumaidi

²⁾Universitas Negeri Semarang, ²⁾Universitas Muhammadiyah Surakarta

¹⁾wi2isna@yahoo.co.id, ²⁾kuma_426@yahoo.com.

Abstrak

Penelitian ini bertujuan mengevaluasi secara komprehensif tentang implementasi pendekatan keterampilan proses (PKP) dalam pembelajaran biologi di SMAN Kota Semarang, dan mengungkap berbagai faktor pendukung dan penghambatnya. Penelitian ini menggunakan model evaluasi kesenjangan. Data dikumpulkan dengan pendekatan *mixed-method*, diawali dengan penggunaan metode kuantitatif dilanjutkan dengan metode kualitatif. Data kuantitatif tingkat penggunaan keterampilan proses oleh peserta didik dikumpulkan menggunakan kuesioner. Kuesioner dikembangkan dengan memodifikasi langkah Madaus *et al.*, Claeys *et al.*, dan Wheeler, melalui tahap review literatur, mengkonstruksi instrumen, menetapkan validitas isi dengan cara validasi pakar, dan uji coba lapangan. Data kualitatif tentang proses pembelajaran biologi dan data kualitatif lainnya, diperoleh melalui observasi, pembuatan catatan lapangan, kategorisasi catatan lapangan, dan wawancara mendalam. Seluruh data dianalisis dengan *explanatory mixed-method analysis*. Hasil penelitian menunjukkan bahwa kualitas implementasi PKP dalam pembelajaran biologi di SMAN kota Semarang belum mencapai kategori baik. Faktor pendukung dan penghambat utamanya adalah tinggi dan rendahnya komitmen guru untuk mengimplementasikan PKP dalam pembelajaran biologi.

Kata kunci: *evaluasi implementasi kurikulum, pendekatan keterampilan proses, pembelajaran biologi, mixed-method*

**AN EVALUATION OF THE PSA IMPLEMENTATION IN BIOLOGY TEACHING AT
STATE HIGH SCHOOLS IN SEMARANG CITY
USING THE MIXED-METHOD APPROACH**

Wini Isnaeni, Kumaidi

Universitas Negeri Semarang, Universitas Muhammadiyah Surakarta

wi2isna@yahoo.co.id, kuma_426@yahoo.com.

Abstract

This study aims to evaluate the implementation of the process skill approach (PSA) in the biology teaching at state high schools in Semarang City, and reveal the various supporting factors and the constraints. The study used a discrepancy evaluation model. The data were collected using a mixed-methods approach, initiated with the use of the quantitative method followed by the use the qualitative method. The quantitative data about the level of the process-skill use by the students were collected using questionnaire. The questionnaire was developed by referring to the steps of Madaus *et al.*, Claeys *et al.*, and Wheeler, through reviewing literature, constructing instruments, and determining the content validity through expert validation, and conducting field trials. The qualitative data about the biology teaching process and the other qualitative data were collected through observation, the preparation of field notes and categorization of field notes, and in depth interview. The whole data were analyzed using the explanatory mixed-method analysis. The result shows that the quality of the implementation of the biology curriculum at state high schools in Semarang City is not in a good category yet. The main supporting factors and the constraints are the the teachers' high and low commitment to implement the PSA in the biology teaching.

Keywords: *evaluation of curriculum implementation, process skill approach, biology learning, mixed-method*

Pendahuluan

Pendekatan keterampilan proses sains (PKP) merupakan pendekatan khas untuk membelajarkan biologi, seperti yang telah dikemukakan dalam kurikulum biologi. Kurikulum adalah seperangkat rencana dan pengaturan mengenai tujuan, isi, bahan pelajaran, sekaligus sebagai pedoman penyelenggaraan kegiatan pembelajaran untuk mencapai tujuan pendidikan tertentu (BSNP, 2006, p.3). Kurikulum yang diimplementasikan, perlu dipantau agar tingkat kesesuaian antara rencana dan pelaksanaannya dapat diketahui (Alfiero *et al.*, 2008, p.13). Evaluasi terhadap implementasi kurikulum juga bermanfaat untuk mengontrol kualitas kurikulum itu.

Perubahan kurikulum di Indonesia sudah dilakukan beberapa kali, namun mutu pendidikan di Indonesia (khususnya untuk bidang sains) masih tetap rendah. Hal ini dapat diketahui dari hasil studi TIMSS dan hasil asesmen PISA, yang menggambarkan performansi peserta didik Indonesia yang sangat rendah di tingkat internasional. Hasil studi TIMSS tahun 1999 menunjukkan bahwa tingkat penguasaan IPA oleh anak-anak Indonesia pada tingkat internasional menempati urutan ke-32 dari 38 negara peserta (Hayat, 2006, p.1). Hasil asesmen PISA tahun 2009 menunjukkan bahwa performansi peserta didik Indonesia (berusia 15 tahun) pada bidang sains menempati urutan ke-57 dari 65 negara peserta (OECD, 2010, p.8).

Hasil asesmen tersebut menggambarkan bahwa perubahan kurikulum di Indonesia dalam kurun waktu 10 tahun terakhir (tahun 1999-2009) ternyata belum dapat menghasilkan lulusan yang mampu bersaing di tingkat internasional. Mengapa perubahan kurikulum selama 10 tahun tersebut belum mampu meningkatkan performansi atau penguasaan sains oleh anak Indonesia?

Kurikulum merupakan jantungnya dunia pendidikan (Puskur, 2007, p.1). Perubahan kurikulum dapat dipandang sebagai cara yang tepat untuk meningkatkan mutu pendidikan di Indonesia. Kunci utama untuk mencapai keberhasilan dalam mengupa-

yaikan peningkatan kualitas program pendidikan ialah penyelenggaraan proses pembelajaran yang berkualitas (Widoyoko, 2008, p.1). Pendapat tersebut sesuai dengan pernyataan Djohar (2003, p.3) bahwa perbaikan kurikulum harus dibarengi dengan perbaikan praksis pembelajaran. Perubahan kurikulum tanpa disertai perubahan praksis pembelajaran tidak akan menghasilkan apa-apa. Uraian tersebut dapat ditegaskan bahwa implementasi kurikulum Biologi harus dilaksanakan dengan menyelenggarakan proses pembelajaran yang optimal, dalam hal ini harus sesuai tuntutan standar proses yang ditetapkan BSNP.

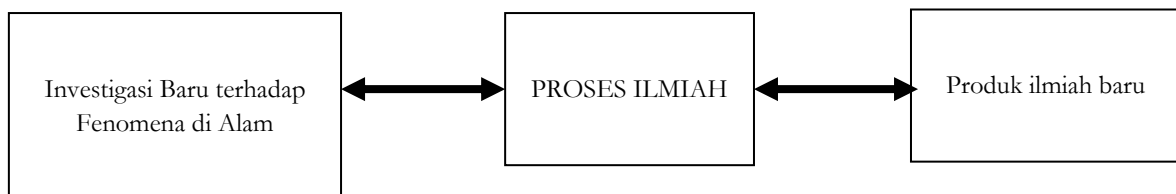
Berkaitan dengan hal itu, upaya meningkatkan kualitas pendidikan harus dimulai dengan meningkatkan kualitas proses pembelajaran. Pembelajaran adalah proses interaksi antara peserta didik, pendidik, dan sumber belajar pada suatu lingkungan belajar (pasal 1 ayat 20 Undang-Undang Sisdiknas Nomor 20 Tahun 2003). Berdasar kepada uraian tersebut maka jelas bahwa proses pembelajaran biologi yang berkualitas merupakan kunci keberhasilan implementasi kurikulum biologi. Oleh karena itu perlu dilakukan evaluasi terhadap implementasi kurikulum biologi, agar kualitas atau efektivitas implementasinya dapat diketahui. Untuk mewujudkan kegiatan pembelajaran yang efektif dan efisien, implementasi kurikulum harus mengacu kepada standar proses (Permendiknas Nomor 41 Tahun 2007 & No. 22 tahun 2006)

Proses pembelajaran biologi yang dikehendaki adalah pembelajaran yang sesuai dengan karakteristik sains/biologi dan karakteristik pendidikan sains (Permendiknas Nomor 41 Tahun 2007), yaitu proses pembelajaran aktif berbasis *exploring* dan *inquiring* (pembelajaran yang berorientasi kepada eksplorasi dan inkuiri). Pembelajaran berorientasi inkuiri dideskripsikan sebagai proses aktif yang melibatkan berbagai kegiatan seperti observasi, merancang investigasi, juga menggunakan alat untuk mengumpulkan, menganalisis, dan menginterpretasikan data, mengajukan pertanyaan, dan membuat prediksi serta mengkomunikasikan hasil

(Smith *et.al.*, 2007, p.170). Pembelajaran seperti yang dimaksud secara luas disebut sebagai pembelajaran yang menerapkan pendekatan keterampilan proses.

Keterampilan proses dapat digolongkan menjadi 2, yaitu keterampilan dasar/*basic skills* dan keterampilan terintegrasi/*integrated skills* (Arena, 1996 dalam Martin *et.al.*, 2005, pp.17-20). Keterampilan dasar meliputi keterampilan observasi, klasifikasi, komunikasi, pengukuran, estimasi, prediksi dan interferensi. Keterampilan terintegrasi/terpadu meliputi keterampilan untuk mengidentifikasi variabel, mengontrol variabel, membuat definisi operasional, membuat hipotesis, melakukan eksperimen, membuat grafik, menginterpretasikan data, pemodelan, dan melakukan penyelidikan.

Melakukan penyelidikan/investigasi berkaitan dengan perolehan produk ilmu baru. Pola saling hubungan antara kegiatan investigasi (terhadap gejala/fenomena di alam), proses ilmiah, dan perolehan produk ilmiah dinyatakan oleh Carin & Sund (1989, p.6) seperti pada Gambar 1. Berdasarkan kepada Gambar 1 dapat dinyatakan bahwa pelaksanaan pembelajaran biologi dengan menekankan penggunaan metode ilmiah atau pendekatan keterampilan proses (PKP) akan memberi peluang kepada peserta didik untuk mengembangkan ilmu yang dimilikinya, dengan menghasilkan produk baru. Oleh karena itu, pengembangan keterampilan proses sains pada peserta didik harus benar-benar dapat dilaksanakan dalam pembelajaran biologi.



Gambar 1. Gambaran Ringkas tentang Saling Hubungan antara Investigasi Gejala, Proses Ilmiah, dan Perolehan Produk Ilmiah (Dimodifikasi dari Carin & Sund, 1989, p.6).

Evaluasi dapat menjadi upaya untuk memperbaiki bagian program yang diselenggarakan dengan kurang baik (Worthen & Sanders, 1981). Evaluasi juga menggunakan metode penilaian yang memerlukan informasi/deskripsi kuantitatif (misalnya skor tes hasil pengukuran) dan informasi/deskripsi kualitatif (misalnya catatan tentang perilaku peserta didik dan pendidik/guru dalam pembelajaran; Setiadi dkk, 2006, p.69). Evaluasi mengandung dua pengertian, yaitu deskripsi kualitatif dari perilaku peserta didik dan deskripsi kuantitatif dari hasil pengukuran (misalnya skor tes). Jadi, evaluasi mencakup pengukuran dan bersifat lebih komprehensif. Hal serupa juga dikemukakan oleh Abd. Rahman A. Ghani (dalam Setiadi dkk., 2006, p.69) yang menjelaskan bahwa evaluasi adalah metode penilaian yang tidak hanya tergantung kepada pengukuran.

Mengingat bahwa evaluasi memerlukan informasi kuantitatif dan kualitatif sekaligus maka dalam evaluasi perlu diterapkan gabungan antara metode kuantitatif dan kualitatif sekaligus (menggunakan pendekatan *mixed-method*). Penggunaan metode evaluasi kuantitatif dalam proses evaluasi (misalnya dalam evaluasi kurikulum) diyakini mampu menyajikan berbagai data yang terkait dengan indikator yang dapat menggambarkan tingkat keterlaksanaan kurikulum yang sedang berlaku, namun tidak mampu menjelaskan berbagai hal yang terjadi di balik suatu gejala/peristiwa, misalnya alasan yang melatarbelakangi seorang guru dalam memilih strategi pembelajaran tertentu. Hal yang tidak terungkap tersebut merupakan informasi kualitatif, yang berperan sangat penting dalam menjelaskan informasi kuantitatif. Hal ini sesuai dengan pernyataan Brannen (2005, p.85) yang mengemukakan

bahwa penelitian kualitatif dapat membantu penelitian kuantitatif, antara lain dengan memberikan informasi dasar tentang konteks dan subjek. Penelitian/data kualitatif juga dapat membantu interpretasi hubungan antar variabel, dan juga dapat digunakan untuk menjelaskan faktor-faktor yang mendasari hubungan yang terbangun. Dengan demikian, maka jelas bahwa dalam evaluasi, tidak terkecuali dalam evaluasi implementasi PKP dalam pembelajaran biologi, sangat diperlukan informasi kuantitatif dan kualitatif bersama-sama (menggunakan pendekatan *mixed-method*).

Berdasar uraian tersebut maka penelitian ini bertujuan untuk: (1) mengungkap sejauh mana kualitas implementasi PKP dalam pembelajaran biologi di SMA Negeri Kota Semarang yang dikaji dengan pendekatan *mixed-method*; (2) mengungkap berbagai faktor pendukung dan penghambat untuk mengimplementasikan PKP dalam pembelajaran biologi. Manfaat dari penelitian ini ialah memberikan informasi komprehensif kepada sekolah dan masyarakat tentang kualitas pembelajaran biologi yang diselenggarakan di SMA Negeri Kota Semarang, serta berbagai faktor pendukung dan penghambatnya.

Metode Penelitian

Penelitian ini menggunakan model evaluasi kesenjangan (*discrepancy*) (Madaus *et al.*, 1986, p.79; Steinmetz, 2000, p.143; Stufflebeam *et al.*, 2002, p.127; Arikunto & Jabar, 2004, p.31), dan menggunakan metode gabungan antara metode kuantitatif dan metode kualitatif atau pendekatan *mixed-method*, tepatnya *explanatory mixed-method design* (Creswell, 2008, p.552).

Penelitian dilaksanakan di SMA Negeri Kota Semarang, pada bulan September 2012 sampai dengan Januari 2013, yaitu pada SMAN AA, dan SMAN AB, SMAN AC, SMAN AD, SMAN AE dan SMA AF di Semarang (nama-nama SMA disamarkan).

Subjek penelitian ini ialah guru dan peserta didik yang terlibat dalam proses pembelajaran Biologi yang diselenggarakan pada SMA negeri di kota Semarang yang

dipilih dengan pertimbangan tertentu. Objek penelitian ini adalah aktivitas peserta didik dan guru, serta proses pembelajaran biologi yang dipilih untuk diobservasi dengan pertimbangan tertentu.

Langkah pertama penelitian ini ialah mengembangkan kuesioner. Kuesioner dikembangkan dengan memodifikasi langkah yang dilakukan Madaus *et al.* (1986, p.83), Claeys *et al.* (2012, p.577), dan Wheeler (2012, p.43), melalui tahap review literatur, mengkonstruksi instrumen, membuktikan validitas isi melalui kesepakatan ahli, dan uji coba lapangan. Uji coba lapangan dilakukan dengan melibatkan jumlah peserta didik yang makin meningkat dari uji coba pertama, kedua, dan ketiga. Data hasil uji coba lapangan dianalisis menggunakan *exploratory factor analysis* (EFA), sampai diperoleh kuesioner final yang layak. Dari hasil uji coba yang ketiga dapat diperoleh kuesioner yang valid dan reliabel. Berdasarkan hasil EFA diketahui dari nilai total kumulatif varian $\geq 50\%$, dan nilai komponen matrik rotasi per butir $\geq 0,3$. Kisaran nilai komponen matrik rotasi per butir untuk keterampilan dasar adalah antara 0,493 sampai 0,904, sedangkan untuk keterampilan terpadu kisarannya antara 0,476-0,837. Reliabilitas instrumen diketahui dari nilai α -Cronbach untuk dimensi keterampilan dasar dan terpadu masing-masing 0,670 dan 0,775. Instrumen tersebut selanjutnya digunakan sebagai alat pengumpul data dalam penelitian ini.

Penelitian diawali dengan pengumpulan dan analisis data kuantitatif, dilanjutkan dengan pengumpulan dan analisis data kualitatif. Hasilnya digunakan untuk memberikan penjelasan/eksplanasi lebih lanjut mengenai fenomena yang belum dapat dijelaskan oleh hasil penelitian kuantitatif.

Jenis data, kategori data, teknik pengumpulan data, instrumen yang digunakan, dan sumber data disajikan pada Tabel 1. Data yang dikumpulkan mencakup data kuantitatif dan data kualitatif. Data kuantitatif berupa kadar inkuiri atau tingkat implementasi keterampilan proses dikumpulkan menggunakan kuesioner. Data ini

menunjukkan tingkat penggunaan keterampilan dasar dan keterampilan terpadu oleh peserta didik. Data kualitatif tentang pelaksanaan proses pembelajaran biologi dikumpulkan melalui proses observasi, pembuatan catatan lapangan dan kategorisasi catatan lapangan, serta wawancara mendalam dengan responden (guru dan peserta didik).

Catatan lapangan adalah dokumentasi tertulis tentang berbagai peristiwa yang ter-

jadi pada proses pembelajaran biologi, diperoleh dengan cara mengobservasi proses pembelajaran biologi yang dilaksanakan oleh para guru yang dipilih dengan pertimbangan tertentu (secara purposif). Pembuatan catatan lapangan dilakukan selama penyelenggaraan proses pembelajaran biologi untuk kompetensi dasar atau materi ajar tertentu.

Tabel 1. Jenis dan Kategori Data/Informasi yang Dikumpulkan, Teknik Pengumpulan Data, Instrumen, dan Sumber Data pada Penelitian Ini

No	Jenis dan kategori data	Teknik dan metode pengumpulan data	Instrumen	Sumber data
1	Tingkat implementasi PKP dalam pembelajaran biologi selama 1 semester-data kuantitatif	Kuesioner (kuantitatif)	Kuesioner	Peserta didik kelas XI IPA
2	Proses pembelajaran biologi untuk KD tertentu - data kualitatif	Observasi (kualitatif)	Lembar untuk catatan lapangan	Proses pembelajaran biologi
3	Tingkat implementasi PKP dalam PBM biologi pada KD terobservasi didasarkan catatan lapangan - data kuantifikasi	Kategorisasi catatan lapangan (kuantifikasi)	Lembar kategorisasi	Catatan lapangan
4	Tanggapan guru PBM biologi yang dilaksanakan (untuk KD tertentu) - data kualitatif dan kuantifikasi	Wawancara (kualitatif) kuesioner (kuantitatif/ kuantifikasi)	Panduan wawancara kuesioner	Guru biologi terpilih Guru anggota MGMP biologi

Seluruh data yang diperoleh dianalisis dengan *explanatory mixed-method analysis*, (Creswell, 2008, p.557), diawali dengan analisis data kuantitatif menggunakan teknik kuantitatif deskriptif, dilanjutkan dengan eksplanasi dan elaborasi terhadap hasil temuan kuantitatif tersebut menggunakan kumpulan data kualitatif yang telah diperoleh dengan teknik-teknik kualitatif.

Hasil Penelitian dan Pembahasan

Hasil penelitian ini terdiri dari dua kategori, yaitu hasil kuantitatif dan hasil kualitatif. Hasil penelitian kuantitatif pada

penelitian ini menginformasikan tingkat implementasi PKP pada berbagai SMA negeri di Kota Semarang. Data ini menggambarkan tingkat penggunaan keterampilan dasar dan keterampilan terpadu yang digunakan peserta didik selama belajar biologi. Data kuantitatif menunjukkan bahwa tingkat implementasi PKP dalam pembelajaran biologi di kelas XI IPA semester gasal adalah 72% (disajikan pada Tabel 2). Dalam hal ini, peserta didik menggunakan keterampilan dasar masing-masing $(71:2)\% = 35,5\%$ dan keterampilan terpadu sebesar $(73:2)\% = 36,5\%$.

Tabel 2. Tingkat Implementasi PKP dalam Proses Pembelajaran Biologi pada Berbagai SMAN di Kota Semarang pada Kelas XI Semester Gasal.

Dimensi dan Sub Dimensi	Persentase Tingkat implementasi PKP dalam pembelajaran Biologi yang diselenggarakan di SMA						Rerata dalam persen
	AD*	AE*	AA*	AB *	AC*	AF*	
Keterampilan Dasar (Pildas)							
Observasi	55	73	40	74	75	100	71
Mengukur	41	89	65	55	57	75	62
Mengklasifikasikan	68	95	72	91	96	100	89
Mengkomunikasikan	90	57	82	85	86	68	80
Memperkirakan	31	72	58	63	45	61	56
Menyimpulkan	54	85	70	83	85	33	72
Penggunaan Pildas/SMA	59	79	65	75	74	73	71 **
Keterampilan Terpadu (Pilter)							
Merencanakan penelit.	58	85	78	81	79	89	79
Merumuskan hipotesis	57	93	72	87	82	100	82
Melakukan penelitian/penyelidikan	39	99	84	95	93	100	74
Menginterpretasikan data	51	98	94	91	90	100	89
Mempresentasikan hasil penelitian/hasil kajian teoretik	81	42	49	84	59	55	64
Penggunaan Pilter/SMA	57	84	75	88	80	89	73 **
Rerata persen penggunaan keterampilan proses per SMA ***	58	81	70	81	78	81	
Rerata persen tingkat implementasi PKP ****	$(71 + 73) : 2 = 72\%$						

Keterangan:

- * : AD: SMAN AD Smg; AE: SMAN AE Smg. AA: SMAN AA Smg; AB: SMAN AB Smg; AC: SMAN AC Smg. AF: SMA AF Smg
- ** : Rerata persen penggunaan keterampilan dasar atau keterampilan terpadu dalam pembelajaran biologi untuk seluruh sekolah yang diobservasi.
- *** : Rerata persen penggunaan keterampilan proses sains dalam pembelajaran biologi untuk seluruh sekolah yang diobservasi.
- **** : Tingkat implementasi PKP 50-70%: cukup; tingkat implementasi > 70%: baik.

Berdasarkan data pada Tabel 2 (baris kedua dari bawah), dapat diketahui bahwa tingkat implementasi PKP di SMA Negeri Kota Semarang relatif bervariasi. Mengacu kepada keterangan tentang kategori tingkat implementasi PKP yang disajikan pada bagian bawah Tabel 2, dapat diketahui bahwa tingkat implementasi PKP dalam pembelajaran biologi pada SMA Negeri di Kota Semarang dapat dikelompokkan menjadi dua kategori, yaitu kategori cukup dan baik. Sekolah yang menunjukkan tingkat implementasi PKP paling rendah ialah SMA Negeri AD (58%), diikuti oleh SMA Negeri AA yang tingkat implementasinya mencapai

70%. Kedua sekolah tersebut menunjukkan tingkat implementasi PKP sebesar $\leq 70\%$, sehingga dapat dinyatakan mencapai kategori cukup. Mengingat bahwa rata-rata tingkat implementasi PKP di Kota Semarang adalah 72% maka dapat dinyatakan bahwa tingkat implementasi PKP dalam pembelajaran biologi di SMA Negeri Kota Semarang secara kuantitatif termasuk kategori baik.

Hasil temuan kuantitatif tersebut perlu dijelaskan dan dielaborasi antara lain menggunakan data kualitatif berupa catatan lapangan. Gambaran tentang isi catatan lapangan secara umum disajikan pada Tabel 3.

Tabel 3. Gambaran Umum Isi Catatan Lapangan Hasil Observasi terhadap Proses Pembelajaran Biologi yang Diselenggarakan di Kelas XI IPA pada Semester Gasal pada SMA Negeri di Kota Semarang

No	Aspek yang diobservasi	Hasil Observasi
1.	Pelaksanaan pembelajaran biologi secara umum	Kegiatan pembelajaran biologi di antara sekolah yang satu dan yang lain sangat bervariasi, namun pada umumnya masih kurang mencerminkan adanya kegiatan yang melibatkan peserta didik dalam penggunaan keterampilan proses sains, terutama sekali dalam hal penggunaan keterampilan terpadu. Jenis keterampilan dasar yang digunakan peserta didik relatif masih sangat terbatas. Kegiatan pembelajaran biologi masih didominasi oleh aktivitas komunikasi (mengkomunikasikan), baik dalam bentuk komunikasi lisan maupun komunikasi tertulis. Kegiatan observasi objek studi berupa benda asli masih sangat terbatas.
2.	Peran guru dalam pembelajaran	Peran guru dalam pembelajaran paling dominan sebagai narasumber utama. Dalam diskusi/presentasi, guru berperan sebagai fasilitator, narasumber, dan evaluator, juga sebagai konsultan. Dalam kegiatan praktikum, guru memainkan sangat banyak peran, yaitu sebagai fasilitator, pembimbing, narasumber, laboran/teknisi laboratorium. Selama observasi, ada guru yang tidak menyelenggarakan praktikum sekalipun, dan dalam keadaan demikian, peserta didik menjadi pasif, dengan aktivitas mendengar dan mencatat.
3.	Metode/strategi pembelajaran yang digunakan	Penggunaan metode ilmiah (keterampilan dasar dan terpadu) yang cukup menonjol, ditemukan pada sekolah yang berlokasi di daerah marginal.
4.	Objek studi yang diobservasi	Objek studi yang diobservasi kebanyakan berupa objek yang dihadirkan kedalam kelas dengan bantuan IT, atau berupa objek tiruan/torso/model, karena materi yang dikaji selama observasi, adalah tentang sistem gerak, sirkulasi, dan pencernaan makanan.
5.	Keterlibatan peserta didik dalam pembelajaran	Dalam kegiatan praktikum peserta didik melakukan pengamatan susunan otot dan rangka, pengujian darah (uji kecepatan pembekuan darah dan atau uji golongan darah), serta uji kandungan zat dalam bahan makanan, dan memnulis laporan praktikum. Diskusi (kelas/kelompok), dan presentasi lisan.

Berdasarkan data pada Tabel 3 dapat diperoleh informasi bahwa kegiatan pembelajaran biologi di SMA negeri Kota Semarang masih didominasi oleh praktik pembelajaran yang berorientasi materi (pembelajaran tekstual). Dalam kegiatan pembelajaran yang dimaksud, guru lebih banyak berperan sebagai narasumber utama. Kegiatan pembelajaran biologi masih didominasi oleh aktivitas komunikasi (mengkomunikasikan), baik dalam bentuk komunikasi lisan maupun komunikasi tertulis, sementara kegiatan observasi objek studi berupa benda asli masih sangat terbatas.

Hasil penelitian kualitatif lainnya ialah hasil wawancara dengan guru mata pelajaran biologi, yang menginformasikan persepsi mereka mengenai implementasi PKP dalam pembelajaran biologi. Hasil wawancara

memberikan informasi bahwa guru tidak dapat melaksanakan kegiatan pembelajaran biologi yang sepenuhnya sesuai dengan rencana (yang tertulis dalam RPP). Dalam hal ini guru memperkirakan tingkat keterlaksanaan rencana yang mereka buat secara bervariasi, ada yang memperkirakan sekitar 50-75% (berarti nilai tengahnya adalah 62,5%), ada yang memperkirakan 75%, dan ada juga yang memperkirakan 90%. Jika diambil titik tengahnya dan dirata-ratakan maka rerata tingkat implementasi PKP dalam pembelajaran biologi menurut perkiraan guru adalah $(62,5+75+90):3=79\%$. Angka 79% ini tidak terlalu jauh berbeda dari tingkat implementasi PKP yang tercantum dalam Tabel 3, sebesar 72%. Apabila keduanya dirata-ratakan maka diperoleh angka 75%.

Keterlaksanaan implementasi PKP dalam pembelajaran biologi baik sebesar 72% maupun 75% menunjukkan bahwa PKP tidak diimplementasikan secara maksimal. Hal ini juga memberikan gambaran proses pembelajaran biologi tidak diselenggarakan secara maksimal. Padahal proses pembelajaran yang berkualitas merupakan kunci utama dalam upaya peningkatan kualitas berbagai program pendidikan (Widoyoko, 2008, p.1). Implementasi PKP yang tidak maksimal dalam pembelajaran biologi juga menggambarkan kurang adanya perbaikan praksis pembelajaran. Sesuai dengan pendapat Djohar (2003, p.3) keadaan semacam itu menyebabkan tidak diperolehnya hasil apapun dari perubahan kurikulum yang dilakukan. Hal itu juga menggambarkan praktik pembelajaran yang tidak sesuai dengan standar proses seperti yang ditetapkan dalam Permendiknas No. 41 Tahun 2007 dan No. 22 Tahun 2006).

Ketidakmaksimalan implementasi PKP dalam pembelajaran biologi disebabkan oleh beberapa alasan, seperti yang diemukakan para guru berikut ini.

... Tugas guru yang terlalu padat, dengan kewajiban mengajar 24 jam per minggu, di luar tugas tambahan lainnya seperti menjadi wali kelas, pembina kegiatan tertentu dan lain-lain. Hal lain yang menjadi kendala untuk menyelenggarakan pembelajaran dengan menerapkan PKP ialah banyaknya materi yang harus disampaikan. Kadang guru merasa seperti "dituntut" untuk menyelesaikan materi pada batas waktu tertentu. Jika dilaksanakan praktikum, akan perlu waktu lebih banyak lagi, maka pelaksanaan praktikum dibatasi ...

... Keterlaksanaannya paling-paling sekitar 75%. Jujur saja, "penjejalan materi" masih berlaku di sekolah ini. Pokoknya materi harus terselesaikan. Kalau dipikir-pikir, dalam KTSP seharusnya tidak perlu ada ujian nasional untuk menentukan kelulusan peserta didik, tetapi cukup dengan ujian sekolah saja. Namun karena masih ada UN (Ujian Nasional), penerapan pendekatan keterampilan proses ya terpaksa diabaikan. Bagaimanapun juga, selama ada UN penyampaian

materi harus benar-benar selesai (terselesaikan semua). Jadi, banyak sedikit, UN itu "mengganggu" penerapan pendekatan keterampilan proses sains dalam pembelajaran biologi...

...Faktor penghambat yang utama ya UN. Di satu sisi, UN memang mampu menyemangati guru, peserta didik, maupun orang tua untuk mencapai tingkat yang lebih baik. Saya setuju kalau UN memang tetap diperlukan untuk menyamakan perkembangan peserta didik di suatu sekolah dengan perkembangan peserta didik di sekolah lain... (peneliti: maksudnya, supaya perkembangan atau kemajuan belajar peserta didik diberbagai sekolah dapat disejajarkan).

Beberapa jawaban guru seperti tersebut, tampak bahwa pembelajaran biologi yang diselenggarakan di SMA negeri Kota Semarang masih berorientasi materi atau bersifat tekstual. Pembelajaran tekstual sebenarnya tidak sesuai dengan karakteristik biologi dan karakteristik pembelajaran biologi. Karakteristik pendidikan sains (termasuk biologi) yaitu pembelajaran aktif berbasis *exploring* dan *inquiring* (berorientasi eksplorasi dan inkuiri; Permendiknas No. 41 tahun 2007). Meskipun demikian, sebagian besar responden berpendapat bahwa mengkaji materi secara klasikal merupakan hal yang penting, karena materi biologi sangat banyak. Dalam keadaan demikian, guru berpendapat bahwa mengkaji materi hingga tuntas adalah hal yang lebih penting daripada mengimplementasikan PKP. Hal ini dipilih untuk dilakukan guru biologi dengan tujuan agar peserta didik memperoleh kesempatan lebih baik dalam mempersiapkan diri menghadapi UN.

Apabila dikaji secara lebih seksama, sikap guru seperti tersebut pada dasarnya memperlihatkan adanya rasa khawatir pada guru, kalau peserta didik tidak melakukan persiapan diri dengan baik dalam menghadapi UN. Dengan demikian, maka dapat dinyatakan bahwa UN sebenarnya bukan faktor penghambat implementasi PKP dalam pembelajaran biologi. Faktor penghambat yang sebenarnya adalah adanya rasa khawatir pada guru (sebagian guru) dalam

menghadapi UN. Pendapat peneliti ini didukung oleh pernyataan responden yang tidak ingin terjebak dalam pembelajaran tekstual dan tak perlu khawatir dalam menghadapi UN seperti berikut ini.

...Saya tidak mau terjebak pada proses pembelajaran tekstual. Yang terpenting, pembelajaran dengan PKP harus tetap jalan (terselenggara) meskipun materi ajar tidak dapat disampaikan semuanya. Saya tidak perlu mengkhawatirkan hal itu (peneliti: UN), karena nyatanya peserta didik tetap dapat mengerjakan soal-soal UN dengan baik (peneliti: maksudnya, peserta didik tetap dapat memperoleh nilai UN yang tinggi). Saya tidak peduli UN, karena UN hanya untuk peserta didik kelas XII, dan sudah ada media tersendiri untuk mempersiapkan mereka. Tetapi saya tahu, teman-teman pengampu mata pelajaran lain yang tampak masih sangat dipengaruhi UN dalam menyelenggarakan pembelajaran...

Berdasar kepada uraian di atas maka dapat dinyatakan bahwa yang menjadi faktor penghambat dalam implementasi PKP dalam pembelajaran biologi mencakup beberapa hal, antara lain ialah padatnya tugas guru, banyaknya materi biologi, dan adanya rasa khawatir pada sebagian besar guru dalam menghadapi UN. Jadi, yang menjadi penghambat dalam hal ini bukanlah UN.

Gambaran tentang implementasi PKP dalam pembelajaran biologi di SMA negeri kota Semarang juga dapat diketahui dari hasil kategorisasi catatan lapangan. Memperhatikan prosedur pengumpulan data yang dilakukan, data tingkat implementasi PKP (kadar inkuiri) hasil kategorisasi catatan lapangan ini dikelompokkan sebagai data kuantifikasi. Data kuantifikasi tingkat implementasi PKP dalam pembelajaran biologi disajikan pada Tabel 4 (kolom ke-3 dan ke-4) yang menunjukkan tingkat penggunaan keterampilan dasar dan keterampilan terpadu oleh peserta didik.

Tabel 4. Proporsi Keterampilan Proses Sain yang Digunakan Peserta Didik dalam Belajar Biologi di Kelas XI IPA Semester Gasal di Beberapa SMA Negeri Kota Semarang, pada Materi ajar/KD Tertentu

Jenis keterampilan yang digunakan	Proporsi penggunaan keterampilan dasar dan terpadu oleh peserta didik		
	Frekuensi *	Keterlaksanaan kegiatan (I) dalam persen	Keterlaksanaan riil (0,72x I) *** dalam persen
Keterampilan Dasar			
- Observasi	77	12.5**	9,0****
- Mengukur	6	1.0	0,7
- Mengklasifikasikan	15	2.4	1,7
- Mengkomunikasikan	333	54.2	39,0
- Memperkirakan	41	6.7	4,8
- Menyimpulkan	25	4.1	3,0
Tingkat Implementasi		81	58
Keterampilan Terpadu			
- Merencanakan penelitian	1	0.2	0,1
- Merumuskan hipotesis	49	7.5	5,4
- Melakukan penelitian/Penyelidikan	32	5.2	3,7
- Menginterpretasikan data	28	4.6	3,3
- Mempresentasikan hasil penelitian/hsl kajian teoretik	10	1.6	1,2
Tingkat implementasi Pil-Ter		19	14

Keterangan:

* : Total kegiatan ada 615 kegiatan;

** : Contoh cara menghitung = $(77/615) \times 100\% = 12,5\%$

*** : Kegiatan yang terlaksana dikalikan dengan 0,72 atau 72% dari yang dilaksanakan.

**** : Keterlaksanaan riil untuk keterampilan mengukur = $12,5 \times 0,72 = 9,0\%$

Data pada Tabel 4 kolom ke-3 menunjukkan adanya ketidak-seimbangan antara tingkat penggunaan keterampilan dasar dan keterampilan terpadu. Dalam hal ini, tingkat penggunaan keterampilan dasar tampak terlalu dominan, terutama sekali pada keterampilan komunikasi. Mengingat bahwa keterampilan proses terdiri atas keterampilan dasar/*basic skills* dan keterampilan terpadu (terintegrasi)/*integrated skills* (Arena, 1996 dalam Martin *et.al.*, 2005, pp.17-20), seharusnya semua jenis keterampilan dapat dikembangkan secara seimbang, bukan didominasi oleh penggunaan keterampilan komunikasi seperti yang ditemukan pada penelitian ini.

Data tentang tanggapan guru mengenai proses pembelajaran biologi diselenggarakan di SMA Negeri Kota Semarang diperoleh dari para guru biologi setelah selesai melaksanakan pembelajaran selama satu semester penuh. Dari mereka, diperoleh berbagai informasi antara lain sebagai berikut. (1) guru biologi di SMAN Kota Semarang sudah berusaha mengimplementasikan PKP dalam pembelajaran, namun tingkat implementasi yang dilaksanakan ternyata berbeda-beda; (2) sebagian besar responden/guru biologi sudah menggunakan alat peraga/media yang cukup bervariasi. Media yang dimaksud mencakup papan tulis, komputer/laptop dengan LCD-nya, gambar/model atau makhluk hidup yang sesungguhnya, LKS dan LDS; (3) para guru sudah memanfaatkan kemajuan teknologi komputer sebagai media.

Untuk menjelaskan lebih lanjut tentang gambaran riil proses implementasi PKP dalam pembelajaran biologi, diperlukan informasi tentang proses pembelajaran yang dilaksanakan guru yang diperoleh melalui observasi, yang merupakan data kualitatif. Penggunaan data kuantitatif dan kualitatif pada penelitian ini sesuai dengan pendapat Abd. Rahman A. Ghani (dalam Setiadi dkk., 2010, p.69) yang mengemukakan bahwa evaluasi merupakan metode penilaian yang melibatkan dua hal, yaitu deskripsi kualitatif dari perilaku peserta didik dan

deskripsi kuantitatif dari hasil pengukuran (misalnya skor tes).

Dalam penelitian evaluasi, penelitian kualitatif dapat membantu penelitian kuantitatif karena informasi kualitatif berperan sangat penting dalam menjelaskan informasi kuantitatif (Brannen, 2005, p.85). Penelitian atau data kualitatif dapat membantu interpretasi hubungan antar variabel, dan juga dapat digunakan untuk menjelaskan faktor-faktor yang mendasari hubungan yang terbangun.

Dari catatan lapangan dan pengamatan terhadap kegiatan yang dilakukan peserta didik selama belajar biologi dapat diketahui bahwa sebagian besar dari kegiatan observasi objek studi yang mereka lakukan ternyata bukan kegiatan mengamati objek/alam, namun lebih banyak mengamati struktur/gambar tertentu yang berasal dari buku teks atau internet. Kegiatan mengobservasi objek asli juga dilakukan peserta didik, namun semua kegiatan tersebut ternyata tidak disertai dengan tuntutan atau instruksi kepada peserta didik untuk melakukan kegiatan lanjutan berupa mengukur dan mengklasifikasikan. Hal ini menyebabkan penggunaan keterampilan mengukur dan mengklasifikasikan menjadi tidak dilakukan peserta didik. Padahal mengukur/melakukan pengukuran merupakan proses dasar dalam sains, yang seharusnya dapat menambah ketepatan pada observasi, klasifikasi dan komunikasi, serta memberi peluang kepada mereka untuk menghasilkan produk ilmu baru, seperti yang dikemukakan oleh Carin & Sund (1989, p.6). Keberhasilan peserta didik dalam menghasilkan produk baru tentu saja tidak terjadi begitu saja, namun memerlukan proses prosedural, menggunakan segenap keterampilan proses yang dimilikinya. Penggunaan keterampilan proses yang tidak menyeluruh atau tidak seimbang diduga kuat akan mempengaruhi kemampuan peserta didik dalam menghasilkan ilmu baru. Oleh karena itu, sudah seharusnya guru melakukan usaha keras untuk mengimplementasikan PKP secara seimbang dalam pembelajaran biologi.

Selama dilakukan penelitian ini, hanya ada seorang guru/responden (disebut sebagai responden khusus) yang menunjukkan tekad kuat untuk mengimplementasikan PKP dalam pembelajaran biologi. Dia menyatakan tidak peduli dengan materi yang banyak, dan selalu berupaya untuk membelajarkan biologi melalui praktikum (dengan mengajak peserta didik untuk menggunakan keterampilan proses sains).

...Saya tidak peduli dengan materi yang banyak, saya selalu mengupayakan praktikum untuk setiap pokok bahasan. Sebagai konsekuensinya, peserta didik perlu mempelajari sendiri materi yang mudah dipahami, sedangkan materi yang dibahas di kelas hanyalah materi yang sulit, yang memerlukan kajian mendalam dengan bantuan guru...

Responden ini juga mengemukakan secara rinci tentang cara yang dilakukan untuk mengatasi kendala implementasi PKP yang ternyata ada pada peserta didik.

...Kondisi peserta didik ini menjadi kendala, karena input peserta didik di sekolah ini (pen: SMAN AE) memang cukup memprihatinkan... Mereka juga memiliki kemampuan komunikasi yang relatif rendah. Cukup banyak peserta didik yang pada awal masuk sekolah ini sulit berkomunikasi. Jumlah mereka mencapai lebih dari 50%. Mereka kurang mampu menjelaskan konsep secara urut, dan pada umumnya menggunakan bahasa Jawa (bahasa tidak standar). Jadi, hal yang paling utama dan pertama harus dilakukan adalah mengajari peserta didik agar mau "ngomong" (melatih kemandirian dan keberanian berbicara/berkomunikasi).

...Hal terpenting untuk dilakukan bagi peserta didik kelas X ialah membangun atau mengembangkan rasa percaya diri peserta didik, melatih disiplin, dan mengembangkan kemampuan bekerjasama. Jika kemampuan itu terbentuk, maka penerapan pembelajaran dengan menerapkan pendekatan keterampilan proses sains pun menjadi mudah. Yang sulit adalah kalau mereka sulit ditata dan sulit berkomunikasi".

Dari penjelasan responden tersebut, tampak jelas bahwa dalam diri seorang guru biologi perlu ada tekad yang kuat untuk dapat menerapkan pendekatan keterampilan proses secara memadai. Tekad akan mendorong seseorang untuk berupaya keras dalam rangka menghilangkan berbagai kendala, agar PKP dapat diimplementasikan dengan baik. Tekad kuat itu penulis sebut sebagai komitmen. Selain dibutuhkan tekad atau komitmen yang kuat, ditemukan sejumlah faktor pendukung bagi guru untuk dapat mengimplementasikan PKP dalam pembelajaran biologi.

Pernyataan lain dari responden khusus ialah sebagai berikut.

...Hal/faktor yang dapat mendukung pelaksanaan pembelajaran dengan pendekatan keterampilan proses, antara lain ialah fasilitas yang dimiliki sekolah. Fasilitas yang dimiliki sekolah ini (pen: SMAN AE) antara lain ialah laboratorium, lingkungan sekitar, ICT, perpustakaan, dan lain-lain. Sejauh dapat digunakan, ya penggunaannya harus dioptimalkan. Salah satu fasilitas yang dimiliki sekolah ini ialah mikroskop binokuler, yang jumlahnya ada dua buah...

Pada saat responden khusus ini berhalangan masuk kelas untuk menyelenggarakan pembelajaran karena tugas lain maka responden ini memilih strategi khusus sebagai solusi yang menurut peneliti sangat bagus dan patut diteladani oleh guru/responden lain.

...Apabila pembelajaran tidak dapat dilaksanakan karena guru mendapat tugas lain, misalnya tugas yang harus dikerjakan di luar sekolah, maka guru akan mencari guru pengganti... Untuk menjamin bahwa pembelajaran yang saya rencanakan terlaksana seperti yang diharapkan, saya melaksanakan coaching, untuk menyampaikan pesan-pesan penting yang terdapat dalam RPP kepada guru pengganti...

Pernyataan responden khusus tersebut mencerminkan adanya pemahaman dan kesadaran yang baik pada guru biologi mengenai hakekat biologi dan hakekat pembelajaran biologi. Hal itulah yang selanjutnya

mendorong munculnya tekad atau komitmen guru untuk menerapkan PKP dalam setiap kegiatan pembelajaran yang diselenggarakannya. Dengan demikian maka jelas bahwa adanya komitmen guru, ternyata merupakan faktor pendukung penting bagi guru untuk menerapkan PKP.

Berdasar kepada uraian tersebut diperoleh informasi bahwa implementasi PKP dalam pembelajaran biologi ternyata belum terlaksana dengan baik. Hal ini tidak akan dapat diungkap dengan baik jika dalam penelitian ini tidak digunakan metode kualitatif yang digabung dengan metode kuantitatif. Menggunakan gabungan kedua metode tersebut dapat diperoleh informasi komprehensif yang menggambarkan kualitas implementasi PKP dalam pembelajaran biologi, beserta berbagai faktor pendukung dan penghambatnya. Faktor pendukungnya ialah fasilitas yang dimiliki sekolah dan adanya komitmen yang tinggi pada guru untuk mengimplementasikan PKP dalam pembelajaran biologi. Sedangkan faktor penghambatnya beberapa hal, yaitu padatnya tugas guru, banyaknya materi biologi, adanya rasa khawatir pada sebagian besar guru dalam menghadapi UN, dan kurangnya komitmen guru untuk mengimplementasikan PKP.

Simpulan dan Saran

Simpulan

Simpulan dari penelitian ini ialah bahwa kualitas implementasi PKP dalam pembelajaran biologi yang diselenggarakan di SMA negeri kota Semarang belum mencapai kategori baik. Hal ini ditandai dengan adanya penggunaan keterampilan dasar yang terlalu dominan oleh penggunaan keterampilan dasar (terutama sekali penggunaan keterampilan komunikasi) dan maraknya praktik pembelajaran tekstual (pembelajaran berorientasi materi).

Faktor pendukung utama implementasi PKP dalam pembelajaran biologi di SMA negeri di kota Semarang adalah adanya fasilitas yang tersedia di sekolah dan komitmen guru yang tinggi untuk mengimple-

mentasikan PKP. Faktor penghambatnya ialah adanya beban tugas mengajar guru yang terlalu berat, materi pelajaran biologi yang sangat banyak, kekhawatiran sebagian besar guru dalam menghadapi UN, dan kurangnya komitmen pada sebagian besar guru untuk mengimplementasikan PKP.

Saran

Beberapa saran yang perlu dikemukakan ialah sebagai berikut. (1) Guru biologi SMA di kota Semarang perlu selalu diingatkan agar tidak terlalu khawatir terhadap kesiapan peserta didik dalam menghadapi UN. (2) Guru perlu diyakinkan dan meyakini bahwa menerapkan PKP dalam pembelajaran biologi merupakan hal yang sangat penting untuk dilakukan. (3) Dalam evaluasi implementasi kurikulum perlu digunakan pendekatan *mixed-method* agar dapat diperoleh informasi yang komprehensif.

Daftar Pustaka

- Alfiero, L., Apanovitch, C., Bearor, J., McNulty, G., Michaud, J., & Simmonds, S. (2008). *Cape elizabeth schools curriculum management plant*. Towson: Baltimore County Public Schools.
- Arikunto, S., & Jabar, C. S. A. (2004). *Evaluasi program pendidikan: pedoman teoretis praktis bagi praktisi pendidikan*. Jakarta: PT Bumi Aksara.
- Brannen, J. (2005). *Memadu metode penelitian kualitatif dan kuantitatif*. Cetakan ke VI. Yogyakarta: Penerbit Fakultas Tarbiyah IAIN Antasari Samarinda bekerja sama dengan Pustaka Pelajar. ISBN: 979-8581-107-9.
- BSNP. (2006). *Panduan penyusunan kurikulum tingkat satuan pendidikan jenjang pendidikan dasar dan menengah*. Jakarta: Badan Standar Nasional Pendidikan.
- Carin A.A. & Sund R.B. (1989). *Teaching science through discovery*. Sixth Edition. Melbourne: Merril Publish Company.
- Claeys, C., Neve, J., Tulkens, P.M., Spine-wine, A. (2012). Content validity and inter-rater reliability an instru-ment to

- characterize unintentional medication discrepancy. Original Research Article. *Drugs Aging*. Springer International Publishing, A.G., 20 (7), 577-591.
- Creswell, J.W. (2008). *Educational Research: Planning, conducting, and evaluating quantitative and qualitative research*. Third Edition. Pearson International Edition. New Jersey: Pearson Education Inc.
- Djohar. (15 Mei 2003). *Gambaran umum program pembaharuan pendidikan MIPA dan bahasa inggris*. Makalah disampaikan pada Penataran Guru Penelitian di YAMB, Yogyakarta.
- Setiadi, H., Ghani, A. R. A., & Suyatno. (2006). *Evaluasi pendidikan: konsep dan aplikasi*. Jakarta: Uhamka Press.
- Hayat, B. (2006). *Prinsip-prinsip dan strategi penilaian kelas*. Jakarta: Pusat Penilaian Pendidikan. Badan Penelitian dan Pengembangan. Departemen Pendidikan Nasional.
- Kemendiknas. (2006a). *Peraturan mendiknas no 22 tahun 2006 tentang standar isi untuk satuan pendidikan dasar dan menengah*. Jakarta: Depdiknas RI.
- Kemendiknas. (2006b). *Peraturan mendiknas no 23 tahun 2006 tentang standar kompetensi lulusan tingkat sekolah menengah pertama dan madrasah tsanawiyah*. Jakarta: Depdiknas RI.
- Kemendiknas. (2007). *Peraturan mendiknas no 41 tahun 2007 tentang standar proses pelaksanaan kurikulum pada jenjang pendidikan dasar dan menengah*. Jakarta: Depdiknas RI.
- Madaus, G.F., Scriven, M., Stufflebeam, D.L. (1986). *Evaluation models: viewpoints on educational and human services evaluation*. Boston: Kluwer Nijhoff Publishing.
- Martin, R., Sexton, C., Franklin, T., & Gerlovich, J. (2005). *Teaching science for all children: inquiry methods for constructing understanding*. New York: Pearson Education, Inc.
- OECD. (2010). *PISA 2009 Result: executive summary*. 1-21.
- Pusat Kurikulum Badan Penelitian dan Pengembangan. (2007). *Naskah akademik kebijakan kurikulum mata pelajaran IPA*. Jakarta: Depdiknas RI.
- Smith, T. M., Desimone, L. M., Zeidner, T. L., Dun, A.C., Bhatt, M., & Rumi-yant-seva, N.L. (September 2007). Educational evaluation and policy analysis. Inquiry-oriented instruction in science: Who teaches that way? *ProQuest Education Journals*, 29 (3), 169-199.
- Steinmetz, A. (2000). *The discrepancy evaluation model. evaluation models: evaluation in education and human services*. Dalam: Stufflebeam, D. L (Ed), Madaus, G.F., Kellaghan T., *Evaluation models: viewpoints on educational and human services evaluation*. ISBN: 978-0-7923-7884-6 (Print) 978-0-306-47559-7 (Online), 49, 127-143.
- Stufflebeam, D.L.; Madaus, G.F.; Kellaghan, T. (2002). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Second Edition. Kluwer Academic Publishers. New York. 509 pgs.
- Undang-undang Sisdiknas nomor 20 tahun 2003
- Wheeler, S.L. (2012). From Discrepancy Evaluation To Response To Intervention (Rti): Are We Ready In Higher Education? *Dissertation*. Published by ProQuest LLC (2012). UMI 3544014. Microform Edition © ProQuest LLC.
- Widoyoko, S. P. (2008). *Peranan evaluasi program pembelajaran dalam meningkatkan kualitas pendidikan*. Makalah penyerta disampaikan dalam Seminar Strategi Peningkatan Kualitas Pendidikan pada tanggal 22 Maret 2008, di PPS UNY, Yogyakarta. Kerjasama PPS UNY- HEPI.
- Worthen B.R.& Sanders J.R. (1981). *Educational evaluation: theory and practice*. Woshington: Charles A. Jones Publishing Company.