

PERFORMANCE ASSESSMENT: SEBUAH DILEMA PENILAIAN HASIL PEMBELAJARAN

**Dwi Setiyadi
FPBS IKIP PGRI Madiun**

Abstract :

This research is attempted to figure out the nature and essence of performance assessment in the point of view of lecturers of higher educational institution around Madiun Municipality. Theories have launched their judgement that performance assessment is regarded better than a sole pencil-and-paper test, in terms that it records the broader scope of students achievement, not only cognitively but also affectively and psychomotorically. In spite of the approval from lecturers that performance assessment is theoretically better than pencil-and-paper test, they (lecturers) are reluctant to implement the performance assessment model, for its complicated natures. The research method applied to get data is survey, so as to draw 'mapping' of lecturers responds on performance assessment. The result of the analysis shows contradiction between what lecturers say and what they do in relation to the implementation of performance assessment.

Key words : performance assessment, dilemma

Pendahuluan

Pada awal tahun 1990-an banyak sekali pembuat kebijakan yang mengalihkan perhatiannya pada penilaian kinerja (performance assessment). Penilaian kinerja merupakan suatu pendekatan yang digunakan untuk mengukur kondisi siswa berdasarkan pada cara yang dilakukan oleh siswa dalam melakukan tugas khusus (dikehendaki). Secara teoretis ketika siswa memilih antara benar dan salah dalam soal pilihan biner, siswa tersebut dikatakan melakukan suatu tugas, meskipun tugas tersebut berupa tugas modest. Tetapi penguji yang menggunakan penilaian kinerja memiliki skema pengukuran dari dalam pikiran yang sangat berbeda dengan soal pilihan biner atau pilihan ganda. Kenyataannya tes yang hanya menggunakan kertas dan pensil tidak begitu cukup memuaskan sehingga memaksa para pendidik untuk menggunakan prosedur penilaian kinerja.

Penilaian kinerja kemudian dianggap sebagai penilaian hasil pembelajaran yang lebih baik dibandingkan penilaian dengan test oleh sebagian besar dosen dengan sebuah pemahaman bahwa performance assessment merupakan penilaian yang lebih dekat dengan realita kemampuan pembelajar. Namun meskipun mereka sepakat bahwa performance assessment merupakan model penilaian yang lebih baik, mereka juga mengakui bahwa performance assessment jarang mereka lakukan dalam menilai hasil pembelajaran dalam kelas mereka.

Penelitian ini diselenggarakan untuk melakukan pemetaan tentang keberterimaan performance assessment dalam pembelajaran yang dilakukan dosen, serta sejauh mana para dosen telah menerapkan performance assessment.

Performance Assessment

Makna dan Pengertian

Meskipun sudah banyak model penilaian yang diberikan kepada siswa dalam berbagai cara, ketika mereka berbicara tentang penilaian kinerja, mereka mulai berpikir mengenai *sebuah penilaian di mana siswa diharapkan mampu memberikan tanggapan asli/murni*. Kebanyakan penguji (seperti guru) melakukan pengamatan terhadap proses penyusunan tanggapan yang dilakukan oleh siswa, karena memang kenyataannya guru perlu mengamati kinerja siswa dan kemudian melakukan keputusan penilaian atas kinerja tersebut. Fitzprick dan Morrison (1971) menyatakan bahwa tidak ada perbedaan yang hakiki antara

penilaian kinerja dengan penilaian jenis lain. Mereka menandakan bahwa perbedaan antara penilaian kinerja dengan penilaian konvensional lain hanya pada sejauh mana sebuah penilaian itu mampu memberikan stimulus atas situasi yang dimaksud - dalam artian bahwa penilaian tersebut mampu memperkirakan lingkup perilaku setiap siswa berkaitan dalam menyusun inferensi yang diharapkan.

Inferensi yang dibuat oleh guru dalam hal ini meliputi sejauh mana setiap siswa memiliki kemampuan bekerjasama. *Keputusan pendidikan* yang dihasilkan oleh guru selanjutnya akan menetapkan apakah sekelompok siswa tertentu memerlukan instruksi tambahan supaya mampu menyelesaikan tugas tersebut atau guru bias langsung melanjutkan penugasan dalam kerangka tujuan instruksional yang lain. Dalam hal ini, sebenarnya tugas guru adalah mengamati untuk mengetahui kemampuan masing-masing siswa untuk bekerjasama dalam menyelesaikan tugas secara bersama-sama.

Tentunya jelas bagi anda bahwa pendidik yang berbeda akan menggunakan istilah *performance assessment* untuk mengacu pada jenis pendekatan penilaian yang berbeda. Misalnya, beberapa pendidik akan menganggap bahwa soal jawaban pendek dan essay merupakan bentuk dari penilaian kinerja. Dengan kata lain mereka akan menyamakan penilaian kinerja dengan setiap bentuk penilaian tanggapan terstruktur. Pendidik yang lain akan lebih ketata dalam menerapkan prosedur pengukuran yang disebut sebagai penilaian kinerja. Misalnya, seorang pembuat penilaian kinerja bersikukuh bahwa penilaian kinerja yang benar harus memiliki setidaknya tiga unsure:

- Kriteria ganda – yaitu, bahwa kinerja siswa harus diukur dan ditetapkan dengan menggunakan lebih dari satu kriteria penilaian. Misalnya, kemampuan seorang siswa berbahasa Spanyol diukur berdasarkan pengucapan, tata bahasa dan kosa kata yang digunakan siswa tersebut.
- Standard kualitas yang tetap – yaitu, bahwa setiap kriteria yang digunakan untuk menilai siswa harus bias digunakan terus menerus untuk menilai kualitas kinerja siswa.

Penilaian dengan penetapan – yaitu, tidak seperti selayaknya *response-selected test* yang bias dilakukan oleh mesin atau komputer, sekali diprogram, tidak lagi membutuhkan sentuhan tangan manusia; penilaian kinerja yang sebenarnya bergantung pada putusan manusia untuk menetapkan apakah kinerja siswa diterima atau tidak.

Beberapa penasehat bagi penilaian kinerja akan lebih memilih sebuah penilaian yang *authentic* bagi siswa dalam artian bahwa penilaian tersebut

lebih bias mewakili kondisi yang sebenarnya daripada dunia sekolah atas berbagai masalah yang ada. Beberapa pembuat penilaian kinerja akan sangat bangga jika mereka bisa membuat penilaian dalam lingkup dunia sekolah yang lebih bersifat respon terstruktur daripada *selected-response*. Pendek kata, para pembuat penilaian kinerja seringkali memberikan pendekatan yang berbeda-beda dalam melakukan pengukuran terhadap siswa berdasarkan bagaimana kinerja mereka.

Anda juga seringkali menjumpai pendidik yang menggunakan istilah lain untuk menjelaskan penilaian kinerja. Misalnya, mereka akan menggunakan istilah *authentic assessment* (karena prosedur penilaian lebih berhubungan dengan kehidupan nyata) atau *alternative assessment* (karena penilaian tersebut memberikan satu alternatif pada test tradisional, tes kertas-dan-pensil).

Mengapa harus memilih penilaian kinerja?

Para penasihat penilaian kinerja banyak memberikan dukungannya terhadap penilaian kinerja karena beberapa factor. Mehrens (1990), seorang spesialis pengukuran pendidikan, menyatakan bahwa ada beberapa alasan mengapa para penasihat mendukung penilaian kinerja. Setidaknya ada tiga hal yang dipercayai oleh Mehrens sebagai alasan utama mengapa mereka mendukung penggunaan penilaian kinerja:

1. Ketidakpuasan terhadap *selected-response test*. Para pembuat penilaian kinerja meyakini bahwa: karena tes pilihan ganda dan pilihan biner benar-salah hanya mengukur sebagian saja dari kemampuan siswa yang sebenarnya, hasil tes tersebut tidak mampu untuk menggambarkan ketrampilan berpikir yang lebih tinggi, seperti kemampuan siswa menyelesaikan masalah, membuat kesimpulan, atau berpikir mandiri.
2. Pengaruh dari psikologi kognitif. Psikologi kognitif mempercayai bahwa siswa harus mendapatkan pengetahuan substansial isi dan pengetahuan prosedural sekaligus dalam suatu tugas. Para psikolog di bidang ini menyatakan bahwa tugas kognitif memerlukan kedua bentuk pengetahuan tersebut, namun demikian setiap tugas memiliki penekanan tertentu yang

berbeda pada kedua jenis pengetahuan tersebut (Snow and Holman, 1989). Karena ada pengetahuan prosedural tertentu yang tidak bisa dinilai melalui selected-response test, banyak psikolog kognitif

yang menggunakan model penilaian kinerja di bidang pendidikan untuk meyakinkan mereka atas kemampuan penilaian kinerja dalam memberikan penekanan instruksional pada perolehan pengetahuan prosedural siswa.

3. Dampak instruksional yang seringkali merugikan dari sebuah tes konvensional. Karena adanya ketentuan-ketentuan yang harus diikuti oleh para pelaku pendidikan, maka guru seringkali menekankan isi/ materi pendidikan secara instruksional dalam sebuah bentuk tes tertentu. Akibatnya jika instruksi diarahkan secara langsung pada tes, skor yang diperoleh siswa bisa saja sangat tinggi meskipun penguasaan ketrampilan atau pengetahuan di lingkup tersebut sangat kurang. Karena kebanyakan pendidik yang menyadari bahwa tes yang didasarkan atas ketetapan yang berlaku seringkali akan berpengaruh secara langsung terhadap materi pembelajaran yang akan disampaikan guru, maka tes kinerja akan menghasilkan tujuan instruksional yang lebih baik, jika dibandingkan dengan tes kertas-dan-pensil. Dengan adanya tujuan instruksional yang lebih memadai dengan tes kinerja, dalam kerangka memenuhi ketetapan yang berlaku, maka aktivitas instruksional guru akan berubah ke arah yang lebih baik.

Ada beberapa alasan yang lain mengapa kita perlu menggunakan penilaian kinerja, yang menurut Mehrens (1992) menjadi motivasi kuat dalam menggunakan model penilaian kinerja. Di antara alasan itu adalah dengan melakukan penilaian kinerja, guru akan mampu membuat kemajuan/perkembangan instruksional yang diawali dengan penilaian kinerja semacam itu. Namun demikian, ia kurang sependapat dengan penilaian kinerja yang dipergunakan sebagai tujuan akuntabilitas edukasional; yaitu, untuk menetapkan seberapa efektif guru memberikan instruksi pada siswanya.

Mengidentifikasi Tugas yang Sesuai dalam Penilaian Kinerja

Penilaian kinerja menghendaki siswa untuk menanggapi tugas yang berskala kecil namun memiliki signifikansi yang lebih tinggi daripada tugas yang besar namun tidak signifikan. Jadi, siswa akan lebih baik diberi tugas untuk melakukan eksperimen dalam pelajaran kimia, daripada diberi 50 butir soal pilihan ganda, lalu kemudian mereka diminta untuk membuat laporan mengenai interpretasinya tentang hasil-hasil yang diperoleh dari eksperimen tersebut serta membuat analisis kritis dari prosedur yang dilakukannya. Dilihat dari sudut pandang guru kimia, gambaran mengenai status pemahaman setiap siswa akan lebih bisa dilihat melalui tanggapan tunggal yang kompleks (dari eksperimen) daripada melalui jawaban atas 50 butir soal tes *mini* pilihan ganda. Dengan adanya signifikansi dari tugas yang digunakan sebagai pendekatan 'ujian kinerja' untuk melakukan penilaian kelas, maka tugas yang akan diberikan harus melalui proses pemilihan yang seksama. Dengan demikian, seorang guru harus: 1) membuat ujian kinerja sendiri, 2) memilih dengan seksama tugas ujian kinerja yang paling tepat, yang bisa didapatkan dengan membandingkan model-model yang dihasilkan oleh pendidik yang lain.

Inferensi dan Tugas

Dari berbagai macam buku tentang penilaian kelas yang ada pada saat ini, hal yang terpenting dalam melakukan penilaian terhadap siswa adalah inferensi apa yang akan ditetapkan atas siswa tersebut, dan keputusan apa yang akan dibuat berdasarkan pada inferensi tersebut. Misalnya, anda adalah seorang guru sejarah yang sedang memikirkan tentang kurikulum sejarah yang akan diberlakukan dalam waktu mendatang. Kemudian, anda menemukan bahwa yang sebenarnya anda inginkan dalam pembelajaran sejarah adalah bagaimana mengajari siswa bagaimana menggunakan pengetahuan masa lalu dalam pelajaran sejarah untuk menyelesaikan masalah yang timbul pada saat ini atau di saat yang akan datang, setidaknya untuk hal-hal yang serupa dari masa lalu. Anda biasanya lalu membuat tes pilihan biner benar-salah sebanyak 1.500 soal, lalu mengukur kemampuan siswa berdasarkan skor yang anda dapatkan dari hasil tes. Saat ini anda mencoba untuk meninggalkan model tes pilihan ganda dan beralih pada penilaian kinerja, dengan menggunakan strategi penilaian kinerja dan memilih satu tugas tes kinerja yang mampu untuk menggambarkan kemampuan setiap siswa dalam menerapkan pengetahuan mereka tentang masa lalu untuk menyelesaikan persoalan di masa kini maupun di masa yang akan datang.

Popham (1994) menunjukkan sebuah grafik yang menerangkan hubungan antara 1) pokok-pokok tujuan edukasional guru, 2) inferensi yang bisa digambarkan guru atas siswanya, 3) tugas tes kinerja yang bisa digunakan untuk memperoleh data yang mampu

mendukung guru dalam membuat inferensi. Sebagaimana anda ketahui, tujuan instruksional guru akan menjadi sumber acuan bagi penentuan inferensi, dan tugas dalam penilaian tersebut bisa digunakan sebagai bukti-bukti atau data yang diperlukan guru untuk meniadakan inferensi yang memadai tentang kemampuan siswa dalam menerapkan pengetahuan sejarahnya dalam menyelesaikan persoalan di saat ini dan yang akan datang. Berdasarkan pada tingkat kemampuan siswa dalam menguasai tujuan instruksional tersebut, guru juga akan dapat menentukan keputusan tentang berapa banyak lagi instruksi yang diperlukan, jika masih ada.

Dilema Penyusunan Generalisasi

Salah satu kesulitan psikometris yang paling serius dalam penilaian kinerja adalah bahwa di dalam penilaian kinerja siswa hanya menanggapi tugas yang berskala kecil dibandingkan dengan ujian kertas-dan-pensil, sehingga sangat sulit untuk membuat generalisasi yang akurat mengenai kemampuan siswa, atau ketrampilan dan pengetahuan yang diperoleh siswa.

Sebagai contoh, anda ingin mengetahui kemampuan siswa secara pasti tentang perkalian angka dua digit. Jika anda hanya memiliki waktu 30 menit untuk melakukan penilaian, anda bisa memberi tugas kepada siswa untuk mengerjakan 20 butir soal perkalian. Dari jawaban yang dibuat oleh siswa atas 20 butir soal perkalian tersebut, anda sudah bisa memberikan penilaian yang adil atas kemampuan siswa mengalikan angka-angka dua digit. Berdasarkan pada hasil kinerja siswa dengan sample yang cukup memadai tersebut, anda akan mampu membuat kesimpulan bahwa si Javier menguasai perkalian angka dua digit, atau bahwa si Fred tidak menguasai perkalian angka dua digit. Kesimpulan tersebut bisa anda buat karena anda telah menarik sample sebaik-baiknya (yang akan anda buat inferensinya) bahwa anda akan mampu membuat inferensi yang meyakinkan tentang kemampuan mereka menyelesaikan soal yang sama. Jika siswa hanya diberi waktu 30 menit untuk menyelesaikan jenis soal yang lebih besar (bukan berupa penilaian psikometrik kelas) mereka mungkin hanya akan mampu menyelesaikan satu soal saja. Misalnya, anda menyuruh siswa untuk menyelesaikan satu soal yang mendasarkan perkalian angka dua digit dengan menggunakan manipulasi, maka sudah untung jika mereka mampu menyelesaikannya dengan menggunakan perumusan aslinya. Kemudian dengan hanya satu tugas saja, bagaimana anda akan mampu menggambarkan kemampuan siswa dalam menerapkan perkalian angka dua digit.

Dan sekarang anda mengerti, seperti itulah permasalahan yang dihadapi oleh penilaian kinerja. Karena jawaban siswa hanya mencakup soal dalam skala kecil, maka guru akan menemukan kesulitan ketika harus menjelaskan kemampuan siswa secara umum. Taruhlah, jika hanya ada satu tes saja, dan kebetulan siswa bisa mengerjakannya dengan baik, lalu apakah anda bisa mengatakan bahwa kemampuan siswa tersebut baik hanya dari satu soal tes saja, atau karena ia beruntung? Sebaliknya, jika siswa tersebut gagal menjawab soal itu, anda akan mengatakan bahwa kemampuannya buruk? Atau ada faktor yang hilang di dalam tes tersebut sehingga ia gagal sedangkan seharusnya ia bisa melakukannya dengan lebih baik dari itu. Sebagai guru kelas anda akan dihadapkan pada permasalahan yang klasik, dilemma dalam memilih dua bentuk soal tes tersebut. Meskipun penilaian kinerja akan dianggap lebih mampu menggambarkan kemampuan siswa (sebagaimana dalam tujuan instruksional), penilaian atas kemampuan siswa dengan menggunakan tes kinerja perlu mendapat perhatian yang lebih cermat. Dan tidak ada satu penyelesaian yang paling sempurna untuk menghadapi suatu dilemma. Namun demikian setidaknya anda tahu caramenghadapi dilemma tersebut secermat mungkin. Dari contoh tersebut, yang perlu anda perhatikan adalah strategi yang cermat dalam memilih tugas yang harus diselesaikan siswa. Di antara pertimbangan-pertimbangan yang penting dalam memilih tugas pada sebuah tes kinerja adalah bahwa tugas tersebut mampu menggambarkan (setidaknya yang terdekat dengan) kemampuan yang sebenarnya. Jika anda mampu menjaga generalisasi sebelum anda melakukan tes, maka anda akan mampu menciptakan inferensi atas siswa berdasarkan atas tes kinerja tersebut.

Kriteria Evaluasi Tugas Tes Kinerja

Berikut adalah daftar yang bisa digunakan untuk mengevaluasi tugas tes kinerja, jika akan membuat atau menilai tes kinerja.

- Generalisabilitas. Keseragaman dalam kinerja siswa di dalam tes tersebut bisa digunakan sebagai generalisasi pada tes-tes yang lain.
- Authenticity. Adanya kemiripan/kesamaan antara kejadian yang ditemui siswa dalam kehidupan nyata.
- Multiple Foci. Apakah tes tersebut mengukur hasil-hasil instruksional ganda atau hanya akan memilih satu macam hasil.
- Teachability. Apakah kemampuan siswa bertambah dengan upaya instruksional yang dilakukan oleh guru.
- Keadilan. Apakah tes tersebut adil bagi semua siswa?
- Feasibilitas. Apakah tugas tersebut bisa diimplementasikan dalam kaitannya dengan biaya, waktu, dan kebutuhan peralatnya.
- Scorability. Apakah mampu menunjukkan tanggapan siswa.

Dalam kondisi tertentu, anda bisa menggunakan hanya sebagian dari ketujuh kriteria tersebut, jadi anda tidak harus menggunakan tujuh kriteria itu sebagai dasar untuk membuat tes kinerja secara bersama-sama. Dalam kenyataannya, satu tes kinerja tidak akan mampu memenuhi seluruh kriteria tersebut di atas.

Mengidentifikasi Kriteria Scoring

Penilaian kinerja didasarkan pada prosedur pengukuran ‘tanggapan-terstruktur’ di mana siswa memberikan tanggapan dengan mengembangkan pengetahuannya bukan memilih tanggapan yang harus diberikannya. Namun demikian, karena semua jenis tes harus melalui prosedur pengukuran, maka penetapan skor pada tes kinerja jauh lebih rumit dibandingkan dengan tes ‘tanggapan-terpilih’. Dalam hal penetapan skor tersebut, penguji harus memiliki *criteria* tertentu agar penilaian tersebut cermat. Berikut adalah penjelasan mengenai kriteria yang harus dipenuhi dalam melakukan penilaian kinerja.

Suatu *criteria*, menurut Webster Dictionary, adalah ukuran standard yang dipakai sebagai dasar membuat keputusan. Dalam melakukan penilaian terhadap tanggapan siswa dalam tes kinerja, anda dilibatkan dalam suatu proses pembuatan keputusan mengenai ketepatan siswa dalam menjawab soal. Kriteria yang digunakan sebagai dasar dalam penetapan skor tersebut akan sangat berpengaruh terhadap kegiatan evaluasi secara keseluruhan. Dalam hal ini, anda mengerti bedanya antara kata *criterion* dan *criteria*.

Kriteria yang digunakan dalam menilai jawaban siswa terdiri dari beberapa hal, misalnya: rubrik, petunjuk penetapan skor, dan dimensi penetapan skor. Kriteria yang digunakan tersebut akan menentukan factor-faktor yang dipertimbangkan ketika melakukan penilaian kinerja siswa. Jawaban siswa seperti apa yang bisa diterima dan seperti apa yang tidak bisa diterima.

Karena penilaian kinerja akan membuahkan jawaban yang lebih kompleks, maka kriteria yang digunakan juga harus lebih luas, sehingga mampu mencakup semua jenis jawaban siswa. Ketika kriteria telah ditetapkan, maka skala penilaian juga harus ditetapkan; yang biasanya terdiri dari 5 (lima) poin. Misalnya: 5 = pujian, 4 = sangat memuaskan, 3 = memuaskan, 2 = kurang, dan 1 = tidak tepat.

Proses intelektual inti dalam menetapkan satu kriteria adalah menemukan factor apa yang dianggap penting yang bisa membedakan antara jawaban yang diterima dan tidak diterima.

Contoh Tugas Tes Kinerja dan System Penetapan Skor

Di bagian ini anda akan melihat contoh tes kinerja dan cara penetapan skor dari jawaban siswa. Gambar 8-3 menunjukkan ada 4 (empat) macam tugas untuk menilai kinerja komunikasi oral siswa. Perlu dicatat bahwa variasi tugas speaking satu-lawan-satu cenderung akan menghasilkan generalisasi mengenai kemampuan komunikasi siswa satu-lawan-satu. Perlu dicatat bahwa petunjuk penetapan skor didasarkan pada kriteria *dimensi*.

Sumber Kesalahan dalam Memberikan Skor Kinerja Siswa

Ada tiga penyebab utama sehingga penilaian kinerja menjadi tidak akurat, yaitu: 1) *skala scoring*, 2) *pemberi skor sendiri*, 3) *prosedur pemberian skor*.

Kesalahan Instrumen

Kesalahan dalam penilaian seringkali disebabkan karena instrumen penilaian tidak memadai, misalnya kriteria penilaian yang tidak jelas.

Kesalahan prosedur

Kesalahan dalam lingkup prosedural ini muncul karena banyaknya kriteria yang harus dinilai, sehingga memicu penyimpangan. Dalam hal ini, semakin sedikit kriteria yang dipakai, maka akan semakin baik hasilnya.

Kesalahan pribadi guru

Kesalahan yang munculnya dari diri pribadi guru, yang terdiri dari tiga macam kesalahan:

- Generosity error* – guru memberikan penilaian yang terlalu tinggi dari yang seharusnya,
- Severity error* – guru memberikan penilaian yang terlalu rendah dari yang seharusnya,
- Central tendency error* – guru cenderung menaruh pada nilai tengah.

Metodologi

Metodologi research yang digunakan dalam penelitian ini adalah survey, yang dilakukan dengan mengumpulkan data atau informasi yang diperlukan untuk menjawab pertanyaan penelitian yang telah ditetapkan. Survey dilakukan dengan menggunakan instrumen yang berupa: 1) questionair atau daftar pertanyaan terkait dengan keberterimaan dan implementasi performance assessment, yang harus dijawab dengan menuliskan infoemasi yang diperlukan, 2) daftar pertanyaan atau interview protocol yang harus disampaikan oleh peneliti secara lisan kepada responden untuk mendapatkan informasi lisan. Data diolah dengan teknik *constant comparative method* (Strauss dan Glasser, 1990). Responden berjumlah 40 yang dipilih dengan teknik purposive sampling yang mendasarkan pada: perguruan tinggi dan jenis mata kuliah.

Diskusi Hasil Analisis: Dilema Performance Assessment dari teori menuju realita

Di luar kelebihan-kelebihan yang dimiliki oleh performance assessment, dalam kenyataannya performance assessment jarang sekali digunakan sebagai pendekatan evaluasi standard sebuah pembelajaran di sekolah. Hal itu disebabkan oleh fakta dilematis untuk menerapkan performance assessment. Di antara dilema-dilema yang terdapat pada performance assessment adalah:

1. Performance assessment hanya melakukan penilaian pada area yang ‘tugas’ siswa yang sangat sempit bila dibandingkan dengan tes konvensional. Jika selected-response test bisa dilakukan siswa untuk melakukan ‘tugas’ yang luas dalam waktu yang relatif singkat, misalnya 20 butir soal dalam waktu 30 menit; maka performance assessment dalam waktu yang sama hanya akan memberikan satu ‘tugas’ saja. Dalam hal ini, ada hal yang perlu dipertanyakan mengenai reliabilitas dari tes tersebut. Apakah hanya dengan melakukan satu ‘tugas’ saja siswa akan dikatakan berprestasi *baik* atau *kurang*? Padahal pada saat mengerjakan tes tersebut belum tentu siswa tersebut berada pada kondisi yang sama dengan kondisi normal.
2. Sebagaimana dijelaskan pada butir (1) di atas, konsekuensinya adalah bahwa performance assessment memerlukan waktu yang cukup banyak. Untuk sampai pada *keputusan pendidikan* yang harus ditetapkan oleh guru, performance assessment seharusnya diterapkan dalam kondisi yang berbeda-beda dengan lingkup topik permasalahan yang bervariasi, sehingga hasil dari performance assessment mencerminkan kemampuan yang sebenarnya dari siswa yang dinilai dalam kerangka penerapan nilai-nilai yang dipelajarinya dalam kehidupan sehari-hari. Jika performance assessment harus diterapkan sebagai pendekatan evaluasi standard di sekolah, maka perlu dipikirkan bentuk ‘tugas’ yang paling sesuai serta alokasi waktu yang memadai.
3. Performance assessment yang baik seharusnya dirancang oleh guru yang bersangkutan, karena gurulah yang paling mengerti kondisi siswa. Dengan demikian guru mengerti bentuk ‘tugas’ yang harus dinilai dalam kondisi tertentu yang representatif untuk mendapatkan hasil yang terdekat dengan kemampuan siswa yang sebenarnya. Dengan demikian, ada satu hal yang dilematis berkaitan dengan hal ini, yaitu: *apakah semua guru memiliki kemampuan, perhatian, dan kemauan untuk melakukan performance*

assessment? Karena performance assessment menuntut guru yang kreatif, peduli terhadap perkembangan siswa, dan memiliki integritas tinggi terhadap ilmu dan tugasnya. Lalu ada berapa guru di negeri ini yang memiliki kriteria tersebut, sedangkan mereka seringkali menjalankan proses pembelajaran sebagai sesuatu yang *instant* dengan berbagai alasannya.

4. Dalam penerapan penilaian hasil belajar siswa, standardisasi sistem penilaian dalam skala yang luas seringkali menjadi kebutuhan. Untuk memenuhi kebutuhan standardisasi sistem penilaian berskala luas, performance assessment akan mengalami hambatan yang cukup signifikan. Penyebab hambatan tersebut antara lain: 1) performance assessment lebih bersifat individual yang memberlakukan perlakuan serta tipe yang berbeda pada setiap siswa disesuaikan dengan kondisi yang ada pada siswa tersebut. 2) Jika memang harus ada standard penilaian dengan pendekatan ini, maka akan melibatkan banyak sekali pihak untuk mendiskusikan secara komprehensif hasil-hasil belajar yang dicapai siswa. 3) Besarnya resiko penyimpangan dalam melakukan performance assessment, misalnya: *penyimpangan instrumen penilaian*, utamanya dalam menetapkan kriteria penilaian yang tepat. Hal ini membutuhkan pemikiran yang mendalam. *Penyimpangan prosedural*, berkaitan dengan jumlah kriteria yang harus dinilai – bisa terlalu banyak atau terlalu sedikit. *Kesalahan bias-personal*, yang berkaitan dengan generosity error, severity error atau central tendency error.

DAFTAR BACAAN

- Anita Lie, 2008, *Cooperative Learning: Mempraktikkan Cooperative Learning di Ruang-ruang Kelas*, Jakarta: PT. Gramedia Widiasarana Indonesia.
- Berns, Robert G. and Patricia M. Erickson, 2001, *Contextual Teaching and Learning: Preparing Student for the New Economy*, Bowling Green, US: Bowling Green University Press.
- Borg, Walter R., and Meredith Damien Gall, 1979, *Educational Research: An Introduction, Third Edition*, New York: Longman Group Limited.
- Clifford, Mattew and Marica Wilson, 2000, *Contextual Teaching, Professional Learning, and Student Experiences: Lesson Learned from Implementation*, Education Brief, Madison, Wisconsin: Center on Education and Work, University of Wisconsin.
- _____, 2007, *What is Contextual Teaching and Learning*, <http://www.cew.wisc.edu/teachnet/ctl/>
- Gagnon Jr, George W., and Michelle Collay, 2001, *Designing for Learning: Six Elements in Constructivist Classrooms*, Thousand Oaks, California: Corwin Press, Inc.
- Genesee, Fred and John A. Usphur, 1997, *Classroom-Based Evaluation in Second Language Education*, Cambridge: Cambridge Language Education.
- H.B. Sutopo, 2006, *Metodologi Penelitian Kualitatif, Dasar Teori dan Terapannya dalam Penelitian*, Surakarta: Universitas Sebelas Maret.
- Johnson, Elaine B., 2007, *Contextual Teaching and Learning: what it is and why it's here to say*, diterjemahkan oleh Ibnu Setiawan, Bandung: Mizan Learning Center.
- Joyce, Bruce, Marsha Weil with Emily Calhom, 2000, *Models of Teaching*, London: Allyn and Bacon.
- Kempler, David, 2000, *Action Learning and Research*, London: British Libraray Cataloguing in Publication Data.
- Mc Neil, John D., 1996, *Curriculum: A Comprehensive Introduction*, University of California, Los Angeles: Harper Collins College Publisher.
- Popham, W. James, 1994, *Classroom Assessment*, London: Allyn and Bacon Co Ltd.
- Richards, Jack C. and Theodore S. Rodgers, 2001, *Approaches and Methods in Language Teaching*, Cambridge: Cambridge University Press.