

MODEL EVALUASI DALAM SUPERVISI PEMBELAJARAN IPA BERBASIS LIMA DOMAIN SAINS

¹⁾Eny Winaryati, ²⁾Suyata, ³⁾Sumarno

¹⁾Universitas Muhammadiyah Semarang (UNIMUS),

^{2,3)}Universitas Negeri Yogyakarta

¹⁾enie.weye@gmail.com

Abstrak

Tujuan penelitian ini adalah: 1) mengembangkan model evaluasi diri dan teman sejawat (EDTS) dalam supervisi pembelajaran IPA berbasis lima domain sains (5 DS); 2) mengevaluasi pembelajaran IPA berbasis 5 DS dalam supervisi melalui model EDTS; 3) menguji keefektifan model. Objek penelitian ini adalah SMP Negeri 7, 11, 14, 15, 19, 29, dan 32. Model EDTS dikembangkan melalui penelitian dan pengembangan (R&D). Subjeknya adalah guru sendiri, gurusejawat, dan kepala sekolah. Alur pelaksanaan evaluasinya adalah: guru sendiri dan guru sejawat mengobservasi kegiatan pra pembelajaran dan proses pembelajaran. Data observasi divalidasi oleh kepala sekolah, ditindaklanjuti dengan diskusi, dan dilanjutkan dengan kegiatan *feedback* dan evaluasi diri. Penelitian ini menghasilkan: 1) model EDTS melalui pendekatan ANTRANINPRO (*Antecedent, Transaction, Interim Product*), dikembangkan dengan metode R&D; 2) Model EDTS disusun untuk memperkuat pelaksanaan supervisi di sekolah, melalui dukungan informasi yang diberikan guru kepada kepala sekolah; 3) berdasarkan hasil evaluasi, diperoleh penilaian kurang baik pada item yang terkait dengan 5DS, terutama domain kreativitas dan aplikasi sains; 4) model EDTS efektif untuk digunakan, berdasarkan data: validitas model (sangat baik), validitas konten (96% relevan), reliabilitas (kriteria baik), model memberi kemudahan user untuk menggunakan, serta memberi dampak positif terhadap perbaikan pembelajaran.

Kata kunci: *model evaluasi, supervisi pembelajaran, lima domain sains*

EVALUATION MODEL IN THE TEACHING SUPERVISION OF NATURAL SCIENCES BASED ON FIVE DOMAINS OF SCIENCE

Abstract

The purposes of this study are: 1. developing a self and peer evaluation model (EDTS model) in the teaching supervision of natural sciences based on five domains of science (DS 5); 2. evaluating the teaching in supervision of natural sciences based on five domains of science (DS 5) through EDTS model; 3. measuring the effectiveness of this model. The objects of this research are SMPN 7, 11, 14, 15, 19, 29, and 32. The EDTS model was developed through research and development (R&D). The subjects are teachers, teacher's peers and headmasters. The steps of the evaluation were: teachers and teacher's peers observed-pre learning activities and learning processes. Observation data were validated by the headmasters, followed by discussion and feedback and self-evaluation. The results of this research are: 1) EDTS models through ANTRANINPRO approach (*Antecedent, Transaction, Interim Product*) is developed by the R & D method; 2) EDTS model is designed to strengthen the implementation of supervision in schools, through the support of information provided by the teacher to the headmaster; 3) Based on the evaluation, items associated with 5DS are considered not so good, especially the domain of creativity and scientific applications; 4) EDTS model is effective to be used, based on the data: the validity of the model (very good), content validity (96% relevant), reliability (good criteria), the model gives the user easiness to use, and gives a positive impact on learning improvement.

Keywords: *models of evaluation, teaching supervision, five domains of science.*

Pendahuluan

IPA merupakan pengetahuan yang telah mengalami uji kebenaran melalui metode ilmiah dengan ciri: objektif, metodik, sistematis, universal, dan tentatif. Pendidikan IPA sekarang ini diarahkan untuk menitikberatkan pada pengembangan taksonomi pendidikan IPA, melalui 5 DS. Ke-5 DS harus selalu terintegrasi dan tidak dapat dipisahkan satu dengan lainnya. Ke-5 DS ini meliputi: pengetahuan, sikap ilmiah, proses ilmiah, aplikasi, dan kreativitas (McCormack, 1999: 24). Hal ini dimaksudkan untuk mengimbangi kemajuan ilmu pengetahuan dan teknologi.

Hasil penelitian memberikan gambaran bahwa guru kurang memperhatikan kebutuhan, permasalahan, dan memantau siswa. Guru kurang mengikuti arah dan aktivitas dalam wilayah kurikulum, serta kurang melakukan penelitian, (Winaryati, 2011b). Hal ini mengindikasikan bahwa kemampuan guru untuk mengelola kelas masih kurang.

Upaya peningkatan kemampuan guru dalam pembelajaran IPA, sangat erat kaitannya dengan program supervisi yang dilakukan baik oleh kepala sekolah maupun pengawas sekolah. Supervisi yang terjadi selama ini belum memberikan hasil sesuai dengan yang diharapkan. Program supervisi pendidikan IPA seharusnya dapat dilakukan oleh supervisor yang memiliki kompetensi: (1). penguasaan materi subjek (*subject knowledge*) sesuai dengan hakikat IPA; (2). penerapan IPA (*subject application*) dalam pembelajaran di kelas; (3). pengelolaan kelas (*class management*) dan (4). penilaian hasil pembelajaran (*assessment*) (Tim IKIP Bandung, 1999, p.3).

Hasil penelitian Winaryati (2011a), tentang “Kompetensi Kepala Sekolah sebagai supervisor pada pembelajaran IPA SMP di Kota Semarang”, ditemukan bahwa: supervisi dilakukan oleh kepala sekolah didasarkan pada minat dan kesempatan yang dimilikinya. Supervisi dilakukan hanya yang umum-umum saja, guru merasa canggung dan grogi ketika disupervisi oleh kepala sekolah. Kepala sekolah tidak menguasai ma-

teri, metode dan media pembelajaran IPA berakibat pemberian umpan balik terlalu umum dan kurang mengarah ke aspek yang dibutuhkan guru. Perlunya melibatkan guru dalam perencanaan pelaksanaan supervisi. Menurut Marzanon (2003) dalam Cooper & Forrest (2009, p.24), bahwa kualitas hubungan supervisor dengan guru, merupakan faktor yang paling penting dalam supervisi dari-pada metode supervisi yang digunakan.

Supervisi akademik yang dilakukan per bidang IPA melalui guru inti dari Kelompok Kerja Guru (KKG), Musyawarah Guru Mata Pelajaran (MGMP), atau bentuk kegiatan lain, tidak berjalan dengan mulus. Realita dari kegiatan ini belum memberikan hasil sesuai dengan yang diharapkan. Realita ini, perlu mempertimbangkan adanya konsep pengembangan program supervisi pendidikan IPA melalui penggalan potensi yang ada pada guru dan teman sejawat pada rumpun mata pelajaran IPA, melalui evaluasi diri dan teman sejawat.

Melalui evaluasi diri dapat membantu supervisi mengatasi persoalan yang dihadapi. Temuan penelitian yang dilakukan oleh Winaryati (2011a, 2011b), bahwa penilaian yang dilakukan terhadap diri sendiri lebih tinggi bila dibandingkan dengan penilaian yang dilakukan oleh orang lain (*peer*). Hal ini juga didukung oleh Ozogul, Olina & Sullivan (2008, pp. 181-201), bahwa skor *self evaluation* lebih tinggi dibandingkan dengan *peer evaluation*. Persoalan ini menjadi bahan pertimbangan untuk dilaksanakannya penilaian oleh teman sejawat.

Sibbald (2009, pp.452-454) dalam bukunya *Lesson study and Self-Efficacy, (report)*, memberikan kontribusi bahwa evaluasi teman sejawat dapat meningkatkan prestasi siswa, meningkatkan kemampuan guru dalam pemilihan metode, terbentuknya komunitas yang saling melengkapi. Hasil penelitian ini juga didukung oleh Meyer (2005) dalam Sibbald (2009), dan Middendorf & Kalish (1995), ditemukan dari penelitiannya bahwa adanya signifikansi berkenaan dengan keterlibatan siswa dalam pembelajaran.

Persoalan supervisi di atas, mendorong peneliti untuk melakukan revitalisasi

konsep supervisi yang selama ini terjadi. Konsep yang digagas adalah evaluasi diri dan teman sejawat (EDTS). Konsep ini menitikberatkan pada evaluasi terhadap diri sendiri dan evaluasi yang dilakukan oleh teman sejawat (rumpun guru IPA). Hal ini didasarkan dari berbagai penelitian yang mendahului pada konteks yang berbeda serta konsep umpan balik 360 derajat, melalui model evaluasi yang relevan.

Model EDTS pada supervisi pembelajaran IPA ini berlandaskan inti dari kegiatan evaluasi dan supervisi. Inti dari kegiatan evaluasi adalah proses dihasilkan-nya informasi sebagai alternatif keputusannya (Stufflebeam, 1973, pp.3-5; Hamm, 1985, pp.256-622; Stake, 1967, pp.2-4; Stufflebeam & Shinkfield, 1985, p.159). Inti dari supervise adalah pembinaan ke arah perbaikan, yang berdampak peningkatan kualitas pembelajaran, (Wiles & Lovell, 1975, p.20-23; Diat Prasajo, 1975, p.31; Evans, 2009, pp.1-23); Stinson, 2000).

Model EDTS yang disusun mengaitkan fungsi evaluasi agar ditindaklanjuti, dan berdampak pada supervisi untuk melakukan perbaikan pembelajaran. Model ini melibatkan guru dalam kegiatan supervisi, melalui evaluasi diri dan teman sejawat, untuk memberikan dukungan informasi kepada supervisor. Peran kepala sekolah sebagai supervisor lebih optimal, melalui kegiatan memvalidasi data antara guru sendiri dengan guru sejawat, serta menjadi fasilitator dalam kegiatan diskusi dan *feedback*. Model disusun berdasarkan persoalan pelaksanaan supervisi akademik yang sekarang ini dilaksanakan; harapannya dapat digunakan untuk memperkuat kegiatan supervisi akademik di sekolah.

Tujuan yang ingin dicapai dalam penelitian ini adalah: (1) mengembangkan evaluasi model EDTS pada supervisi pembelajaran IPA berbasis lima domain sains; (2) mengevaluasi pembelajaran IPA berbasis 5 DS dalam supervisi, melalui EDTS; (3) mengetahui keefektifan model EDTS pada supervisi pembelajaran IPA.

Metode Penelitian

Model EDTS, dikembangkan melalui penelitian dan pengembangan (R&D). Tahapan R&D yang digunakan *Analysis, Definition, Design, Demonstration, Development*.

Penelitian ini dilakukan dari bulan Mei 2012 - Februari 2013. Objek penelitian adalah: SMP Negeri 14, 29, 15, 7, 19, 32, dan 11.

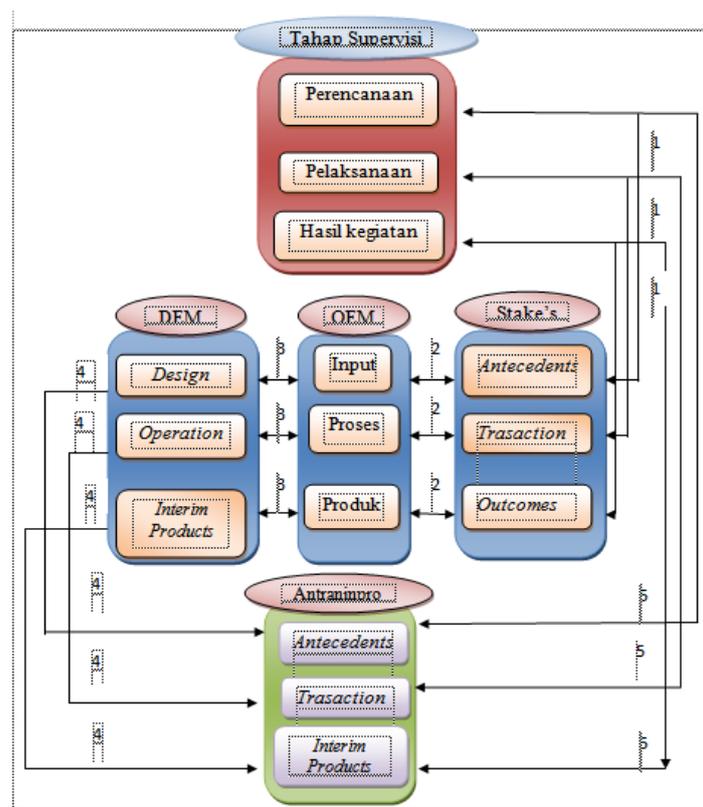
Kegiatan uji coba melibatkan guru IPA sendiri, guru sejawat, dan kepala sekolah. Rangkaian kegiatan untuk satu kali uji coba adalah: (1) observasi terkait dengan persiapan guru mengajar; (2) observasi pada saat guru mengajar; (3) diskusi; (4) *feedback*; dan (5) evaluasi diri. Setiap kegiatan observasi, peneliti ikut melakukan penilaian, sebagai triangulasi sumber.

Data dianalisis secara *mixed method approach*, yaitu gabungan pendekatan kualitatif dan kuantitatif. Data kuantitatif digunakan pada saat melakukan analisis statistik, di antaranya: pengujian validitas (muka dan isi), dan reliabilitas. Data kualitatif diperoleh melalui saran, dan pendapat yang diperoleh dari kegiatan FGD, *expert judgement*, dan kegiatan uji coba.

Hasil Penelitian dan Pembahasan

Pengembangan Model Evaluasi

Pemilihan model evaluasi didasarkan dari tahapan supervisi akademik, yang meliputi tiga tahapan praobservasi, observasi pembelajaran, dan pascaobservasi. Hal ini diperkuat oleh Isaac & Michael (1981, pp.2,3), bahwa karakteristik keunggulan model evaluasi, menekankan pendekatan sistem (*input, proses dan output*). Berdasarkan tahap supervisi tersebut, model evaluasi yang relevan adalah *Stake's Countenance Model*. Stake membagi objek evaluasi menjadi 3 kategori, yaitu: (1) *Antecedents* (input); (2) *Transaction*, (proses); (3) *Outcomes* (efek samping dari sistem) (Fernandes, 1984:8-10). Alur pendekatan modelnya sebagai berikut.



Gambar 1. Tahapan Pengembangan Model Evaluasi ANTRANINPRO

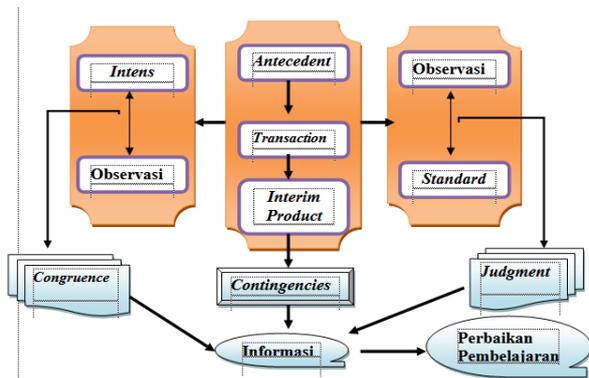
Berdasarkan penelitian yang dilakukan oleh (Yonnie Chyung, 2001, pp.36-49; Yonnie Chyung, 2006; Kaufman, 2005, pp.6-16); Owston (2008, pp.605-617), diperoleh gambaran bahwa sebelum melakukan penetapan model evaluasi yang relevan, terlebih dahulu ditetapkan elemen organisasi yang tepat agar pemetaan evaluasi yang diharapkan tepat. Pemetaan ini dilakukan dengan menggunakan *Organisational Element Models* (OEM). Tahapan evaluasi Stake bila ditetapkan dengan tahapan OEM (*input*, proses, produk, *output*, dan *outcome*), maka tahap evaluasi Stake yang relevan adalah *Input* (*Antecedent*), Proses (*Transaction*), dan Produk (*Outcomes*), (Provus, 1969, pp.2,3; Kaufman & Thomas, 1980, p.137).

Deskripsi *Outcomes* model Stake adalah dampak dari pengalaman instruksional/ suatu program. Deskripsi ini kurang relevan dengan maksud dari pascaobservasi/ umpan balik dari supervisi. Hal inilah yang menjadi landasan perlunya dimodifikasi dengan model evaluasi yang lain.

Outcomes model Stake adalah dampak dari pengalaman instruksional/ suatu program. Deskripsi ini kurang relevan dengan maksud dari kegiatan pascaobservasi/ umpan balik dari kegiatan supervisi yang bersifat sementara untuk ditindaklanjuti. Hal itu menjadi landasan perlunya dimodifikasi dengan model evaluasi yang lain. Model evaluasi yang memiliki karakteristik identik dengan model Stake adalah (*The Discrepancy Evaluation Models* (DEM)), (Provus, 1969, pp.2-3). Kedua model evaluasi ini, sama-sama menekankan konsep kesenjangan.

Penetapan tahapan yang relevan dari tahapan *outcomes*, dapat ditetapkan dengan menggunakan tahapan dari OEM (*Organizational Elements Model*). OEM “menyediakan kerangka merancang dan menerapkan cara yang efektif untuk mencapai hasil akhir yang diinginkan”, (Yonnie Chyung, 2001, pp.36-49) *product* dari OEM, relevan dengan tahap *interim products* dari DEM (Provus, 1969, pp.2,3; Kaufman & Thomas, 1980). Tahap *interim products* ini sejalan

dengan tahap supervisi yang diharapkan, dan deskripsi produk dari OEM. Deskripsi dari *interim products* adalah menggambarkan hubungan antara proses program dan produk sementara. (Kaufman & Susan (1980, p.137). Kombinasi dua model evaluasi ini dinamakan ANTRANINPRO (*Antecedent, Transaction, Interim Products*). Jalannya model evaluasi tertera pada gambar berikut ini.

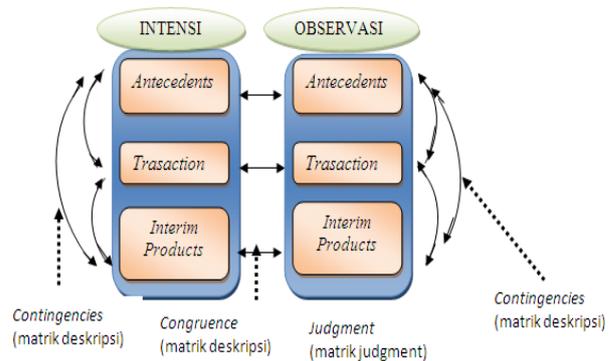


Gambar 2. Tahapan Proses Evaluasi

Informasi hasil evaluasi, terdiri atas: (1) hasil evaluasi kesesuaian antara intensi dengan observasi keseluruhan tahap evaluasi, dan setiap tahap kegiatan. Setiap tahap kegiatan meliputi *antecedent*, *transaction*, dan *interim product*. Data evaluasi diperoleh dari matrik penilaian; (2) Tahap *congruencies* yaitu deskripsi kesesuaian pada *antecedent*, *transaction* dan *interim product*, antara intensi dengan observasi; (3) tahap *contingencies* yaitu: hubungan antara *antecedent vs transaction*, *transaction vs interim product*, dan *antecedent, transaction*, dan *interim product*, baik pada intensi dan observasi. Berikut gambar 6 (enam) cell model ANTRANINPRO.

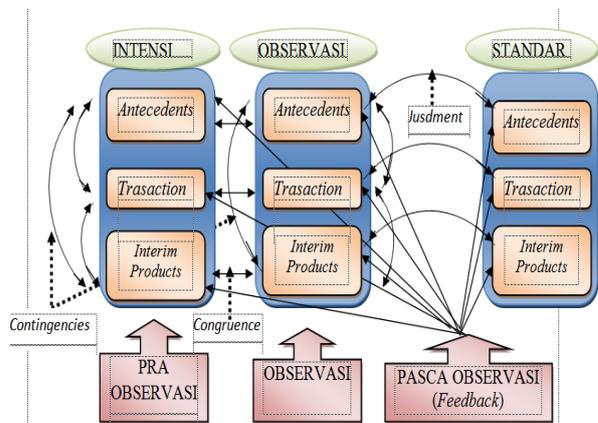
ANTRANINPRO merupakan metode pendekatan dari model EDTS. ANTRANINPRO membantu penilai dalam mengumpulkan, mengorganisir, dan menafsirkan data kuantitatif dan kualitatif untuk mencapai tujuan. Matrik Stake mencakup (*intens, observations, standards, dan judgments*). Evaluator merumuskan *intent* (maksud) dan standar. Standar disarikan dari indikator Instrumen Supervisi Akademik dimodifikasi, Instrumen Penilaian Kinerja Guru (IPKG I dan IPKG II), dipadukan dengan 5

DS, *A Guidebook for peer evaluation* (Savoie, 2010, pp.11-28), dan beberapa peraturan/keputusan dari pemerintah. Penilaian dilakukan oleh diri sendiri, guru serumpun. Peran supervisor kepala sekolah sebagai fasilitator dalam proses supervisi, dan menjadi validator dalam *judgment*. Analisis deskripsi informasi, diidentifikasi melalui *congruence* (kesesuaian antara yang dimaksudkan dengan yang terjadi) dan *contingencie* (hubungan logis antara *Antecedents*, *Transaction*, dan *Interim Products*).



Gambar 3. Enam Cell model ANTRANINPRO

Penilaian, terdiri dari 2 matrik, yaitu matrik penilaian dan matrik deskripsi. Matrik penilaian, berisi kegiatan penilaian dengan cara mencentang pada nilai 0 (nol), 1 (satu) atau 2 (dua), pada item yang sesuai. Matrik deskripsi, berisi kegiatan penilaian dengan cara melingkari item narasi yang sesuai. Gambaran model EDTS dengan pendekatan model evaluasi ANTRANINPRO sebagai berikut:



Gambar 4. Model evaluasi ANTRANINPRO

	Description Matrix		Judgment Matrix	
	Intens	Observations	Standards	Judgement
R a t i o n a l	C o n g r u e n c e	C o n g r u e n c e	ANTECEDENT	
	C o n g r u e n c e	C o n g r u e n c e	TRANS-ACTION	
	C o n g r u e n c e	C o n g r u e n c e	OUTCOMES	

Gambar 5. Tahapan Pelaksanaan Model Evaluasi ANTRANINPRO

Perbaikan terhadap Model EDTS

Data revisi terhadap model diperoleh melalui: (1) hasil kegiatan FGD (ahli dan praktisi); (2) saran dari kegiatan uji coba.

Saran dan pendapat ahli ini diperoleh melalui kegiatan *Focuss Group Discussion* (FGD) pakar/ahli yang terlibat dalam kegiatan FGD adalah 5 (lima) pakar/ahli, meliputi pakar evaluasi, pembelajaran sains, pengukuran dan supervisi. Hasil saran dari ahli digunakan sebagai rujukan untuk melakukan perbaikan terhadap model.

Data perbaikan berdasarkan saran oleh ahli meliputi: (1) penilaian terkait 5 DS dilakukan pada silabus, komponen-komponen RPP, nilai KKM, proses pembelajaran dan hasil belajar siswa; (2) konstruk teori dibangun dari beberapa literatur, kemudian peneliti melakukan penekanan terhadap konten yang terkait dengan penelitian; (3) petunjuk pelaksanaan model EDTS diperjelas meliputi: peran kepala sekolah pada tiap tahap, petunjuk dan alur evaluasi, penambahan kelebihan dan kekurangan model, tindak lanjut hasil rekomendasi; (4) perbaikan terkait dengan instrumen model EDTS meliputi: mengemblok tabel yang tidak harus diisi dengan warna tertentu, skala penilaian telah disamakan menjadi nilai 0,1,2, penggunaan istilah-istilah yang umum dan mudah dipahami, seperti: kata *antecedent* diperjelas dengan bahasa Indonesia perencanaan; (5) perbaikan terkait dengan supervisi pembelajaran IPA: (a) teknik penyama-

an persepsi penilaian antarguru; (b) Konsep teori diawali dengan penjelasan rinci meliputi: maksud dan tujuan, cara menilai, 5 DS, peran guru sendiri dan teman sejawat; c) Untuk memperkecil terjadinya bias pengamatan/observasi untuk penilaian antarguru, peneliti ikut terlibat pada saat observasi sebagai triangulasi sumber.

Selain FGD oleh para ahli, juga FGD oleh praktisi. Praktisi yang terlibat adalah guru IPA dan kepala sekolah SMP. FGD oleh MGMP IPA terjadi 3 (tiga) kali, yaitu: FGD awal (menjelang penyusunan instrumen), tengah (setelah instrumen jadi), dan akhir (sebelum kegiatan simulasi dimulai).

Saran dan pendapat oleh Musyawarah Guru Mata Pelajaran (MGMP) IPA SMP kota Semarang, sebagai berikut: (1) Dukungan disusunnya model EDTS; (2) saran pemberian penilaian terdiri dari 3 skor; (3) Perlunya diperhatikan terkait dengan pelaksanaan model EDTS, mengingat jam tatap muka guru sangat padat; (4) perbaikan terhadap petunjuk model; (5) perlunya memberi keterangan pada istilah-istilah asing; (6) perlunya pencantuman penilaian ulangan harian; (7) penetapan nilai KKM berbasis 5 DS. (pengetahuan sains $\geq 85\%$, proses sains $\geq 75\%$, kreativitas sains $\geq 65\%$, sikap sains $\geq 85\%$, aplikasi sains $\geq 65\%$); (8) uji coba sebaiknya kelas 1 atau kelas 2; (9) teknik pengisian deskripsi, guru jangan menulis sendiri, tapi tinggal mencentang dan menambah bila ada yang kurang.

Hasil FGD dengan pengurus MKKS SMP kota Semarang, dihasilkan beberapa saran sebagai berikut: (1) kegiatan EDTS guru yang dilibatkan sebaiknya hanya 1 orang sebagai teman sejawat; (2) meski hanya 2 orang guru dan 1 KS, kegiatan *feedback* tetap dapat dilakukan.

Revisi terhadap model, juga diperoleh dari kegiatan uji coba. Saran dapat muncul pada saat kegiatan observasi, diskusi, atau pasca diskusi, baik oleh guru, atau kepala sekolah. Saran lebih mengarah pada panduan model dan alur evaluasi. Kedua hal ini sempat terangkat pada FGD ahli, dan *expert judgment* yang membutuhkan untuk diperhatikan. Hasil rekomendasinya sebagai beri-

kut: (1) Saran dari guru sejawat, pada silabus perlu ditambah komponen KARAKTER, karena komponen ini sudah menjadi komponen wajib; (2) perbaikan terhadap beberapa kata kalimat, dengan harapan agar mudah dipahami; (3) petunjuk pada matrik deskripsi bukan dicentang tapi dilingkari pada angka yang ada; (4) halaman depan sebaiknya sudah tertulis nama guru sendiri dan teman sejawat; (5) sebaiknya temuan yang diangkat untuk didiskusikan tidak hanya 2 tingkat saja, namun bila terjadi kesenjangan maka hal itu harus diangkat untuk didiskusikan juga; (6) perlu adanya penyingkatan kata, pada kategori terkait dengan petunjuk penilaian. Misanya: tidak baik (TB); (7) perlu ditambahkan tulisan nilai 0, 1, 2 pada lembar *feedback* agar tidak membingungkan pengisian, ketika akan mengisi nilai terkait dengan pemasukannya pada kolom; (8) perlunya dilakukan kegiatan pelatihan bagi guru dan kepala sekolah sebelum kegiatan pelaksanaan model EDTS dilakukan; (9) proses diskusi sebai dilakukan melalui kegiatan klarifikasi; (10) konsistensi penggunaan kata PESERTA DIDIK dari pada siswa; (11) pada saat kegiatan diskusi masing-masing guru membawa lembar penilaiannya, dan kepala sekolah memegang data rekapitulasi data penilaian. Diskusi dilakukan dengan cara klarifikasi data perbedaan yang muncul.

Mengevaluasi Pembelajaran IPA dalam Supervisi melalui EDTS

Informasi dari Kegiatan Evaluasi

Informasi dari hasil evaluasi, terdiri atas: (a) kesesuaian antara *intens* (yang diharapkan) dengan *observation* (pelaksanaan) pada matrik penilaian; (b) kesesuaian antara *instens* dengan *observation* pada matrik deskripsi (*congruencies*); (c) hubungan antartahap evaluasi, baik *intens* dan *observation* (*contingencies*).

Kesesuaian Intens dengan Observation pada Matrik Penilaian

Data penilaian secara kuantitatif, dikonversikan ke dalam data kualitatif, untuk

mendapatkan analisis teks. Penetapan data kualitatif dilakukan dengan cara kategorisasi. Skor penilaian dari model EDTS adalah: 0, 1, 2, maka dihasilkan data kategori sebagai berikut:

- $0 - \leq 0,5$: Tidak Baik (TB)
- $0,5 - \leq 1$: Kurang Baik (KB)
- $1 - \leq 1,5$: Baik (B)
- $1,5 - \leq 2$: Sangat Baik (SB)

Hasil penilaian kesesuaian *intens* dengan *observation*, pada matrik penilaian diperoleh temuan sbegagai berikut: (1) Diperoleh data pada tahap *intens*: *self*= 1,55 (SB), *peer*= 1,51 (SB). Data pada tahap *observation*/pelaksanaan, adalah: *self* = 1,54 (SB), *peer* = 1,50 (B); (2) Ada perbedaan penilaian antara *self* dengan *peer*. Penilaian *self* relatif lebih tinggi dibanding penilaian oleh *peer*; (3) Penilaian yang dilakukan oleh *self*, terdapat kesesuaian antara yang diharapkan (*instens*) dengan yang dilaksanakan (*observation*), dengan kategori sangat baik; (4) Penilaian yang dilakukan oleh *peer*, terdapat ketidaksesuaian, antara yang diharapkan (*instens*) dengan kategori sangat baik, sedang yang dilaksanakan (*observation*), dengan kategori baik. Namun perbedaannya tidak terlalu jauh; (5) Ditemukan penilaian tidak baik, antara *intens* (pada dimensi perencanaan nilai KKM berbasis 5DS), dengan *observation* (terkait dimensi capaian nilai KKM berbasis 5 DS). Hal ini berarti, kesesuaia antara yang diharapkan dengan pelaksanaan, nilainya tidak baik. Belum adanya kesesuaian antara yang diharapkan (perencanaan nilai KKM berbasis 5DS), dengan yang dilaksanakan, (capaian nilai KKM berbasis 5 DS); (6) Ditemukan adanya ketidaksesuaian antara yang diharapkan (*intens*) dengan yang dilaksanakan (*observation*); (7) Pada tahap *antecedent* (perencanaan): Sangat Baik (pada *intens*), Baik (pada *observation*); (8) Pada tahap *interim product* (hasil) dengan nilai: Baik (pada *intens*), Kurang Baik (pada *observation*).

Kesesuaian Intens dengan Observation pada Matrik Deskripsi (Congruencies)

Tahap *Congruencies* yaitu deskripsi kesesuaian pada *antecedent* (antara persiapan

dengan pelaksanaan), *transaction* (antara perencanaan dengan pelaksanaan proses pembelajaran) dan *interim product* (antara perencanaan dengan dampak pembelajaran). Pada intinya *congruencies* adalah kesesuaian antara intensi (yang diharapkan) dengan observasi (pelaksanaan), pada matrik deskripsi. Matrik deskripsi adalah suatu penilaian dengan cara

memberi nilai 0 (jika tidak ada/terjadi) dan nilai 1 (jika ada/terjadi), sehingga penetapan kategori skornya adalah:

- 0- ≤0,25 : Tidak Baik (TB).
- 0,25 - ≤0,5 : Kurang Baik (KB).
- 0,5 - ≤0,75 : Baik (B).
- 0,75 - ≤1 : Sangat Baik (SB).

Tabel 1. Data *Congruence*

CONRUENCIES							
	Dimensi	Nilai <i>Self</i>	Rata-rata	Kategori	Nilai <i>peer</i>	Rata-rata	Kategori
Antecedent	Silabus	0,79		SB	0,86		SB
	Kesiapan realisasi RPP	0,48	0,62	KB	0,43	0,57	KB
	Dokumen Pendukung	0,61		B	0,43		KB
	Pengelolaan Pembelajaran	1,00		SB	1,00		SB
Transaction	Knowledge	0,76		SB	0,76		SB
	Proses Sains	0,86	0,85	SB	0,90	0,79	SB
	Kreativitas	0,81		SB	0,43		KB
	Sikap Sains	0,90		SB	1,00		SB
Interim Product	Penerapan	0,76		SB	0,67		B
	Penilaian Pembelajaran	0,90		SB	0,57		B
	Penilaian berbasis 5 DS	0,46	0,56	KB	0,4	0,40	KB
	Nilai KKM berbasis 5 DS	0,31		KB	0,23		TB
	Rata-rata	0,72		B	0,64		B

Hasil penilaian kesesuaian *intens* dengan *observation*, pada matrik deskripsi (*Congruencies*), diperoleh beberapa temuan: (1) Secara keseluruhan, penilaian *self* relatif lebih tinggi dibandingkan dengan penilaian oleh *peer*; (2) Ditemukan beberapa dimensi yang memiliki nilai kurang baik, yaitu: (a) Terkait dengan kesiapan realisasi RPP pada pembelajaran yang berbasis 5 DS. Hal ini berarti belum adanya kesesuaian antara RPP dengan kesiapan guru mengajar, kesiapan siswa dan kesiapan peralatan yang ada, (b) Terkait dengan dokumen pendukung. Ditemukan belum adanya kesesuaian terkait dengan ketersediaan dokumen pendukung dengan realisasi pelaksanaannya, mencakup: jurnal guru, modul/bahan ajar, LKS, bank soal, (c) Terkait dengan kreativitas siswa. Belum adanya kesesuaian antara perencanaan dengan proses pembelajaran ter-

kait dengan kreativitas peserta didik, terutama pada indikator: pemilihan dan penggunaan media yang mengarah pada pertumbuhan pengembangan ide peserta didik, kesesuaian sumber belajar dengan perkembangan peserta didik sehingga mampu menumbuhkan kreativitasnya, untuk menghubungkan dengan masa depannya, (d) Terkait dengan penilaian berbasis 5 DS. Ditemukan belum adanya kesesuaian perencanaan teknik penilaian yang digunakan dengan pelaksanaan untuk melakukan penilaian berbasis 5 DS, (e) Terkait dengan nilai KKM berbasis 5 DS, ditemukan belum adanya kesesuaian antara perencanaan nilai KKM berbasis 5 DS dengan pelaksanaan, meliputi: pengetahuan siswa, proses sains, kreativitas, sikap sains, dan aplikasi sains; (3) Berdasarkan data penilaian di atas, nilai

kurang baik/tidak baik, karena belum terlaksananya 5 DS.

Hubungan antara Antecedent, Transaction dan Interim Product (Contingencies)

Hubungan antara *antecedent vs transaction*, *transaction vs interim product*, dan *antecedent*,

transaction dan *interim product*, baik pada *intens* dan *observation*, semua hasil penilaian tidak baik. Hal ini memberikan gambaran tidak adanya hubungan antara masing-masing hubungan di atas.

Tabel 2. Data *Contingencies*

Dimensi	CONTINGENCIES							
	INTENS				OBSERVATION			
	Nilai <i>Self</i>	Kategori	Nilai <i>Peer</i>	Kategori	Nilai <i>Self</i>	Kategori	Nilai <i>Peer</i>	Kategori
Antecedent VS Transaction	0,24	TB	0,21	TB	0,21	TB	0,19	TB
Transaction VS Interim Product	0,14	TB	0,05	TB	0,21	TB	0,14	TB
Antecedent, Transaction dan Interim Product	0,24	TB	0,24	TB	0,00	TB	0,00	TB
Rata-Rata	0,21		0,17		0,14		0,11	

Dapat ditarik kesimpulan bahwa tidak ada hubungan antara perencanaan dengan pelaksanaan; pelaksanaan dengan hasil; perencanaan, pelaksanaan dengan hasil; terutama terkait dengan 5 DS. Temuan ini menjadi kelebihan dari model EDTS. Penilaian melalui matrik penilaian tidak ditemukan masalah, namun melalui matrik dekripsi, dapat terungkap.

Informasi yang dihasilkan dari kegiatan evaluasi diperkuat melalui pendekatan triangulasi sumber dan triangulasi jenis data. Triangulasi sumber berasal dari guru sendiri, guru sejawat dan peneliti. Peneliti berfungsi sebagai *cross check* data, bila terdapat perbedaan antara guru sendiri dan guru sejawat. Setelah data diklarifikasi, maka informasi yang dihasilkan digunakan untuk perbaikan terhadap pembelajaran.

Triangulasi data terdiri dari data observasi guru sendiri, data observasi guru sejawat, dan data diskusi. Klarifikasi data pada saat diskusi, mempunyai fungsi ganda. Fungsi pertama untuk perbaikan terhadap instrumen model, dan kedua untuk menge-

tahui kualitas pembelajaran guru. Pada saat mencocokkan data yang berbeda, maka satuan guru akan memberikan alasan. Terjadilah proses penyelarasan kesepakatan antarguru IPA dengan sepengetahuan kepala sekolah, apakah menurunkan atau menaikkan nilainya. Melalui proses diskusi ini, kepala sekolah dapat mengetahui kejadian pembelajaran yang terjadi.

Informasi diperoleh dari kegiatan feedback dan evaluasi diri

Feedback pada saat diskusi dan pasca diskusi. *Feedback* dari guru sendiri, guru sejawat dan kepala sekolah ditujukan kepada guru IPA sendiri. Instrumen *feedback* lebih mengarah pada substansi terkait dengan 5 DS, kegiatan pembelajaran, dan respon terhadap proses diskusi.

Data penilaian *feedback* oleh guru sendiri, kurang baik (14%), baik (71%), sangat baik (15%). Penilaian *feedback* oleh guru sejawat, baik (57%), sangat baik (43%). Penilaian *feedback* oleh kepala sekolah, baik (71%), sangat baik (29%). Nilai total penilai-

an *feedback* adalah: baik (71%), sangat baik (29%). Hasil *feedback* belum mendapatkan nilai sangat baik, karena kontens *feedback* lebih banyak terkait dengan 5 DS. Di lapangan 5 DS, sebagian sudah terlaksana, meskipun belum lengkap. Domain yang belum, sebagian besar adalah kreativitas dan aplikasi.

Informasi hasil evaluasi dapat diperoleh dari kegiatan evaluasi diri, yang ditulis oleh guru IPA sendiri. Data evaluasi diri diperoleh dari kegiatan klarifikasi pada tahap diskusi dan data *feedback*. Data *feedback* diperoleh dari kepala sekolah, guru sejawat dan guru IPA sendiri. Informasi ini digunakan untuk melakukan perbaikan terhadap pembelajaran yang dilakukannya. Guru sendiri menulis item perbaikan apa saja yang perlu dilakukan pada pembelajaran berikutnya.

Berikut data evaluasi diri yang disampaikan oleh guru IPA sendiri, dilengkapi frekuensinya sebagai berikut: (1) guru perlu membuat modul dan bank soal yang disusun sendiri (29%); (2) perlunya kesesuaian indikator kompetensi dengan indikator penilaian (100%); (3) materi perlu dirinci, sesuai urutan proses pembelajaran (29%); (4) pengelolaan laboratorium belum profesional, karena semuanya dilaksanakan oleh guru sendiri (29%); (5) indikator dalam RPP supaya lebih dirinci lagi, sehingga dapat diukur (29%); (6) mengkaitkan materi pembelajaran dengan hal-hal yang mutakhir, masa depan siswa, dan persoalan yang dihadapi siswa (71%); (7) penulisan metode harus dirinci, sesuai dengan kegiatan yang harus dilakukan (14%); (8) perlunya mengoptimalkan lingkungan sebagai sumber belajar dan laboratorium pembelajaran (71%); (9) laboratorium perlu lebih ditingkatkan terkait dengan fungsinya sebagai sumber belajar (29%); (10) perlu adanya kesesuaian antara perencanaan yang tertuang di RPP, dengan pelaksanaannya (29%); (11) siswa perlu membuat laporan setiap selesai kegiatan (14%); (12) perlu adanya kesesuaian alat evaluasi yang direncanakan dengan pelaksanaan (14%); (13) siswa perlu lebih diberi dorongan untuk aktif dalam ber-

diskusi (14%); (14) perlunya memberikan tambahan muatan di atas standar minimal KD (14%); (15) pemberian penilaian pada sebelum dan saat pembelajaran perlu dilakukan, baik melalui lisan atau tulisan (14%); (16) pemilihan media lebih diarahkan pada media yang mampu mengembangkan ide siswa, dan masa depan siswa (43%); (17) pada pembelajaran metode “demonstrasi”, semua siswa sebaiknya dilibatkan secara aktif, agar memahami apa yang disampaikan oleh guru (14%); (18) perlunya mengungkap apersepsi siswa dan memotivasi siswa (14%); (19) lebih meningkatkan rasa percaya diri dalam proses kegiatan belajar mengajar walaupun bukan bidang studinya (Fisika), (29%); (20) perlu kesesuaian pembagian waktu mengajar, antara RPP dengan pelaksanaan (14%); (21) sebaiknya guru menekankan pentingnya materi terhadap penerapan/aplikasi dalam kehidupan sehari-hari yang dialami siswa (14%); (22) guru perlu melibatkan kreativitas peserta didik (14%).

Temuan oleh Peneliti

Keterlibatan peneliti sebagai *cross check* data, mengfungsikan diri sebagai triangulasi sumber pada kegiatan observasi. Dari kegiatan ini, peneliti memperoleh banyak temuan. Baik pada saat persiapan sebelum pembelajaran dimulai, atau pada saat proses pembelajaran berlangsung; diantaranya: 1) Terjadinya pemahaman yang kurang tepat, dari masing-masing komponen RPP. Meskipun SMP berada dalam 1 (satu) wilayah, namun guru dalam mengimpretasikannya dapat berbeda; 2) Ditemukannya RPP yang tidak dibuat sendiri oleh guru, artinya guru hanya mem-*fotocopy* RPP yang dimiliki oleh temannya. Hal inilah yang memberikan gambaran terjadinya ketidaksesuaian antara realita dengan pelaksanaan; 3) Belum adanya kesesuaian antara indikator kompetensi, dengan indikator penilaian, dari semua RPP yang telah disiapkan oleh guru; 4) Ditemukan urutan penempatan komponen RPP yang beragam.

Pada saat guru mengajar, diperoleh beberapa temuan, diantaranya: (1) Penyam-

paian konsep yang kurang tepat; (2) Kekurangpercayaan guru biologi ketika menyampaikan materi fisika; (3) Adanya kecenderungan guru untuk menyampaikan materi yang pokok-pokok, terutama yang terkait dengan soal-soal UAN. Hal ini memberikan kecenderungan guru kurang memanfaatkan fungsi laboratorium secara optimal; (4) Guru masih lemah untuk mengkaitkan pembelajaran IPA dengan lingkungan dan kehidupan sehari-hari siswa; (5) Adanya kecenderungan kekurangan kemampuan guru untuk menyampaikan materi secara komprehensif, dengan mengkaitkan materi lainnya yang kemungkinan berhubungan. Konsep inilah yang akan tercapai manakala 5 DS diimplementasikan; (6) Kemampuan siswa dalam menyampaikan pendapat dan berdiskusi masih lemah, sehingga suasana diskusi kurang hidup; terutama pada kelas satu; (7) Pada kegiatan demonstrasi, siswa kurang fokus dan kurang menghargai. Hal ini memberi gambaran, perlunya siswa diberi pembelajaran berdiskusi, bekerja sama dan saling menghargai.

Selain temuan-temuan lapangan di atas, juga diperolehnya temuan terkait dengan karakteristik guru dalam menilai. Temuan yang diperoleh peneliti adalah sebagai berikut: (1) Guru sendiri melakukan penilaian, relatif menilai lebih tinggi dibandingkan dengan guru sejawat; (2) Masih terdapatnya subyektifitas yang tinggi oleh guru sejawat dalam melakukan penilaian, misalnya rasa *pekewah*, *sungkan*, dll. Persoalan ini, dapat diatasi oleh model EDTS. Kelebihan model EDTS ini adalah, karena adanya matrik dalam bentuk penilaian dan deskripsi. Pada saat penilaian dalam bentuk angka tidak terdapat persoalan, namun akan diperoleh kesenjangan, dan persoalan pada matrik deskripsi. Sehingga kepala sekolah masih dapat diperoleh temuan persoalan; (3) Masih adanya keraguan memberi nilai, karena ada indikasi disilang dan dilingkari.

Menguji keefektifan model EDTS

Suatu model yang disusun agar dapat disebarluaskan pemakaiannya, harus diuji

keefektifannya. Kualitas model dapat dilihat dari data validitas termasuk reliabilitasnya (konsistensi), dan kepraktisan (keduanya sebagai prasarat keefektifan), serta efektivitas (kriteria evaluasi model). Kepraktisan harus memenuhi kebutuhan, keinginan, dan kendala kontekstual pada sasaran. Efektivitas harus memberikan dampak positif pada peningkatan kualitas peserta didik, profesionalisme guru dan interaksi keduanya untuk mencapai tujuan. Efektivitas dapat dilakukan melalui evaluasi formatif proses pengambilan keputusan (perencanaan, pengumpulan data, dan pelaporan), McKenney, Nieveen, & Akker, 1999, pp.25-36). Evaluasi formatif sebagai evaluasi proses *prototyping*, dilakukan melalui partisipasi pengguna dan ahli. Partisipasi pengguna diperoleh dari daftar informasi melalui pengamatan, kuesioner, *logbook*, saran *user* maupun pengisian instrumen. Partisipasi ahli dapat melalui penilaian dan saran. Hasil dari evaluasi formatif digunakan untuk revisi yang berdampak pada kualitas model, (Nieveen, 2007, pp.82-102).

Validitas dan reliabilitas

Data validitas diperoleh dari hasil penilaian oleh *expert*. Validasi meliputi: model dan isi. Validitas adalah sejauh mana ketepatan dan kecermatan suatu alat ukur dalam melakukan fungsi ukurnya, atau memberikan hasil ukur yang sesuai dengan maksud dilakukannya pengukuran. *Expert judgment* terdiri dari: 2 pakar bidang evaluasi, 2 pakar pengukuran, dan 1 pakar pendidikan sains.

Validitas Muka

Validitas tampak adalah penilaian terhadap format penampilan dari model EDTS. Apakah model telah meyakinkan dan memberi kesan mampu mengungkap apa yang hendak diukur, maka dapat dikatakan bahwa validitas muka (*Face validity*) telah terpenuhi. Penilaian validitas muka dianalisis secara kuantitatif dan kualitatif. Penilaian ahli terkait dengan model EDTS, disajikan pada Tabel 3.

Tabel 3. Penilaian Ahli terhadap Model

Dimensi	Instrumen	Ahli 1	Ahli 2	Ahli 3	Ahli 4	Ahli 5	Rerata model	Kategori model
Model Evaluasi	Cakupan model	3,8	3,2	3,1	3,1	3,6	3,4	Sangat Baik
	Kemanfaatan	4,0	3,4	3,0	3,6	3,8	3,6	Sangat Baik
	Kepraktisan	3,5	3,0	3,3	3,0	3,3	3,2	Sangat Baik
	Keefektifan	3,7	3,3	3,0	3,3	3,3	3,3	Sangat Baik
Panduan Model	Petunjuk	3,3	3,0	3,0	3,0	3,0	3,1	Baik
	Alur Evaluasi	3,3	3,0	3,0	3,0	3,0	3,1	Baik
Instrumen	Isi Instrumen	3,8	3,6	3,0	3,0	3,2	3,3	Sangat Baik
	Bahasa Instrumen	3,5	3,0	3,0	3,3	3,3	3,2	Sangat Baik
Rerata/ahli		3,6	3,2	3,0	3,2	3,3	3,3	Sangat Baik
Katagori/ahli		Sangat Baik	Sangat Baik	Baik	Sangat Baik	Sangat Baik	Sangat Baik	

Rata-rata penilaian terhadap model EDTS adalah sangat baik, kecuali panduan model (petunjuk dan alur evaluasi) nilainya baik. Nilai “sangat baik” terhadap model evaluasi (cakupan model, kemanfaatan, kepraktisan, keefektifan), dan instrumen (isi dan bahasa instrumen). Dari kelima ahli, 4 (empat) ahli rerata skornya berada pada kategori SANGAT BAIK. Data selengkapnya tertera pada Tabel 3.

Validitas Isi

Validitas isi merupakan pengujian terhadap isi instrumen, melalui *professional judgment*. Tujuannya adalah untuk menilai apakah item-item instrumen telah mencakup keseluruhan kawasan isi objek yang hendak diukur. Apakah setiap butir instrumen yang ada dalam model, isinya telah relevan.

Tabel 4. Rekapitulasi Data Penilaian Validitas Isi

MATRIK PENILAIAN						
Indikator	No item	Ahli 1	Ahli 2	Ahli 3	Ahli 4	Ahli 5
Tujuan pembelajaran	1	1	2	2	2	2
Kegiatan pembelajaran	1	1	2	2	2	2
Perencanaan	1	2	2	1	2	2
	2	2	2	1	2	2
	3	2	2	1	2	2
Isi perencanaan	1	2	2	1	2	2
Kesiapan laboratorium	1	1	2	2	2	2
Pengelolaan laboratorium	1	1	2	2	2	2
Pendayagunaan media	1	1	2	2	2	2
Nilai	1	1	2	2	2	2
MATRIK DESKRIPSI						
Kreativitas	24	2	2	2	1	2
Penerapan	34	1	2	2	2	2
Hubungan Antecedent dengan Transaction	13	1	2	2	1	2
	14	1	2	2	1	2
	15	2	2	2	1	2

Skala penilaiannya adalah: 2 (dua) jika relevan, 1 (kurang relevan), 0 (tidak relevan). Jika kurang/tidak relevan, apa saran yang direkomendasikan oleh ahli. Penilaian dilakukan 57 item butir pada tahap pra pembelajaran, dan 82 butir pada tahap proses pembelajaran. Jadi pada matrik penilaian ada 139 butir (untuk matrik penilaian), 56 butir untuk matrik deskripsi. Matrik deskripsi terdiri dari *congruencies* dan *contingencies*. *Congruencies* terdiri dari 38 yang dianalisis dan *contingencies* terdiri dari 18 unit yang dianalisis.

Data memberikan gambaran bahwa 211 dinilai relevan oleh ahli, dan hanya 8 butir instrumen yang dinilai kurang relevan. Hanya 1 (satu) dari 5 (lima) ahli yang menilai 1 (kurang relevan). Sebagian besar pakar telah menyetujui.

Saran ahli terkait dengan item yang perlu diperbaiki, didasarkan dari hasil penilaian yang kurang relevan dengan nilai 1 (satu), sebagai berikut: (1) Tujuan pembelajaran memiliki kesesuaian dengan indikator kompetensi. Karakteristik pembelajaran IPA adalah adanya 5 DS dalam perencanaan dan pembelajaran. Hal inilah yang mendasari perlunya penambahan 5 DS dalam item tujuan pembelajaran; (2) Kegiatan pembelajaran perlu adanya penambahan yang berisi pendahuluan, inti dan penutup. Penambahan kalimat ini akan memperkuat instrumen, maka peneliti perlu melakukan penambahan kalimat; (3) Perencanaan penilaian harus ada baik pada sebelum, pada saat dan setelah pembelajaran. Butir instrumen ini tetap dipertahankan, karena penilaian pembelajaran IPA dapat dilakukan melalui tes tulis maupun non tes (lisan, penugasan, tanya jawab, diskusi, dll). Guru dapat mengetahui sejak awal kemampuan siswa, sehingga dapat mengambil alternatif pemecahan agar siswa dapat lebih memahami; (4) Isi perencanaan: ada perencanaan tentang ulangan harian. Butir ini perlu adanya penambahan baik di kelas maupun di laboratorium (baik dalam ruangan maupun lingkungan/alam terbuka. Pembelajaran IPA dapat dilakukan dalam ruang maupun di luar kelas, maka perlu adanya penambahan

kalimat pada instrumen tersebut dengan: baik dalam kelas maupun di laboratorium; (5) Teori belajar konstruktivis adalah suatu pembelajaran yang secara aktif membangun makna belajar pada siswa. Menurut perspektif konstruktivis, siswa dipahamkan dengan pengetahuan sebelumnya dan memahami temuan untuk membangun pengetahuan baru. Konstruktivis memberi rujukan dan dihimbau untuk memberikan peluang yang menantang bagi siswa agar mengembangkan pengetahuan tentang dunia sekitar mereka. Mencermati pengertian ini maka item “Guru mampu meng-*explore* emosi dan pengetahuan peserta didik, sehingga terbentuk pembelajaran konstruktivis”, ini tetap dipertahankan; (6) *Active learning*, adalah pembelajaran yang menghendaki agar melibatkan siswa dalam pembelajaran, pemecahan masalah, pemilihan media, dan memberikan usul serta saran. Guru mendekati siswa pada realita pengalaman, temuan persoalan di lapangan, mendorong siswa untuk menumbuhkan rasa ingin tahu yang tinggi, serta memiliki kemampuan untuk mengaitkan dan mengaplikasikan sains dalam kehidupannya. Hal ini akan mengarah pada pengembangan kreativitas dan aplikasi sains siswa. Terkait dengan keterangan ini, maka item “pemilihan dan penggunaan media yang mengarah pada penumbuhan pengembangan ide peserta didik” dan ”materi disesuaikan dengan perkembangan mutakhir menuntut guru mendorong peserta didik berfikir pada tingkat yang lebih tinggi”, masih dipertahankan; (7) Hubungan antara: (a) kesiapan peserta didik meliputi apersepsi, motivasi untuk keperluan pembelajaran; (b) kesiapan guru untuk mengimplementasikan 5 DS; (c) ketersediaan sarana dan prasarana untuk mendukung pembelajaran berbasis 5 DS, dengan pelaksanaannya pada proses pembelajaran, sangat penting untuk dinilai. Bila antara kesiapan dengan pelaksanaan tidak ada hubungan, maka tujuan pembelajaran tidak akan terpenuhi. Hal inilah yang mendasari item ini tetap dipertahankan.

Saran Ahli Terkait dengan Model dan Instrumen

Saran dari ahli meliputi: saran terkait pembelajaran sains, terkait pelaksanaan 5 DS, dan instrumen. Berdasarkan saran yang disampaikan dilakukan perbaikan.

Perbaikan terkait dengan instrumen model, ada penambahan 5 butir instrumen yaitu: 1 butir pada tahap prapembelajaran, dan 4 butir instrumen pada tahap proses pembelajaran. 5 (lima) item tambahannya adalah sebagai berikut: (1) Pada tahap pra pembelajaran item “kegiatan pembelajaran” butir instrumennya ditambah “Berisi eksplorasi, elaborasi dan konfirmasi”. Jumlah butir sebelumnya 57 butir menjadi 58 butir instrumen; (2) Pada tahap proses pembelajaran ada penambahan 4 (empat) butir instrumen. Sehingga jumlah butir yang semula 82 butir menjadi 86 butir instrumen. Penambahan butir-butir instrumennya adalah sebagai berikut: (a) Pada dimensi pengelolaan laboratorium, ada penambahan 2 butir instrumen yaitu: Tersedia kotak PPPK dalam ruang laboratorium, penempatan peralatan dan bahan telah tertata dengan rapi, (b) Pada dimensi pendayagunaan media, ada penambahn 1 butir yaitu: guru terampil memanfaatkan lingkungan sebagai laboratorium dan sumber belajar; (c) Pada dimensi

kegiatan pembelajaran, ada penambahan 1 butir instrumen yaitu: kegiatan pembelajaran sistematis, dan runtut melalui proses, eksplorasi, elaborasi, dan konfirmasi.

Reliabilitas Antarrater (Kappa)

Reliabilitas ini dipakai untuk menilai konsistensi dua orang rater dalam menilai performansi individu melalui *checklist* yang menghasilkan data nominal. Mengetahui kesamaan antara observer terhadap emosi subjek. Data yang dihasilkan adalah data nominal, yaitu 0 (tidak ada/tidak sesuai), 1 (kurang sesuai), 2 (ada/sesuai). Semakin banyak kemiripan hasil penilaian antara satu rater dan rater lainnya maka koefisien reliabilitas yang dihasilkan akan tinggi.

Fleiss (1981: 378-382) mengkategorikan tingkat reliabilitas antarrater menjadi empat kategori, yaitu:

- Kappa < 0,4 : buruk (*bad*)
- Kappa 0,4 – 0,6 : cukup (*fair*)
- Kappa 0,6 – 0,75: memuaskan (*good*)
- Kappa > 0,75 : istimewa (*excellent*).

Berdasarkan hasil perhitungan untuk menguji reliabilitas antarrater, dan dibandingkan dengan koefisien kappa, (Gwet, (2012: 224), diperoleh data tertera pada tabel berikut.

Tabel 5. Rekapitulasi Hasil Analisis Reliabilitas antarrater

		Matrik Penilaian		Matrik Deskripsi		Kategori
		Supervisi Prapembelajaran	Supervisi Proses Pembelajaran	Congruencies	Contingencies	
SMPN 14	K	0,684	0,616	0,678	0,684	K=0,67 (Good)
	Sig	0,002	0,000	0,000	0,002	
SMPN 29	K	0,609	0,609	0,722	0,609	K=0,61 (Good)
	Sig	0,005	0,000	0,000	0,005	
SMPN 15	K	0,605	0,626	0,622	0,769	K=0,66 (Good)
	Sig	0,000	0,000	0,000	0,001	
SMPN 07	K	0,750	0,620	0,725	0,750	K=71 (Good)
	Sig	0,001	0,000	0,000	0,001	
SMPN 32	K	0,675	0,831	0,630	0,675	K=0,7 (Good)
	Sig	0,002	0,000	0,000	0,002	
SMPN 19	K	0,620	0,610	0,657	0,620	K=0,63 (Good)
	Sig	0,004	0,000	0,000	0,004	
SMPN 11	K	0,620	0,630	0,604	0,620	K=62 (Good)
	Sig	0,004	0,000	0,000	0,004	

Untuk mengetahui ada dan tidaknya hubungan antara rater guru sendiri dan guru sejawat, didasarkan pada pengambilan keputusan berdasarkan: H0: tidak ada hubungan antara rater guru sendiri dan guru sejawat. H1: Ada hubungan antara rater guru sendiri dan guru sejawat.

Jika nilai sig >0,05 maka H0 diterima, demikian juga sebaliknya. Berdasarkan perhitungan di atas baik nilai sig untuk tiap unit, total tiap unit yang dianalisis, maupun antarrater *self* dan *peer* guru SMP, dan totalnya didapatkan data nilai sig-nya di bawah 0,05. Nilai kappa berada pada rentang 0,6–0,75 (memuaskan/*good*). Hal ini dapat disimpulkan bahwa “ada hubungan antarrater” antara guru sendiri dan guru sejawat, atau dengan kata lain antarrater “ada kesepakatan”.

Kepraktisan model EDTS

Kepraktisan harus memenuhi kebutuhan, keinginan, dan kendala kontekstual pada sasaran. Kemudahan user menggunakan model melalui intervensi pengguna dan ahli, serta evaluasi *try out*. Perbaikan model berdasarkan saran dari hasil uji coba di lapangan. Setiap selesai satu kali uji coba, selalu diperoleh saran, baik yang disampaikan oleh guru atau kepala sekolah. Perbaikan model dilakukan secara terus menerus, sehingga dihasilkan model yang dapat memberi kemudahan bagi user untuk menggunakan. Perbaikan model meliputi alur pelaksanaan, instrumen, teknik pemakaian model, dan persoalan lapangan.

Keefektifan model EDTS

Keefektifan model dapat dilihat dampak positif yang dihasilkan dan sesuai dengan yang diharapkan/tujuan (Akker, 2007, pp.37-52). Model EDTS pada supervisi pembelajaran IPA akan efektif, bila dapat memberikan dampak positif terhadap perbaikan pembelajaran. Keefektifan dapat dicermati dari kemudahan *user* mengolah data evaluasi untuk menghasilkan informasi,

serta menindaklanjuti informasi untuk perbaikan pembelajaran.

Hasil uji coba model EDTS, telah menghasilkan data *feedback* dan evaluasi diri yang digunakan untuk perbaikan pembelajaran IPA. Kegiatan *feedback* meliputi proses tindak lanjut pascaobservasi (setelah guru melakukan observasi), maupun *feedback* pascakegiatan diskusi (sebagai dasar guru melakukan evaluasi diri). Melalui supervisi akademik yang berlangsung selama ini, data observasi belum ditindaklanjuti oleh kepala sekolah dengan maksimal (Winaryati, 2011-b). Melalui model EDTS persoalan ini dapat diatasi.

Ketiga kriteria kualitas model ini menjadi landasan, bahwa model EDTS dapat dijadikan alternatif pemecahan masalah terkait dengan pelaksanaan supervisi akademik. Pelaksanaan supervisi maupun PKG selama ini adalah guru senior mensupervisi/menilai guru junior, dan guru senior disupervisi/dinilai oleh kepala sekolah. Sementara model EDTS ada penambahan: (1) penilaian oleh guru sendiri; (2) dihasilkannya dukungan informasi dari guru bagi supervisor; (3) proses validasi oleh kepala sekolah berdasarkan kesenjangan data yang ada pada guru; (4) ada kegiatan klarifikasi melalui kegiatan diskusi; (5) terlaksananya proses *feedback*/tindak lanjut; (6) evaluasi diri yang dilakukan oleh guru sendiri, sebagai bentuk komitmen untuk melakukan perbaikan pembelajaran berikutnya.

Perbaikan pembelajaran merupakan bagian dari tujuan supervisi. Data informasi perbaikan pembelajaran menjadi landasan bahwa melalui model EDTS dengan pendekatan ANTRANINPRO efektif untuk mengevaluasi supervisi pembelajaran IPA. Melalui model EDTS ini, guru akan semakin tahu kelebihan dan kekurangannya, sehingga profesionalisme guru meningkat, kualitas pembelajaran semakin baik, dan berdampak pada kualitas lulusan.

Simpulan dan Saran

Simpulan

Berdasarkan uraian, analisis data dan pembahasan, dapat disimpulkan hal-hal sebagai berikut: (1) penelitian ini menghasilkan model evaluasi baru yaitu model evaluasi diri dan teman sejawat (EDTS), yang dikembangkan dengan pendekatan AN-TRANINPRO (*Antecedent, Transaction, Interim Product*); (2) model EDTS disusun untuk memperkuat pelaksanaan supervisi di sekolah. Melalui laporan diri yang diberikan oleh guru sendiri dan guru sejawat, dihasilkan informasi diri pada guru sendiri, sebagai dukungan informasi kepada kepala sekolah (supervisor), untuk memperkuat pelaksanaan supervisi akademik. Berdasarkan hasil evaluasi diperoleh penilaian kurang baik pada item yang terkait dengan lima domain sains. Domain yang masih menjadi persoalan bagi guru untuk diaplikasikan, adalah domain kreativitas dan aplikasi sains.

Model EDTS efektif untuk digunakan, berdasarkan data: (a) validitas model (sangat baik), validitas konten (96% relevan), reliabilitas (kriteria baik), diperkuat dengan pendekatan triangulasi sumber dan jenis data, dan konten dari model EDTS dilengkapi dengan instrumen berbasis lima domain sains, dengan tahapan evaluasi yang rinci; (b) model EDTS memiliki nilai kepraktisan, meliputi kesesuaian keterlaksanaan model dengan panduan model, memberi kemudahan user menggunakan, waktu yang dibutuhkan kepala sekolah untuk mensupervisi lebih sedikit, penilaian lebih obyektif, serta dapat digunakan untuk menyelesaikan persoalan supervise; (c) model EDTS memiliki nilai keefektifan, karena memberi dampak positif terhadap perbaikan pembelajaran, melalui kegiatan *feedback* dan evaluasi diri.

Penelitian ini memberikan sumbang-an teori baru yaitu: (1) Menghubungkan *mapping handbook* supervisi dengan evaluasi. Kegiatan evaluasi menghasilkan informasi, data informasi yang dihasilkan dioptimalkan oleh supervisi untuk perbaikan pembelajaran berikutnya; (2) model evaluasi ANTRA-

NINPRO, yang terdiri dari matrik penilaian (*judgement*) dan matrik deskripsi (*congruencies* dan *contingencies*); (3) memperkuat dan memperbaiki pelaksanaan PKG, karena penilaian melibatkan guru sendiri dan guru sejawat, serta menempatkan kepala sekolah sebagai fasilitator dalam diskusi dan validator terhadap data penilaian; (4) model supervisi berbasis mata pelajaran, yaitu supervisi yang diselarasakan dengan karakteristik mata pelajaran.

Rekomendasi

Berdasarkan kesimpulan di atas, rekomendasi yang dapat dirumuskan sebagai berikut: (1) Berdasarkan rekomendasi hasil penilaian terhadap kualitas model, dapat disimpulkan bahwa model EDTS perlu untuk direkomendasikan pelaksanaannya pada supervisi di sekolah, melalui berbagai bentuk kegiatan, seperti: pelatihan, *workshop*, diskusi, dll; (2) Berdasarkan persoalan supervisi yang sekarang ini dilaksanakan, maka model EDTS direkomendasikan penggunaannya, untuk memperkuat kegiatan supervisi akademik di sekolah, agar kualitas pembelajaran meningkat.

Daftar Pustaka

- Akker, J.V.D. (2007). Curriculum Design Research dalam an introduction to educational design research. *Proceedings of the seminar conducted at the East China Normal University, Shanghai (PR China), November 23-26*. P. 37-52.
- Cooper, N., & Forrest, K. (2009). *Essential guide to educational supervision in postgraduate medical education*. Chennai, India: Wiley-Blackwell. Macmillan Publishing Solutions.
- Diat Prasojo, L. (2011). *Supervisi pendidikan*. Yogyakarta: Gova Media.
- Evans, C.S. (2009). *Essential guide to educational supervision in postgraduate medical education* edited by N. Co-

- per and K. Forrest. Blackwell Publishing Ltd. ISBN: 978-1-405-17071-0.
- Fernandes, H.J.X. (1984). *Evaluation of educational program*. Jakarta: National Education Planning Evaluation and Curriculum Development.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (Edisi kedua). New York: John Wiley.
- Gwet, K.L. (2012). *Handbook, Inter-Rater Reliability*. 3th Ed. USA: Gaithersbourg.
- 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.
- Isaac, S., & Michael, W.B. (1981). *Handbook in research and evaluation*. 2th Ed. San Diego, California: EdITS Publisher.
- Kaufman, R. (2006). Defining and delivering measurable value: A mega Thinking and planning primer. *Performance Improvement Quarterly*, 18(3) pp. 6-16. Florida State University.
- Kaufman, R., & Thomas, S. (1980). *Evaluation without fear*. New York. London: New Viewpoints
- Middendorf, J., & Kalish, K. (1995). *Peer evaluation of teaching*. TRC Newsletter. Campus international consulting.
- McCormack, A.J. (1999). *Trends & issues in science curriculum. Science curriculum resourcebook: a practical guide for K-12 science curriculum*. New York: Kraus International.
- McKenney, S., Nieveen, N., & van den Akker, J. (2002). Computer Support for curriculum developers CASCA-DE. *ETRE&D*, Vol. 50, No. 4, pp. 25–35 ISSN 1042–1629
- Nieveen, N. (2007). Formative evaluation in educational design research. *Proceedings of the seminar conducted at the East China Normal University, Shanghai (PR China), November 23-26*. Hal. 89-102.
- Owston, R. (2008). Models and methods for evaluation. *York University, Toronto, Canada*.
- Ozogul, G., Olina, Z., & Sullivan, H. (2008). Teacher, self and peer evaluation of lesson plans written by preservice teachers. *Education Tech Research Dev* 56:181–201.
- Provus, M., Malcolm. (1969). *The discrepancy evaluation models. An approach to local program improvement and development*. Pittaburgh Public School.
- Savoie. (2010). *A Guidebook for peer evaluation*. College of the Arts. Valdosta State University
- Sibbald, T. (2009). The relationship between lesson study and self-efficacy. (Report). *Thames Vallery District School Board. School Science and Mathematics: Gale Cengage Learning*. Diunduh 10 Oktober 2010.
- Stake, RE. (1967). The countenance of educational evaluation. *Teacher's Coolege Record*. Vol. 68, no:7.
- Stinson. (2000). "Quality supervision.", "The quality of supervisor-provider interactions in zimbabwe," *Operations Research Results* vol. 1, no. 5 (Bethesda, MD: Quality Assurance Project).
- Stufflebeam, D.L., & Shinkfield, A.J. (1984). *Systematic evaluation a self-instructional guide to theory and practice*. Boston: Kluwer-Nijhoff Publsing.
- Stufflebeam, D.L. (1973). Educational evaluation: theory and practice. In B. R. Worthen & J. R. Sanders (Eds.), *Evaluation as Enlightenment for Decision-*

- Making*. Boston: Kluwer-Nijhoff Publishing
- Tim IKIP Bandung. (1999). *Supervisi akademik*. Bahan Sajian untuk Pelatihan Pengawas SLTP dan SMU.
- Wiles, K., & Lovell, J.T (1975). *Supervision for better schools*. New Jersey: Prentice-Hall.
- Winaryati, E. (2011a). Kompetensi kepala sekolah sebagai supervisor pada pembelajaran IPA SMP di kota Semarang. *Univ Muhammadiyah Semarang (UNIMUS)*.
- Winaryati, E. (2011b). Peran guru IPA SMP dalam supervisi akademik untuk mewujudkan pembelajaran bermakna di kota Semarang. *Univ Muhammadiyah Semarang (UNIMUS)*.
- Yonnie Chyung. (2001). Improve the motivational appeal of online instruction for adult learners: what's in it for me?. [*The American Journal of Distance Education \(AJDE\)*](#). Volume 15; Issue 3, p. 36-49.
- Yonnie Chyung., Winiecki, D., & Fenner, J.A. (2006). Evaluation of effective interventions to solve the dropout problem in adult distance education. *Instructional & Performance Technology*. College of Engineering Boise State University.