

MODEL EVALUASI PEMBELAJARAN BERBASIS *KAIZEN* DI SEKOLAH MENENGAH ATAS

¹Primardiana Hermilia Wijayati, ²Suyata, ³Sumarno
¹Universitas Negeri Malang, ^{2,3}Universitas Negeri Yogyakarta
¹phwijayati@gmail.com, ³sumarno.wonosidi@gmail.com

Abstrak

Tujuan penelitian adalah mengembangkan model evaluasi pembelajaran yang hasilnya dapat digunakan untuk mendeteksi kekeliruan dan melakukan koreksi sendiri, serta sistem informasinya. Pengembangan dilakukan melalui pendekatan kuantitatif dan kualitatif dalam empat tahap, yaitu penelitian, penciptaan dan desain, rekayasa dan pengemasan, tahap pengujian dan evaluasi. Konstruksi model evaluasi dimodifikasi dari model Marzano dan divalidasi melalui *focus group discussion* dan teknik Delphi. Subjek penelitian terdiri dari guru dan siswa di enam SMAN dan satu SMA swasta di Malang. Hasil penelitian: 1) model evaluasi didukung oleh instrumen evaluasi diri guru, evaluasi teman sejawat, dan evaluasi siswa; 2) karakteristik instrumen evaluasi mencakup validitas, reliabilitas, dan kepraktisan telah teruji; 3) sistem informasi hasil evaluasi disajikan dalam bentuk *bar chart* yang memuat informasi kelebihan dan kekurangan guru, rencana perbaikan guru, dan saran perbaikan untuk guru dan sekolah; 4) instrumen evaluasi berbentuk *software* disertai manualnya.

Kata kunci: *model evaluasi, kaizen, pembelajaran, sekolah menengah atas*

A TEACHING EVALUATION MODEL BASED ON *KAIZEN* AT SENIOR HIGH SCHOOL

Abstract

This study aimed at developing an evaluation model to find out problems in the teaching process in order to prevent mistakes and to make a self-correction for assuring a continuous teaching improvement, and to deliver its information system. This was a development study which used quantitative and qualitative approaches carried out in four phases, namely initial research, invention and design, engineering and packaging, and test and evaluation. The construct of this evaluation model was adapted from Marzano's evaluation model and had been validated by a focus group discussion and Delphi technique. The subjects of this study were senior high school teachers and students from six public senior high schools (SMAN) and a private senior high school in Malang. This study shows the following conclusions: 1) The evaluation model based on *kaizen* has been equipped by self-evaluation instrument, evaluation instrument by colleagues and students; 2) The characteristics of the evaluation model based on *kaizen* for quality assurance of teaching at senior high school consisted of validity, reliability, and practicability which had been verified. The validity evidence had been obtained through focus group discussion and Delphi techniques, and had been analyzed using *content validity ratio* (CVR) and *content validity index* (CVI) proposed by Lawshe. The construct validity was elaborated through the structural equation modeling based on component; 3) The results of model fit test using GesCA-program developed by Heungsun Hwang indicate that the model is supported by the data shown by the value of GFI 0,997, and SRMR 0.076; 4) The information system of the evaluation results is presented in bar chart which accommodates teachers' strengths and weaknesses, teachers' planning of improvement, and suggestion of improvement for teachers and school.

Keywords: *development model, evaluation model, teaching process*

Pendahuluan

Pembelajaran adalah proses interaksi peserta didik dengan guru dan sumber belajar pada suatu lingkungan belajar. Tujuan utama sebuah pembelajaran adalah pencapaian prestasi siswa yang tinggi. Mutu pembelajaran dapat dicapai apabila guru yang menjadi ujung tombak pembelajaran memiliki mutu yang baik pula. Menurut Business Council of Australia (2008, p.7) mutu pembelajaran merupakan penggerak utama keberhasilan belajar siswa. Direktorat Tenaga Kependidikan (2008, p.1) menjelaskan bahwa begitu pentingnya peran guru dalam mentransformasikan *input-input* pendidikan, sampai-sampai banyak pakar menyatakan bahwa di sekolah tidak akan ada perubahan atau peningkatan kualitas tanpa adanya perubahan dan peningkatan kualitas guru.

Mutu adalah apapun yang bisa dikembangkan. Dalam konteks ini, mutu berhubungan tidak hanya dengan produk dan jasa tetapi dengan cara orang-orang bekerja, cara mesin dijalankan, dan cara sistem dan prosedur saling berkaitan (Imai, 2008, p.49). Mutu meliputi semua aspek tingkah laku manusia. Mendefinisikan mutu dalam mengajar sangatlah sulit, karena selalu menuntut pertimbangan nilai yang menimbulkan banyak perbedaan pendapat (Berliner, 2005, p.206). Dalam penelitiannya yang dilakukan di perguruan tinggi, Deepwell (2007, p.35) mengemukakan bahwa mutu dipandang berbeda oleh orang yang berbeda dan mempertahankan mutu merupakan ciri dari stabilitas. Menurut Dinham, Ingvarson, & Kleinhenz (2008) mutu pembelajaran merupakan penggerak utama keberhasilan belajar siswa. Sementara Jackson dan Bruegmann (2009) mendukung pernyataan bahwa guru yang baik di sekolah berdampak pada kinerja guru lain dan juga berdampak pada prestasi siswa.

Sistem mutu selalu membutuhkan rangkaian umpan balik (Sallis, 2007, p.236). Menurut Marzano, Frontier & Livingstone (2012, pp.61-62) umpan balik penting untuk menentukan keberhasilan praktik yang terfokus, karena umpan balik merupakan ma-

sukan kepada guru apakah usaha mereka benar-benar mengembangkan keahlian pembelajaran. Umpan balik merupakan informasi yang disediakan oleh agen (misalnya guru, teman sebaya, buku, orang tua, sendiri, pengalaman) mengenai aspek dari prestasi seseorang atau pemahaman (Hattie & Timperley, 2007, p.81). Umpan balik yang demikian disebut oleh Antonioni (1996), Bracken et al. (1997), Fleenor & Prince (1997), dan Chappelow (1998) sebagai penilaian kinerja 360 derajat. Penilaian kinerja yang multisource merupakan proses penilaian dengan mengkombinasikan *upward*, *downward*, *lateral* dan *self-assessment*. Umpan balik yang diberikan oleh banyak sumber yang benar-benar mengetahui keseharian seorang guru dapat menjadi umpan balik yang lebih objektif.

Umpan balik sangat bermanfaat bagi guru karena umpan balik yang disampaikan memberikan informasi fokus perubahan yang dikehendaki sekaligus strategi dan perilaku yang seharusnya diperhatikan oleh guru pada pembelajaran berikutnya. Fink (2003, p.14) mengungkapkan bahwa umpan balik yang berkualitas tinggi mempunyai karakteristik berikut: sering, langsung, dapat membedakan, dan disampaikan dengan penuh kasih. Umpan balik yang disampaikan secara langsung, sering, dan disampaikan dengan penuh empati dapat mengubah pendirian dan pendapat guru, khususnya guru yang semula alergi terhadap masukan, menjadi sebuah aktivitas yang ditunggu-tunggu dan diharapkan oleh guru untuk peningkatan kinerjanya.

Umpan balik mulai diberikan ketika aktivitas perencanaan pembelajaran dilakukan (*input*) dilanjutkan ketika proses pembelajaran berlangsung (*process*), dan umpan balik dari kedua aktivitas tersebut dapat menjadi bahan refleksi guru untuk tindak lanjut perbaikan dalam pembelajaran berikutnya. Dengan demikian, evaluasi terhadap pembelajaran tidak hanya dilakukan oleh pengawas sebagai orang luar yang bekerja jauh dari guru, melainkan dilengkapi pula dengan penilaian orang yang berada di lingkungan kerja guru.

Menurut Cahyana (2007) ada dua faktor yang dapat menjelaskan mengapa upaya perbaikan mutu pendidikan selama ini kurang atau tidak berhasil. Pertama, strategi pembangunan pendidikan selama ini lebih bersifat *input oriented*. Strategi yang lebih bersandar kepada asumsi bahwa bilamana semua *input* pendidikan telah dipenuhi, maka secara otomatis lembaga pendidikan (sekolah) akan dapat menghasilkan *output* (keluaran) yang bermutu sebagaimana yang diharapkan. Kedua, pengelolaan pendidikan selama ini lebih bersifat *macrooriented* diatur oleh jajaran birokrasi di tingkat pusat. Akibatnya, banyak faktor yang diproyeksikan di tingkat makro-pusat tidak terjadi atau tidak berjalan sebagaimana seharusnya di tingkat mikro-sekolah. Ochuba (2009) mengemukakan bahwa inspeksi yang wajar dan teratur berpotensi untuk mempromosikan efektivitas sekolah menuju peningkatan mutu sekolah yang diberikan kepada masyarakat. Selain itu, mutu pembelajaran menurun disebabkan oleh tidak efektif dan tidak efisiennya pengawasan sekolah.

Kebijakan pemerintah yang lebih mengacu pada isu-isu yang terjadi di luar kelas, dan tidak mengungkap apa yang sebenarnya terjadi di dalam kelas merupakan salah satu faktor yang menghambat peningkatan mutu pendidikan. Surya (2007) menguraikan bahwa penataan struktur ruang kelas tempat guru bertugas membuat guru bekerja secara individual dan berada di lingkungan kerja yang terisolasi. Direktorat Tenaga Kependidikan telah mempertimbangkan faktor bahwa sampai saat ini pekerjaan guru masih cukup tertutup. Bahkan atasan guru seperti kepala sekolah dan pengawas sekali pun tidak mudah untuk mendapatkan data dan mengamati realitas keseharian *performance* guru di hadapan siswa. Tidak jarang terjadi guru berusaha menampakkan kinerja terbaiknya baik pada aspek perencanaan maupun pelaksanaan pembelajaran hanya pada saat dikunjungi. Selanjutnya ia akan kembali bekerja seperti sedia kala, kadang tanpa persiapan yang matang serta tanpa semangat dan antusiasme yang tinggi (Direktorat Tenaga Kepen-

didikan Dirjen PMPTK, 2008). Oleh karena itu, mulai tahun 2013 pemerintah melakukan evaluasi terhadap kinerja guru yang dilakukan oleh kepala sekolah atau teman sejawat yang telah mendapat pelatihan.

Untuk menilai apakah sekolah mengerjakan tugasnya perlu digali informasi secara internal dan eksternal (Pang, 2006; Fullan, 2007). Menurut *Accomplished California Teachers* (2010, p.15) tujuan utama evaluasi harus selalu peningkatan mutu pembelajaran dan promosi terhadap belajar siswa yang lebih baik. Evaluasi pembelajaran yang dilakukan secara eksternal sering memberikan dampak perbaikan dalam jangka pendek, seperti yang telah diteliti oleh Pang (2006) dan Ochuba (2009). Karena itu, untuk melengkapi evaluasi eksternal yang bersifat *top-down*, perlu dilakukan evaluasi internal yang bersifat *bottom-up*. Reformasi *bottom-up* berfokus pada sekolah masing-masing atau guru secara individu dan dilakukan untuk membuat pembelajaran lebih profesional dengan memberikan berbagai kesempatan untuk pengembangan profesional (Stronge, 2006, p.187). Evaluasi merupakan alat utama dan penting untuk membuat guru menyadari praktiknya, memberikan tantangan kepada para guru untuk memikirkannya, dan mendorong mereka menganalisis dan mengevaluasinya, dan mengimplementasikan perubahan sesuai dengan kebutuhan.

Pengertian *kaizen* adalah peningkatan terus-menerus dan berkesinambungan (*continuous improvement*) atau usaha perbaikan berkelanjutan untuk menjadi lebih baik dari kondisi sekarang secara terus-menerus, termasuk perbaikan diri setiap orang – manajemen paling atas, manajer, dan pekerja (Imai, 2008, p.29; 1998, p.1; 1996, p.4). *Kaizen* menghasilkan pemikiran berorientasi proses, karena proses harus diperbaiki sebelum kita memperbaiki hasil. Selain itu, *kaizen* berorientasi pada manusia dan diarahkan pada upaya yang dilakukan manusia itu sendiri (Imai, 2008, p.61). Keberhasilan the Academy of the Pacific Rim Charter Public School (APR) di Boston dalam menerapkan *kaizen* dibangun bersama oleh

direktur, kepala sekolah, guru, siswa, dan semua komponen yang terlibat di sekolah (Blasdale, 2004). Keberhasilan ini menuntut komunitas yang memiliki komitmen luar biasa sebagai pendidik yang berdedikasi terhadap setiap siswa dan terhadap pekerjaan yang tidak terputus dalam menciptakan, melanjutkan dan memperkokoh sekolah yang sukses. Penelitian yang dilakukan oleh Emilian (2005) tentang penerapan *kaizen* pada Program Master Eksekutif (PME) di Sekolah bisnis di Rensselaer di Hartford (Connecticut), salah satu unit di Rensselaer Polytechnic Institute (Troy, NY) menunjukkan bahwa penerapan *kaizen* berhasil meningkatkan masing-masing matakuliah PME setelah dilakukan modifikasi sesuai dengan keadaan. Perubahan yang terjadi setelah pelaksanaan *kaizen* adalah: (1) menghapus makna ganda di dalam silabus terkait dengan kriteria penilaian (misal, partisipasi kelas) dan tugas; (2) menghapus variasi di dalam silabus seperti format, deskripsi matakuliah, tujuan matakuliah, atau pernyataan “integritas akademik”; (3) menghapus duplikat materi pengajaran seperti studi kasus atau artikel jurnal yang digunakan di dua matakuliah; (4) menjamin mahasiswa mendapat peluang yang cukup untuk memperoleh level yang ditunjukkan dengan perubahan dari ujian akhir saja, atau mid tes dan ujian akhir, ke dalam 4-12 tingkatan tugas dan (5) mengidentifikasi hubungan antara matakuliah untuk menyampaikan lebih banyak program PME yang secara tematik konsisten berfokus pada strategi berpikir dan kepemimpinan.

Adapun tujuan penelitian ini adalah: (1) menghasilkan model evaluasi yang hasilnya dapat digunakan untuk mendeteksi kekeliruan dan memberi peringatan awal, serta dapat digunakan untuk melakukan koreksi sendiri dalam menjamin perbaikan pembelajaran yang kontinyu; (2) menghasilkan sistem informasi dan indikator yang diperlukan untuk mendukung model evaluasi tersebut.

Spesifikasi produk yang dikembangkan berdasarkan kajian dari berbagai kriteria, komponen, dan indikator proses pem-

belajaran adalah model evaluasi, sistem informasi, dan pedoman pelaksanaan evaluasi oleh unsur pendukung pendidikan yang dapat menjamin mutu pembelajaran di sekolah menengah atas.

Model evaluasi yang dikembangkan merupakan hasil modifikasi dari model yang dikemukakan oleh Marzano, Frontier & Livingston (2011) dipadukan dengan *kaizen*. Penentuan ini didasarkan pada beberapa pertimbangan, antara lain: (1) evaluasi pembelajaran yang tercantum di dalam Permenegpan & RB No. 16 Tahun 2009 mencakup perencanaan pembelajaran, pelaksanaan pembelajaran, dan penilaian pembelajaran, sedangkan peran kolega dalam meningkatkan profesionalisme pembelajaran belum tercakup di dalam evaluasi; (2) elemen-elemen yang tercakup di dalam masing-masing ranah pada model yang diusulkan Marzano dkk. dapat menjadi elemen-elemen penjabaran dari indikator-indikator kinerja yang terdapat di dalam instrumen PKG; (3) orientasi evaluasi adalah pada proses pembelajaran yang dilakukan oleh guru, sejalan dengan *kaizen* yang berorientasi pada manusia dan diarahkan pada upaya yang dilakukan manusia itu sendiri; (4) dengan bantuan elemen-elemen yang diadaptasi dari model Marzano dkk, model evaluasi ini dapat membantu guru dan sekolah mengenali masalah pembelajaran secara rinci; dan (5) menentukan rencana perbaikan berdasarkan elemen-elemen yang telah dikenali.

Untuk mengevaluasi pembelajaran yang bermutu, ada empat ranah yang menjadi penyebab bagi ranah lainnya, yaitu ranah: (1) strategi dan perilaku kelas, (2) perencanaan dan persiapan, (3) refleksi pengajaran, dan (4) kolegialitas dan profesionalisme guru (Marzano, 2011, p.29). Tekanan utama di dalam Model Marzano dkk adalah apa yang terjadi di dalam kelas, yakni penggunaan strategi dan perilaku guru untuk meningkatkan prestasi siswa. Tekanan ini membedakan model ini dari model evaluasi guru lainnya. Ranah strategi dan perilaku di kelas berhubungan langsung dengan apa yang dilakukan guru di dalam

kelas. Limbach, Duron & Waugh (2008) mengemukakan bahwa belajar aktif dapat membuat pelajaran lebih dapat dinikmati baik bagi guru maupun siswa, dan yang paling penting, dapat menyebabkan siswa berpikir kritis.

Menurut Liker & Hoseus (2008: p.xxix) tanpa alat-alat pada tingkat proses, masalah tidak akan terlihat, menjadikan orang-orang enggan untuk mengembangkan kemampuan mereka dalam berpikir dan menemukan masalah. Karena itu, untuk mendukung evaluasi dikembangkan instrumen evaluasi diri, teman sejawat dan siswa yang memuat elemen-elemen pembelajaran secara rinci. Evaluasi diri guru adalah proses ketika guru membuat keputusan tentang kecukupan dan keefektifan pengetahuan, kinerja, kepercayaan, dan pengaruh bagi tujuan perbaikan diri mereka sendiri (Freddano & Siri, 2012, p.1143). Di dalam evaluasi diri, guru sendiri yang mengidentifikasi, menginterpretasi, dan memutuskan informasi terhadap praktik pembelajaran yang telah dilakukan. Hasil evaluasi siswa bisa digunakan secara formatif untuk memperbaiki metode mengajar guru (Al-Abbadi et al., 2009, p.181). Marlow (2009) menghadirkan dan mempresentasikan kolega yang profesional dan berpengalaman bersama-sama dalam sebuah konferensi sebagai cara untuk meningkatkan identitas profesional mereka.

Penelitian yang dilakukan oleh Douglas & Douglas (2006) tentang evaluasi mutu pembelajaran di Riverbank Business School United Kingdom membandingkan hasil evaluasi pembelajaran yang diberikan oleh mahasiswa, teman sejawat dan mahasiswa anonim. Review teman sejawat digunakan untuk memantau mutu perbaikan pembelajaran. Hasil review teman sejawat menunjukkan bahwa observasi menyebabkan dosen melakukan pembelajaran yang berbeda, tidak natural, dibuat-buat, kelas tidak berjalan sebaik biasanya, ada juga dosen yang tidak suka dilihat dan cenderung berfokus pada kritik positif, sementara dosen baru merasa senang diamati karena mereka merasa sedang dalam taraf belajar dan mengembangkan gaya pembelajarannya. Pe-

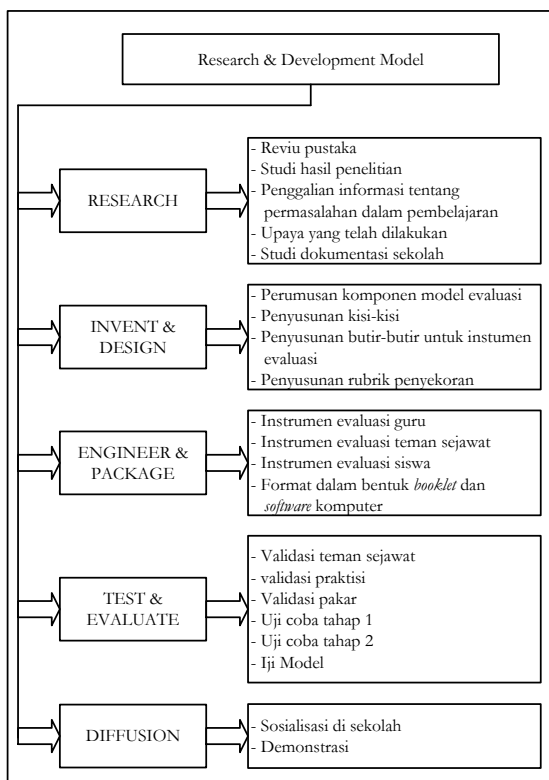
nelitian yang dilakukan oleh Helterbran (2008) bertujuan untuk mengetahui persepsi mahasiswa tentang pembelajaran yang efektif yang dilakukan oleh 283 dosen pada tiga program studi di tiga universitas di Pennsylvania, dengan data yang diperoleh melalui program online. Hasil penelitian menunjukkan bahwa mahasiswa mengkritik dosen yang basis pengetahuannya kurang atau mereka tidak mengikuti perkembangan terbaru di bidang pendidikan. Mahasiswa berharap dosen memiliki pengetahuan yang mendalam dan kemampuan untuk menyampaikannya dengan cara yang bermakna, dan tidak hanya mengandalkan pada materi yang disajikannya dalam PowerPoint, buku, catatan kuliah, atau paket materi lainnya. Kepribadian seorang dosen yang diharapkan oleh mahasiswa adalah yang menunjukkan semangat, antusiasme terhadap pembelajaran dan bidang pendidikan, memiliki "*sense of humor*", dan humanis. Kualitas profesional/pembelajaran yang disukai oleh mahasiswa adalah yang mampu mengorganisasikan dan mampu mengajar. Wolbring (2012) melakukan penelitian untuk melihat apakah evaluasi mahasiswa terhadap pembelajaran bias karena ketidakhadiran, dan apakah prosedur yang menyelaraskan ketidakhadiran mengubah ranking perkuliahan. Penelitian Wolbring meninjau masalah dari perspektif *missing data* dan hasil empiris saat ini melalui model regresi untuk menentukan faktor mana yang secara simultan berhubungan dengan kehadiran mahasiswa di kelas dan peringkat perkuliahan. Hasilnya menunjukkan bahwa tidak ada peringkat perkuliahan dan perankingan yang bias. Evaluasi mahasiswa terhadap pembelajaran lebih layak untuk mengidentifikasi tinggi rendahnya mutu perkuliahan daripada menentukan ranking perkuliahan.

Hasil penelitian yang dilakukan oleh Olive, Mathers & Laine (2009) menjelaskan bahwa kepala sekolah mengetahui bahwa guru yang efektif adalah kontributor terbesar terhadap peningkatan *outcome* siswa. Yang tidak selalu jelas adalah bagaimana kepala sekolah membantu guru menyesuaikan praktik pengajarannya dengan diversifi-

kasi gaya belajar yang semakin meningkat yang mereka temukan di kelas.

Metode Penelitian

Penelitian ini merupakan pengembangan model menggunakan *Research, Development and Diffusion* (R-D-D) dari Havelock (1969) dengan tahapan seperti pada diagram berikut ini.



Gambar 1. Model Pengembangan Instrumen Evaluasi

Pengumpulan data dilakukan dalam dua tahap, yaitu tahap penelitian dan tahap pengembangan. Pada tahap penelitian, data dikumpulkan melalui (a) wawancara, (b) observasi, (c) angket, dan (d) studi dokumentasi, sedangkan data pada tahap pengembangan diperoleh melalui (e) teknik Delphi, (f) *focus group discussion* (FGD), dan (g) Uji coba 1 dan 2.

Penelitian dan pengembangan produk dilakukan mulai Februari 2012 sampai Desember 2012 di SMAN 1, SMAN 3, SMAN 4, SMAN 6, SMAN 7, SMAN 8, dan SMA Laboratorium UM di Kota Malang.

Tahap Penelitian

Pada tahap penelitian (*research stage*), peneliti melakukan reviu pustaka, yakni mengkaji berbagai teori yang terkait dengan proses pembelajaran, jaminan mutu, pelaksanaan mutu menggunakan *kaizen*, sehingga dapat ditentukan komponen-komponen mutu proses pembelajaran dan indikatornya. Selanjutnya, peneliti melakukan observasi ke tujuh SMA di Malang, serta wawancara dengan para kepala sekolah, wakil kepala sekolah bidang kurikulum (Wakakur), unit penjamin mutu atau tim pengembang akademik, guru bidang studi dan bimbingan konseling, dan seorang pengawas sekolah, dengan jumlah total responden sebanyak 35 orang.

Wawancara dan observasi dilakukan untuk mengamati secara langsung berbagai upaya sekolah dalam memperbaiki dan meningkatkan mutu proses pembelajaran, mengidentifikasi kelebihan, kelemahan, permasalahan, dan usaha yang telah dilakukan dalam pelaksanaan penjaminan mutu. Di sekolah yang telah melaksanakan *lesson study* (LS) peneliti memanfaatkan kesempatan untuk aktif mengikuti kegiatan LS diawali dengan aktivitas persiapan (*pre conference*), mengamati pembelajaran guru di kelas, dan refleksi. Di samping itu, peneliti mengamati aktivitas sivitas akademika sehari-hari, dari pagi sebelum jam pelajaran dimulai, sampai jam pelajaran berakhir.

Tahap Pengembangan

Penciptaan dan Desain (Invent & Design)

Tahap pengembangan model diawali dengan merumuskan komponen-komponen model evaluasi untuk mengungkapkan masalah yang dialami di dalam pembelajaran. Selanjutnya, dari komponen-komponen tersebut disusun kisi-kisi, butir-butir, dan rubrik penyekoran untuk instrumen evaluasi yang meliputi instrumen evaluasi diri guru, evaluasi oleh teman sejawat, dan evaluasi oleh siswa.

Rekayasa dan Pengemasan (Engineer dan Package)

Instrumen yang telah disusun pada tahap sebelumnya dan direvisi pada tahap pengujian dan evaluasi, selanjutnya dikemas dalam bentuk *booklet* dan atau *software* berbasis komputer sehingga sekolah dapat dengan mudah menggunakannya dan responden mudah mengisinya. Informasi yang diberikan oleh responden guru dan siswa melalui pengisian instrumen evaluasi berbasis komputer bisa langsung dibaca oleh kepala sekolah atau yang berwenang sebagai bahan supervisi, umpan balik, dan perbaikan atau peningkatan pembelajaran selanjutnya.

Pengujian dan Evaluasi (Test dan Evaluate)

Tahap pengujian dan evaluasi instrumen bertujuan untuk: (1) menilai kesesuaian indikator dengan butir, (2) menentukan butir yang memiliki kualitas yang baik secara teoritis dan empiris, (3) mengetahui keandalan instrumen, dan (4) mengumpulkan data kelebihan dan kelemahan sekolah dalam pembelajaran.

Instrumen yang dikembangkan dalam penelitian ini meliputi instrumen evaluasi diri guru, evaluasi teman sejawat, dan evaluasi siswa. Instrumen yang telah divalidasi oleh pakar dan praktisi selanjutnya diujicobakan sebanyak dua kali.

Subjek uji coba dalam penelitian ini seluruhnya berjumlah 937 orang, meliputi: (1) 6 orang guru, 6 orang teman sejawat, dan 90 orang siswa pada uji coba tahap 1; dan (2) 53 orang guru, 105 teman sejawat, dan 677 siswa pada uji coba tahap 2.

Teknik Analisis Data

Untuk melihat validitas isi berdasarkan pertimbangan panelis digunakan analisis *content validity ratio* (CVR) dari Lawshe (1975) dengan rumus berikut.

$$CVR = \frac{n_e - n/2}{n/2}$$

Keterangan:

n_e = jumlah panelis yang berpendapat butir relevan

n = jumlah total panelis

Analisis reliabilitas instrumen dilakukan menggunakan formula reliabilitas komposit (CR), sedangkan reliabilitas antar-rater menggunakan korelasi Spearman. Analisis model persamaan struktural digunakan *software* program SmartPLS dari Hansmann & Ringle, sedangkan untuk uji kecocokan model digunakan *software* program *generalized structured component analysis* (GesCA) yang dikembangkan oleh Heungsun Hwang (2011). Model SEM SmartPLS dan uji kecocokan model GesCA digunakan untuk mengatasi jumlah sampel minimal yang dipersyaratkan oleh SEM parametrik yakni antara 100 – 150 subjek (Anderson & Gerbing, 1988, dalam Schumacker & Lomax, 2004, p.49), atau setiap satu kasus diwakili oleh minimal lima sampai sepuluh sampel (Bentler & Chou, 1987, dalam Schumacker & Lomax, 2004, p.50).

Teknik analisis data kualitatif digunakan untuk mengolah data hasil observasi dan wawancara menggunakan metode interaktif model Miles dan Huberman (1994, p.23), yakni diawali dengan kegiatan menuliskan hasil observasi dan wawancara, mengedit, mengklasifikasi, mereduksi dan menyajikan.

Hasil Penelitian dan Pembahasan

Pengembangan model evaluasi berbasis *kaizen* terhadap penjaminan mutu pembelajaran dilakukan melalui proses yang panjang. Diawali dengan proses kajian teoretik, dilanjutkan dengan pengumpulan data dan pengamatan di sekolah tentang nilai-nilai *kaizen* yang telah dilaksanakan di sekolah, permasalahan-permasalahan yang muncul dalam proses pembelajaran, dan upaya-upaya yang telah dilakukan, kemudian akhirnya disusun model evaluasi beserta instrumen pendukungnya untuk membantu mengatasi permasalahan yang ada di lapangan. Dengan memadukan model teoritik dan temuan empirik, akhirnya dihasilkan model evaluasi pembelajaran dan sistem infomasi-nya yang didukung oleh tiga buah instrumen evaluasi, yakni evaluasi diri guru, evaluasi teman sejawat dan siswa.

Instrumen yang dikembangkan melalui tahap FGD dan teknik Delphi divalidasi secara statistik dengan analisis CVR dan CVI. Hasil analisis CVR menunjukkan bahwa dari 168 butir pada instrumen evaluasi diri guru, ada dua butir yang kualitasnya tidak baik, dan 3 butir berada pada ambang nilai minimum, yaitu sebesar 0,538. Butir pada instrumen teman sejawat menunjukkan satu butir yang kualitasnya tidak baik, dan pada instrumen siswa ada 8 butir yang kualitasnya tidak baik.

Dari hasil analisis CVR selanjutnya dilakukan analisis validitas instrumen dengan menggunakan *content validity index* (CVI). Hasilnya menunjukkan bahwa ketiga instrumen valid. Hasil CVI disajikan pada tabel berikut.

Tabel 1. *Content Validity Index Instrumen Evaluasi Berbasis Kaizen*

<i>Content Validity Index</i> (CVI) Instrumen Evaluasi		
Guru	Teman Sejawat	Siswa
0,928	0,952	0,864

Instrumen pada uji coba kedua dibuat dalam dua format, yakni format *paper-based instrument* dan *computer-based instrument*. Catatan waktu yang diperlukan untuk pengisian instrumen evaluasi diri guru berbasis komputer berkisar antara 15 menit sampai paling lama 30 menit; instrumen evaluasi teman sejawat antara 10 menit sampai 20 menit, dan instrumen evaluasi siswa antara 15 menit sampai 30 menit.

Analisis reliabilitas ketiga instrumen evaluasi dilakukan dengan menggunakan *software* program statistik SmartPLS. Hasilnya disajikan pada Tabel 1.

Reliabilitas antar-rater digunakan untuk menilai derajat yang dilakukan oleh rater yang berbeda dalam membuat estimasi yang konsisten terhadap fenomena yang sama. Menurut Multon (2012) koefisien korelasi Product Moment dari Pearson adalah

statistik yang paling banyak digunakan untuk menghitung derajat konsistensi antar-rater yang independen. Tingkat reliabilitas yang dapat diterima untuk korelasi Pearson adalah 0,70. Korelasi Pearson dapat dilakukan apabila asumsi yang mendasarinya terpenuhi, yaitu berdistribusi normal. Jika data tidak berdistribusi normal, Multon (2012) mengusulkan untuk menggunakan koefisien Spearman.

Tabel 2. Reliabilitas Ketiga Instrumen Evaluasi

Ranah	Reliabilitas		
	Guru	Teman Sejawat	Siswa
Strategi dan perilaku di Kelas	0,846	0,999	0,913
Perencanaan dan persiapan pembelajaran	0,91	0,930	-
Refleksi pembelajaran	0,93	0,934	-
Kolegialitas dan profesionalisme	0,878	0,866	-
Instrumen	0,911	0,956	0,913

Hasil analisis reliabilitas antar-rater terhadap instrumen evaluasi diri guru, evaluasi teman sejawat, dan evaluasi siswa menunjukkan bahwa hasil evaluasi antara guru dan siswa memiliki konsistensi yang tinggi dan signifikan. Kondisi ini ditunjukkan oleh koefisien korelasi terhadap semua subjek yang berada di atas 0,760, sedangkan korelasi antara guru dengan teman sejawat hanya mencapai 26,92% yang memiliki koefisien korelasi yang secara statistik signifikan, namun korelasi antara hasil evaluasi diri guru dan teman sejawat tergolong rendah, bahkan negatif. Adapun korelasi antara evaluasi teman sejawat dan siswa menun-

jukkan moderat dengan koefisien korelasi di atas 0,40. Kondisi seperti ini dapat disebabkan oleh kondisi antara guru dan teman sejawat yang berbeda dalam mengenali perilaku dan strategi guru di kelas. Siswa sebagai subjek didik yang setiap hari mengikuti pembelajaran bersama guru dapat memberikan evaluasi lebih riil daripada teman sejawat yang bisa jadi hanya satu atau dua kali mengamati guru di kelas.

Jika ditinjau dari korelasi antar-rater pada masing-masing sekolah, diperoleh gambaran bahwa korelasi guru dan teman sejawat pada strategi dan perilaku pembelajaran di SMAN 1 dan SMAN 4 tinggi, yakni masing-masing sebesar 0,928 dan 0,975, sedangkan di SMAN 6 dan SMAN 8 menunjukkan moderat, yakni sebesar 0,50. Sementara hasil evaluasi terhadap guru-guru di SMAN 3 dan di SMA Laboratorium UM menunjukkan negatif. Ada beberapa faktor yang menjadi penyebab kondisi seperti ini. Sebagai contoh di SMAN 1 setiap tahun sekolah mengundang 5 sampai 6 pengawas untuk melakukan supervisi kelas untuk memenuhi persyaratan ISO. Di samping itu, program *co-teacher* untuk guru bahasa Jerman tampaknya memberi dampak positif terhadap guru-guru mapel lain dalam meningkatkan proses pembelajarannya di kelas. Di SMAN 4 kondisi tersebut tampaknya merupakan dampak positif dari pelaksanaan *Lesson Study* (LS) berbasis mata pelajaran, sehingga guru-guru dalam MGMP berusaha untuk selalu memperbaiki proses pembelajarannya di kelas. Di samping itu, kepala SMAN 4 ikut langsung mengamati proses pembelajaran di kelas, memimpin proses refleksi kegiatan LS, dan memantau perkembangan perbaikan yang dilakukan guru pada *open class* kedua.

Hasil korelasi evaluasi diri dan teman sejawat untuk perencanaan dan persiapan pembelajaran di SMAN 4 menunjukkan tinggi, yakni sebesar 0,811, sedangkan di SMAN 1, SMAN 3, SMAN 6, SMAN 8, dan SMA Laboratorium UM menunjukkan rendah bahkan negatif. Faktor penyebab kondisi ini tampaknya karena fungsi MGMP dalam membuat perencanaan dan persiapan

pembelajaran masih belum sesuai dengan yang diharapkan, karena perencanaan dan persiapan pembelajaran dilaksanakan satu tahun sekali dalam bentuk *workshop* persiapan pembelajaran, sedangkan persiapan yang dilakukan bersama-sama secara rutin setiap akan mengajar atau setiap akan menyampaikan satu kompetensi dasar masih belum dilakukan.

Pelaksanaan refleksi pembelajaran yang tampaknya sudah berjalan dengan baik dilakukan oleh guru-guru di SMAN 8, ditunjukkan oleh koefisien korelasi sebesar 0,709, sedangkan di sekolah lain masih rendah. Tampaknya Program LS yang dilaksanakan di SMAN 8 memberi dampak yang positif terhadap kegiatan refleksi pembelajaran, karena kepala sekolah memimpin langsung aktivitas refleksi disertai dengan pakar dari Universitas Negeri Malang sebagai pendamping kegiatan LS. Namun karena kesibukan kepala sekolah yang sangat tinggi, kepala sekolah tidak selalu dapat mengikuti kegiatan observasi kelas.

Di antara ketiga sekolah yang sudah melaksanakan program LS, SMA Laboratorium UM menunjukkan kelebihan dalam kolegialitas dan profesionalisme. Hal ini tampak dari koefisien korelasi yang tinggi untuk indikator-indikator yang mengukur kolegialitas dan profesionalisme antara hasil evaluasi diri dan teman sejawat.

Dengan demikian, dampak positif dari pelaksanaan LS di tiga sekolah berbeda. Instrumen evaluasi berbasis *kaizen* dapat mengidentifikasi kelebihan masing-masing sekolah ditinjau dari indikator-indikator yang mengukur strategi dan perilaku di kelas, perencanaan pembelajaran, refleksi, dan kolegialitas dan profesionalisme.

Untuk mengetahui reliabilitas instrumen evaluasi secara keseluruhan, dilakukan agregasi terhadap ketiga instrumen tersebut, yakni dengan melihat nilai median butir yang mengukur indikator yang sama pada ketiga instrumen evaluasi. Melalui *principal component analysis* (PCA) diketahui bahwa konstruk mutu pembelajaran pada instrumen evaluasi berbasis *kaizen* memiliki dimensi lebih dari satu (multidimensi). Dari

hasil uji tersebut tampak bahwa indikator-indikator mutu pembelajaran yang semula terdiri dari empat ranah mengelompok menjadi tiga dimensi. Dimensi pertama meliputi indikator nomor: (11) perencanaan dan persiapan untuk kompetensi dasar dan materi, (12) perencanaan dan persiapan untuk pemanfaatan bahan dan alat, (13) perencanaan dan persiapan untuk kebutuhan spesifik siswa, (14) evaluasi kinerja sendiri, (15) pengembangan dan pelaksanaan rencana pengembangan profesi, (16) pembentukan lingkungan yang positif, (17) pertukaran ide mengajar dan strateginya, dan (18) promosi pengembangan sekolah. Dimensi kedua memuat indikator nomor: (1) tujuan belajar dan apresiasi, (2) aturan dan prosedur, (3) interaksi dengan pengetahuan baru, (4) praktik dan pendalaman pengetahuan, dan (5) penerapan pengetahuan (praktik dan pengujian hipotesis), sedangkan dimensi ketiga terdiri dari indikator nomor: (6) keterlibatan siswa, (7) kepatuhan terhadap aturan dan prosedur, (8) hubungan yang efektif dengan siswa, (9) sikap respek, dan (10) kepuasan belajar.

Dimensi pertama terbentuk dari ranah perencanaan dan persiapan (indikator nomor 11 s.d 13), refleksi pembelajaran (indikator nomor 14 dan 15), dan kolegialitas dan profesionalisme (indikator nomor 16 s.d 18), sehingga mendapat nama baru sebagai dimensi profesionalisme perencanaan. Dimensi kedua dan ketiga terbentuk dari ranah strategi dan perilaku di kelas. Dimensi kedua memuat elemen-elemen yang terkait dengan strategi pembelajaran, sehingga dimensi kedua diberi nama dimensi strategi pembelajaran (indikator nomor 1 s.d 5), sedangkan dimensi ketiga memuat elemen-elemen yang terkait dengan perilaku yang menyertai strategi pembelajaran (indikator nomor 6 s.d 10), sehingga dimensi ketiga diberi nama dimensi perilaku pembelajaran.

Perubahan konstruk seperti ini dapat disebabkan oleh fakta di lapangan bahwa guru melakukan perencanaan pembelajaran bersama dengan teman sejawat dalam MGMP hanya ketika sekolah melaksanakan *workshop* persiapan tahun pelajaran baru

yang dilakukan pada awal tahun ajaran baru, sedangkan persiapan pembelajaran untuk setiap pertemuan dilakukan sendiri oleh guru. Di samping itu, pada sekolah-sekolah yang masih belum melaksanakan program LS refleksi terhadap proses pembelajaran dan perencanaan jarang dilakukan bersama dengan teman sejawat. Diskusi tentang permasalahan yang terjadi selama proses pembelajaran berlangsung nonformal dan bersifat insidental ketika guru berada di ruang guru. Kondisi seperti ini menyebabkan elemen-elemen yang secara teoritis berada pada ranah perencanaan dan persiapan pembelajaran, refleksi pembelajaran, dan kolegialitas dan profesionalisme bergabung menjadi satu konstruk. Hal ini sebenarnya sudah disebutkan di dalam teori Marzano dkk. bahwa hubungan antarranah dalam model evaluasi bersifat kausalitas, dan pada kondisi nyata di lapangan tidak mungkin dilakukan pemisahan secara nyata.

Analisis reliabilitas instrumen dilakukan menggunakan formula reliabilitas komposit (CR). Tabel berikut menyajikan koefisien reliabilitas tiap dimensi.

Tabel 3. Statistik Deskriptif Hasil Pengukuran Konsep Mutu Pembelajaran

Dimensi	Reliabilitas
Strategi Pembelajaran	0,907
Perilaku Pembelajaran	0,911
Profesionalisme Perencanaan	0,906

Hasilnya diperoleh koefisien reliabilitas komposit untuk instrumen evaluasi berbasis *kaizen* sebesar 0,912. Hal ini menunjukkan bahwa konstruk instrumen evaluasi berbasis *kaizen* memiliki konsistensi internal yang tinggi.

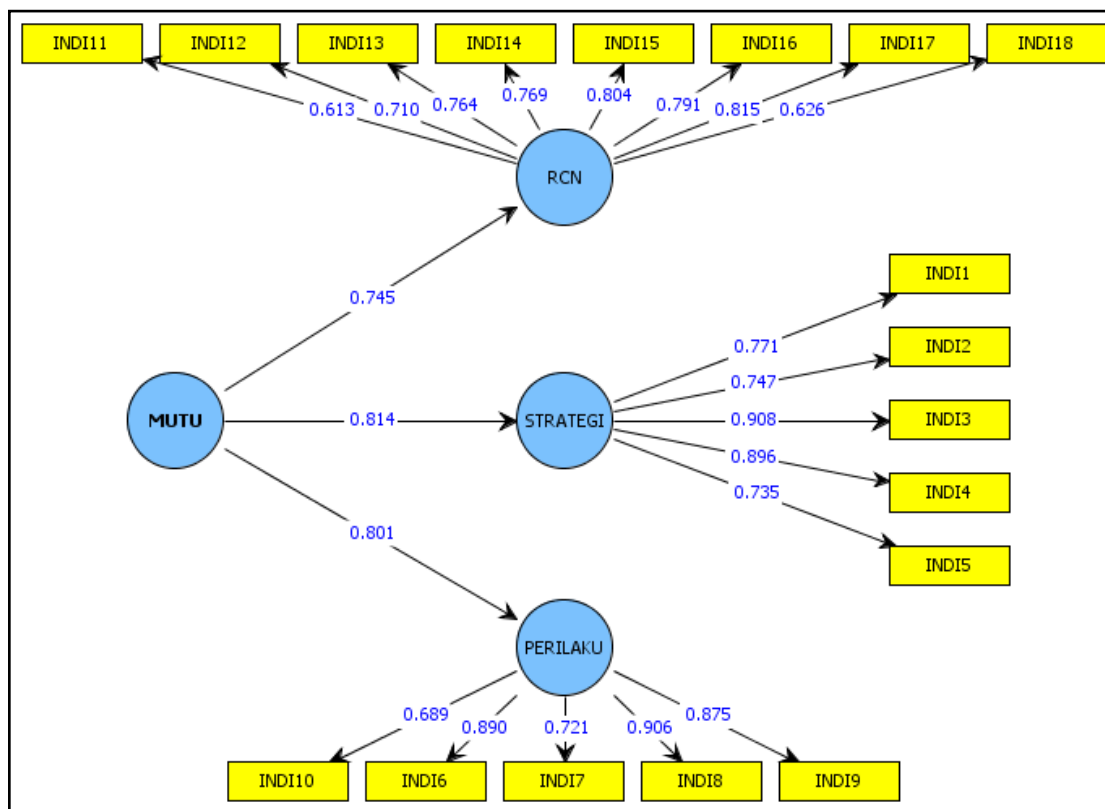
Hasil uji kelayakan model evaluasi berbasis *kaizen* disajikan Tabel 4.

Tabel 4. Uji Kecocokan Model Evaluasi Berbasis *Kaizen* terhadap Penjaminan Mutu di Sekolah Menengah Atas

Model Fit	
FTT	0,607
AFIT	0,572
GFI	0,997
SRMR	0,076
NPAR	75

Pada tabel tersebut tampak bahwa nilai GFI sebesar 0,997, dan SRMR sebesar 0,076, maka dapat disimpulkan bahwa model yang dikembangkan secara teoritik didukung oleh data empirik.

Hasil analisis menunjukkan bahwa model teoretik mengalami perubahan pada model empirik, yakni indikator-indikator yang tercakup pada masing-masing ranah mengalami perubahan. Gambar berikut menunjukkan model empirik beserta koefisien korelasi masing-masing indikator terhadap variabel laten.



Gambar 2. Model Empirik Instrumen Evaluasi Berbasis *Kaizen*

Konstruk mutu pembelajaran dibentuk oleh konstruk profesionalisme perencanaan (Rcn), strategi pembelajaran (strategi), dan perilaku pembelajaran (perilaku). Secara teoritik konstruk profesionalisme perencanaan meliputi indikator-indikator yang tercakup dalam ranah perencanaan dan persiapan pembelajaran, ranah refleksi pem-

belajaran, dan ranah kolegialitas dan profesionalisme, sedangkan konstruk strategi pembelajaran dan perilaku pembelajaran secara teoritik mencakup indikator-indikator yang terdapat pada ranah strategi dan perilaku di kelas. Hasil bentukan dimensi baru disajikan pada Tabel 5.

Tabel 5. Dimensi Baru Berdasarkan Analisis Faktor Eksploratori dan Uji Kecocokan Model

Dimensi Baru	Indikator
Strategi Pembelajaran	<ul style="list-style-type: none"> ▪ Tujuan belajar dan apresiasi ▪ Aturan dan prosedur ▪ Interaksi dengan pengetahuan baru ▪ Praktik dan pendalaman pengetahuan ▪ Penerapan pengetahuan (praktik dan pengujian hipotesis)
Perilaku Pembelajaran	<ul style="list-style-type: none"> ▪ Keterlibatan siswa ▪ Kepatuhan terhadap aturan dan prosedur ▪ Hubungan yang efektif dengan siswa ▪ Sikap respek ▪ Kepuasan belajar
Profesionalisme Perencanaan	<ul style="list-style-type: none"> ▪ Kompetensi dasar dan Materi ▪ Pemanfaatan Bahan dan Alat ▪ Kebutuhan spesifik Siswa ▪ Evaluasi kinerja sendiri ▪ Pengembangan dan pelaksanaan rencana pengembangan profesi ▪ Pembentukan Lingkungan Yang Positif ▪ Pertukaran ide mengajar dan strateginya ▪ Promosi Pengembangan Sekolah

Sistem informasi evaluasi berbasis *kaizen* (Evakai) disajikan melalui program komputer. Program Evakai memuat instrumen evaluasi diri evaluasi teman sejawat, dan evaluasi siswa. Instrumen berbasis komputer dibuat dalam bentuk web menggunakan *Hypertext Preprocessor* (PHP), dan

Structured Query Language (SQL). Adapun *software* yang digunakan untuk mendukung sistem operasi program adalah Xampp. Fungsi Xampp itu sendiri adalah sebagai *server* yang bisa berdiri sendiri (*localhost*).

Program Evakai menampilkan hasil evaluasi dari ketiga responden secara langsung, baik dalam bentuk *chart* maupun deskripsi tentang kelebihan dan kelemahan guru menurut hasil evaluasi diri guru, siswa, dan teman sejawat. Di samping itu, hasil evaluasi sekaligus menampilkan rencana perbaikan yang akan dilakukan oleh guru, dan saran perbaikan pembelajaran dari siswa dan teman sejawat guru. Dengan demikian, baik guru maupun kepala sekolah dapat melihat secara rinci, elemen-elemen mana sajakah yang masih menjadi kelemahan guru dan sekolah, sehingga informasi hasil evaluasi tersebut dapat menjadi dasar untuk merencanakan perbaikan selanjutnya.

Pemantauan perbaikan dapat dilakukan melalui dua cara, pertama (1) kepala sekolah atau teman sejawat yang memiliki kompetensi unggul dapat melakukan pemantauan terhadap upaya perbaikan yang dilakukan oleh guru, dan (2) hasil upaya perbaikan dapat dilihat dari hasil evaluasi kedua pada akhir semester, baik melalui evaluasi diri guru, teman sejawat, maupun dari siswa.

Dari elemen-elemen kelemahan yang disajikan melalui hasil evaluasi, guru bersama tim (teman sejawat) melakukan pengelompokan prioritas perbaikan ke dalam empat kuadran: *easy to do-high impact*, *easy to do-low impact*, *hard to do-high impact*, dan kuadran *hard to do-low impact*. Dalam melaksanakan perbaikan yang segera dilakukan, guru memilih permasalahan yang termasuk ke dalam kuadran *easy to do-high impact*, sedangkan permasalahan yang termasuk ke dalam *hard to do-high impact* dapat disampaikan kepada kepala sekolah untuk dipecahkan bersama-sama. Adapun permasalahan yang termasuk ke dalam *kuadran hard to do-low impact* diabaikan. Perbaikan yang dilakukan oleh guru dapat dipantau melalui grafik hasil evaluasi pertama dan hasil evaluasi berikutnya. Hasil evaluasi yang dimiliki guru dapat

menjadi bahan pijakan bagi guru untuk selalu menyadari kelebihan, kekurangan, dan elemen-elemen yang akan diperbaiki.

Program evaluasi berbasis *kaizen* dapat digunakan minimal dua kali dalam satu semester, yaitu, pada tengah semester, dan akhir semester. Informasi yang diperoleh dari hasil evaluasi pada tengah semester dapat langsung digunakan oleh guru untuk segera mengetahui letak kelebihan dan kekurangannya dalam pembelajaran, sehingga sebelum semester berakhir guru sudah melakukan perbaikan.

Simpulan

Hasil penelitian dan pembahasan terhadap model evaluasi berbasis *kaizen* yang dikembangkan dapat disimpulkan sebagai berikut.

1. Hasil empirik menunjukkan bahwa model evaluasi berbasis *kaizen* dapat digunakan untuk mendeteksi kekeliruan dan memberi peringatan awal, serta dapat digunakan untuk melakukan koreksi sendiri. Model ini mencakup tiga konstruk, yakni: a) strategi pembelajaran; b) perilaku pembelajaran; dan c) profesionalisme perencanaan. Komponen dan indikator model evaluasi didasarkan pada karakteristik *kaizen* yakni mengenali masalah yang muncul dalam proses pembelajaran.
2. Sistem informasi yang digunakan untuk mendukung model evaluasi adalah; a) Tiga buah instrumen berbasis komputer, yakni instrumen evaluasi diri guru, instrumen evaluasi oleh siswa, dan instrumen evaluasi oleh teman sejawat; b) melalui instrumen berbasis komputer hasil evaluasi dapat langsung diketahui oleh guru atau kepala sekolah, baik elemen-elemen kelebihan maupun kekurangan. Di samping itu, masing-masing penilai memiliki kesempatan untuk menyampaikan saran perbaikan bagi diri sendiri, teman sejawat, maupun untuk sekolah. Dengan demikian, kepala sekolah memiliki informasi lengkap sebagai landasan untuk melakukan

supervisi, dan menindaklanjuti saran perbaikan; c) sebagai panduan untuk pengguna, *software* evaluasi berbasis *kaizen* dilengkapi dengan buku manual evaluasi.

Saran Pemanfaatan dan Diseminasi

1. Model evaluasi berbasis *kaizen* yang telah dikembangkan dapat dipergunakan oleh sekolah untuk mengevaluasi pembelajaran dalam rangka memperbaiki dan meningkatkan mutu pembelajaran. Di samping itu, model ini dapat digunakan oleh sekolah sebagai dasar untuk melengkapi informasi p-enilaian kinerja guru. samping itu, model ini dapat digunakan oleh sekolah sebagai dasar untuk melengkapi informasi penilaian kinerja guru.
2. Model evaluasi berbasis *kaizen* telah diujicobakan pada tujuh sekolah menengah atas di Malang, namun belum semua guru dan siswa melakukan evaluasi, sehingga proses diseminasi dapat dilakukan pada semua sekolah untuk semua guru melalui forum Musyawarah Kerja Kepala Sekolah (MKKS).

Keterbatasan Penelitian

Uji kelayakan evaluasi berbasis *kaizen* melibatkan jumlah subjek yang dievaluasi terbatas. Oleh karena itu, hasilnya hanya bermakna bahwa: (1) informasi kelayakan model diperoleh dari hasil agregasi butir berupa indikator, dan (2) tidak semua teman sejawat yang mengevaluasi guru pernah mengamati guru melakukan proses pembelajaran di dalam kelas, sehingga analisis butir pada ranah strategi dan perilaku di kelas diwakili oleh responden dalam jumlah yang lebih sedikit.

Daftar Pustaka

- Accomplished California Teachers. (2010). *A quality teacher in every classroom: Creating a teacher evaluation system that works for California*. Stanford, CA.: National Board Resource Center, Stanford University.

- Al-Abbadi, Ibrahim., Alkhateeb, Fadi., Khanfar, Nile, et al. (2009). Pharmacy students' perceptions of the teaching evaluation process in Jordan. [Versi Elektronik]. *Education, Business and Society: Contemporary Middle Eastern Issues*, 2 (3), 181-190.
- Antonioni, D. (1996). Designing an effective 360-degree appraisal feedback process. *Organizational Dynamics*, 25, 24-31.
- Berliner, D.C. (2005). The near impossibility of testing for teacher quality [Versi Elektronik]. *Journal of Teacher Education*, 56 (3), 205-213.
- Blasdale, S. (2004). From theory to practice: Kaizen and the Academy of the Pacific Rim. *Journal of Education*, 185 (2), 1-9.
- Bracken, D.W., Dalton, M.A., Jako, R.A. et al. (1997). *Should 360-degree feedback be used only for developmental purposes?* Greensboro, North Carolina: Center for Creative Leadership
- Business Council of Australia. (2008). *Review of the National Innovation System*.
- Cahyana, Ade. (2007). *Upaya peningkatan mutu sekolah melalui otonomi satuan pendidikan*. Retrieved at July 7 2010 from ade.psp.depdiknas.go.id/.../Revised3_peningkatan_mutu_pendidikan_mellotonomisasi_220220101.doc
- Chappelow, C. T. (1998). *360-degree feedback*. In McCauley, C. D., Moxley, R., & Van Velsor, E. (Eds.). *The center for creative leadership handbook of leadership development* (pp. 29-65). San Francisco: Jossey-Bass.
- Deepwell, F. (2007). Embedding quality in e-learning implementation through evaluation [Versi elektronik]. *Educational Technology & Society*, 10 (2), 34-43.
- Dinham, S., Ingvarson, L., & Kleinhenz, E. (2008). *How we can raise the quality of school education so that every student benefits? Teaching talent the best teacher for Australia's classroom*. Australia: The Business Council of Australia.
- Direktorat Tenaga Kependidikan. (2008). *Penilaian kinerja guru*. Jakarta: Departemen Pendidikan Nasional.
- Douglas, J. & Douglas, A. (2006). Evaluating teaching quality. *Quality in Higher Education*, Vol. 12 (1), 3-13.
- Emiliani, M.L. (2005). Using *kaizen* to improve graduate business school degree programs [Versi elektronik]. *Quality Assurance in Education*, 13 (1), 37-52.
- Fink, L. D. (2003). *A self-directed guide to designing courses for significant learning*. Diambil pada tanggal 28 Oktober 2004 dari <http://www.byu.edu/fc/pages/tchlrnpages/fink/fink1.doc>.
- Fleener, J.W. & Prince, J.M. (1997). *Using 360-feedback in organizations. An annotated bibliography*. Greensboro, North Carolina: Center for Creative Leadership
- Freddano, M. & Siri, A. (2012). Teacher training for school self-evaluation. *Procedia - Social and Behavioral Sciences* 69, 1142 – 1149.
- Fullan, M. (2007). *The new meaning of educational change*. New York, N.Y.: Teacher College, Columbia University.
- Hansmann, K.W. & Ringle, C.M. (2004). *SmartPLS Manual*. Hamburg, Germany: Förderverein Industrielles Management an der Universität Hamburg.
- Hattie, J. & Timperley, H. (2007). The Power of feedback. *Review of Educational Research*, 77 (1), 81-112.
- Havelock, R.G. (1969). *Planning for innovation. A comparative study of the literature on the dissemination and utilization of scientific knowledge*. Ann Arbor – Michigan: University of Michigan.
- Helterbran, V.R. (2008). The ideal professor: Student perceptions of effective instructor practices. *Education*, 129 (1), 125-138.
- Heungsun Hwang. (2011). *GesCA user's manual*. Diambil dari www.sem-gesca.org pada tanggal 15 April 2013.

- Imai, Masaaki. (2008). *The kaizen power*. (Translation Sigit Prawato). Yogyakarta: Think.
- _____. (1998). *Gemba kaizen. pendekatan akal sehat, berbiaya rendah pada manajemen*. (Translation Kristianto Jahja). Jakarta: Yayasan Toyota-Astra dan Divisi Penerbitan Lembaga PPM. (The original book was published in 1997)
- _____. (1996). *Kaizen: The key to japan's competitive success*. New York, N.Y.: Random House, Inc.
- Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics* 1 (4) 85-108.
- Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563-575.
- Liker, J.K & Hoseus, M. (2008). *Toyota culture, budaya Toyota, jantung dan jiwa Toyota way*. (Translation Dian Rahadyanto Basuki). New York: McGraw-Hill. (Buku asli diterbitkan tahun 2008).
- Limbach, B., Duron, R., & Waugh, W. (2008). Become a better teacher: Five steps in the direction of critical thinking. *Research in Higher Education Journal*, 1, 1 – 13.
- Marlow, M. (2009). Supporting teacher professional identity through mentoring activities. *Research in Higher Education Journal*, 2, 1-9.
- Marzano, R.J. et al. (2012). *Becoming a reflective teacher. The classroom strategies series*. Bloomington, IN: Marzano Research Laboratory.
- Marzano, R.J., Frontier, T. & Livingston, D. (2011). *Effective supervision. Supporting the art and science of teaching*. Alexandria, VA: ASCD.
- Menpan. (2009). *Peraturan Menteri Negara Pendayagunaan Aparatur Negara Dan Reformasi Birokrasi (Permennegpan & RB) Nomor 16 Tahun 2009, tentang Jabatan Fungsional Guru dan Angka Kreditnya*.
- Miles, B.M. & Huberman, A.M. (1994). *Qualitative data analysis*. Beverly Hills: SAGE Publication.
- Multon, K.D. (2012). Interrater reliability. *Encyclopedia of research design*. Ed. Neil J. Salkind. Thousand Oaks, CA: Sage, 2010 pp. 627-629. Sage reference Online.
- Ochuba, V.O. (2009). Improving the quality of education in Nigeria through effective inspection of schools [Versi elektronik]. *Education*, 129 (4), 734-741.
- Olive, M., Mathers, C. & Laine, S. (2009). Effective evaluation. Principal leadership. *Research Library*, 9 (7), 16 – 21.
- Pang, N.S.K. (2006). *Managing school change through self-evaluation in the era of globalization*. Paper to be reviewed for the 19th Annual World ICSEI Congress Fort Lauderdale, Florida January 3-6, 2006.
- Sallis, E. (2007). *Total quality management in education. Manajemen mutu pendidikan*. (Translation Ahmad Ali Riyadi & Fahrurrozi). London: Kogan Page Limited. (The original book was published in 2002).
- Schumacker, R.E. & Lomax, R.G. (2004). *A beginners,s guide to structural equation modeling*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Stronge, J.H. & Hindman, J.L. (2006). *The teacher quality index: A protocol for teacher selection*. Alexandria, VA: ASCD.
- Surya, Mohamad. (2007). *Mendidik guru berkualitas untuk pendidikan berkualitas*. Paper presented in general course at Graduate Program and PGSD of PGRI University in Yogyakarta.
- Wolbring, T. (2012). Class attendance and students' evaluations of teaching: Do no-shows bias course ratings and rankings? [Versi online]. *Evaluation Review*, 36(1), 72-96.