

Australian Council for Educational Research (ACER)

Sharing Session #2: Penilaian Pendidikan Berbasis Kinerja

9th May, 13.30 – 15.30 (120’)

Kerangka Diskusi

Topik untuk *Sharing Session* ini adalah mengenai **Penilaian Berbasis Kinerja** di Ruang Kelas: (i) **Apa yang dimaksud dengan PBK**, (ii) **bagaimana cara kita mendesain penggunaannya di ruang kelas**, serta (iii) **apa saja ragam teknik penerapannya**.

Topik ini menarik untuk diangkat karena sepertinya ada *demand* untuk memahami dengan lebih baik cara penerapannya di kegiatan pembelajaran. Seperti sudah kita ketahui bersama, Kurikulum 2013 menerapkan model pembelajaran yang menekankan pada pencapaian kompetensi peserta didiknya. Metode pembelajaran seperti ini tentunya memerlukan metode penilaian yang berbeda dari sebelumnya. Karena tidak semata menghitung jumlah jawaban benar atau salah di atas kertas, tapi bagaimana guru mampu menilai aplikasi pengetahuan yang sudah dipelajari oleh siswa.

Isi Sesi

Apa itu Penilaian berbasis Kinerja?

Penilaian Berbasis Kinerja (PBK) yang didesain dengan baik dapat memungkinkan guru secara obyektif menilai kompetensi siswa, yang terdiri atas: (i) pengetahuan, (ii) sikap; dan (iii) keterampilan. Menilai kompetensi tentunya tidak dapat semata diukur melalui “*paper and pencil test*” karena **mengetahui bagaimana melakukan sesuatu tidak sama dengan dapat melakukan sesuatu**. Maukah Anda naik pesawat yang diterbangkan oleh pilot yang paham teori, namun belum pernah menerbangkan pesawat sama sekali?

Bagaimana mendesain penggunaan Penilaian berbasis Kinerja di ruang kelas?

Ada tiga hal yang perlu dilakukan. **Pertama, menyusun tugas berstruktur**. Menurut Khattri, Reeve dan Kane (1998), tugas berstruktur memiliki instruksi yang memandu pengerjaannya, seperti topik, waktu (spesifik atau dapat disesuaikan), sifat dari produk yang diberikan (produk atau proses), sumber daya atau media yang perlu digunakan (disediakan sekolah atau dibawa sendiri oleh siswa), serta derajat kontrol siswa (terbatas, diberikan pilihan atau kontrol penuh).

Menurut Silberman (1996), dalam menyusun tugas, kita dapat mengukur:

- Pengetahuan terkait dengan cara mengumpulkan informasi dan penemuan konsep-konsep

- Pembelajaran tingkah laku (keterampilan) terkait dengan perkembangan kompetensi murid dalam menuntaskan suatu tugas, menyelesaikan permasalahan, serta mengekspresikan dirinya
- Pembelajaran afektif (sikap) yang bertujuan untuk mengukur dan mengklarifikasi perasaan dan minat para siswa

Kedua, menentukan kriteria kinerja. Beberapa hal yang perlu diperhatikan saat menentukan kriteria kinerja adalah:

- Tingkah laku spesifik yang harus muncul dalam penyelesaian tugas
- Seluruh aspek, komponen, ataupun tugas-tugas yang akan dievaluasi
- Kriteria harus singkat, jelas dan masuk akal

Metode untuk menyusun kriteria menurut Airasian (2005), adalah sbb:

- Tentukan komponen tugas yang akan dinilai
- Jumlah komponen tidak terlalu banyak
- Diskusi dengan guru lain
- Kriteria diekspresikan dalam bentuk perilaku atau produk yang memungkinkan untuk diamati
- Urutkan sesuai dengan keurutan pemunculannya
- *Don't start from scratch.* Tidak perlu membuat sendiri dari nol, karena sudah banyak kriteria yang pernah digunakan dan didesain.

Kriteria harus jelas dan terukur, seperti contoh pada *slide 15* di *Hand Out Material 1* untuk contoh tugas yang ada di *slide 14*.

Ketiga, menentukan metode penilaian. Ada beberapa metode penilaian yaitu:

- *Checklist,*
sebuah daftar apa saja yang perlu dilakukan seseorang dalam tugasnya, sebagaimana dicontohkan pada *slide 17, Hand Out Material 1*. Paling mudah untuk dilakukan, tetapi harus rinci.
- *Rating,*
Bentuknya dapat dikembangkan dari *checklist* dengan menyertakan frekuensi (lihat *slide 20, Hand Out Material 1*) atau aspek lainnya sebagai dasar penilaian. Dapat menggunakan skala Likert (1 - 5), Thurstone (1 – 11), atau skala lainnya tergantung tujuan penilaian. Contohnya untuk skala 1 – 100, perbedaan antara skor 70 dan 75 harus memiliki arti.
- Rubrik.
Bentuknya dapat *generic* atau *task specific*, dan sifatnya dapat *analitic* atau *holistic*.

Rangkuman Hasil Diskusi

Bagaimana caranya memberikan tugas yang berkaitan dengan pemahaman suatu konsep yang abstrak?

Cari referensi sebanyak-banyaknya. Tidak perlu terlalu takut salah dalam memberikan tugas, karena apabila ada kelemahan dalam desain tugas, hal tersebut akan dan harus jadi masukan dan konsiderasi dalam desain tugas berikutnya.

Mungkinkah menggunakan seluruh ketiga perangkat *scoring* (*checklist*, *rating* dan *rubric*) secara bersamaan dalam penilaian?

Biasanya hanya salah satu saja yang dipakai, dan terkadang ada dua yang dipakai dalam satu tugas.

Terkait pelaporan penilaian, bagaimana agar pelaporan hasil penilaian dapat sesederhana mungkin, namun informatif sehingga mampu menggambarkan proses belajar bagi orangtua?

Salah satu alternatif dalam mengukur perkembangan anak adalah dengan membandingkan capaian pembelajaran satu anak dari waktu ke waktu. Misalnya membandingkannya pada awal semester dan di akhir semester.

Orang tua perlu diberi pemahaman bahwa ada dua jenis model penilaian. Yang pertama, membandingkan dengan siswa lain (*norm-referenced test reporting*); dan yang kedua, membandingkan dengan kriteria (*criterion-referenced test reporting*).

Pada jenis tes yang pertama, skoring tidak dapat memberikan gambaran apapun, kecuali bila ada pembandingnya (misalnya dengan nilai rata-rata).

Pada jenis tes yang kedua, kriteria yang akan digunakan perlu ditetapkan dan disetujui sebelumnya. Orangtua dapat lebih memahami kriteria yang digunakan melalui rapat guru dan orangtua atau bentuk wadah komunikasi lainnya. Guru dapat memberikan informasi terkait progres siswa dalam belajar dan meminta kerjasama orangtua dengan memberitahukan apa yang dapat orangtua lakukan di lingkungan rumah.

Terkait dengan aturan sekolah yang belum mengakomodir penggunaan penilaian berbasis kinerja, guru dapat memulai penggunaannya dengan mengajak guru lain di sekolah yang sama atau di sekolah lain untuk menggunakannya. Cara lainnya adalah dengan mengundang pakar/ahli ke sekolah untuk menjelaskan.

Perlu diingat, peran kegiatan penilaian di dalam kegiatan pembelajaran dapat berupa: (1) *assessment of learning*; (2) *assessment for learning* (termasuk *feedback loop* walaupun dalam bentuk kelas, namun perlu validasi); (3) *assessment as learning*. Rancangan *assessment for learning* memang tidak mudah, tetapi nanti penggunaannya akan sangat mudah.

Bagaimana memberikan tugas berstruktur yang menghasilkan produk penilaian yang baik mengingat kemampuan siswa yang berbeda-beda?

Setiap anak harus diberikan tugas yang sama dan setara dan hasilnya bisa dibandingkan. Namun, pendidik harus menyadari bahwa kemampuan anak berbeda-beda, dan yang utama bukanlah agar semua anak berada di tingkat pembelajaran yang sama, namun agar setiap anak mengalami kemajuan pembelajarannya.

Menggunakan pendekatan *inter-rater reliability*, yaitu ketika lebih dari seorang penilai (minimal 2 orang) menilai seorang siswa menggunakan instrumen penilaian (contohnya rubrik) yang sama. Bila pemahaman masing-masing penilai berbeda, maka instrumen tersebut harus direvisi kembali untuk memperkecil perbedaan persepsi antar penilai. *Inter-rater reliability* sangat berguna bila penilaian dilakukan secara subjektif, seperti penilaian tugas seni atau bahasa.

Terkait keharusan untuk mengacu pada standar capaian pembelajaran, bagaimana dengan kriteria yang digunakan di sekolah non-formal, termasuk pendidikan di rumah (*home-schooling*), yang mungkin tidak menggunakan standar umum tetapi dapat menghasilkan anak yang berprestasi?

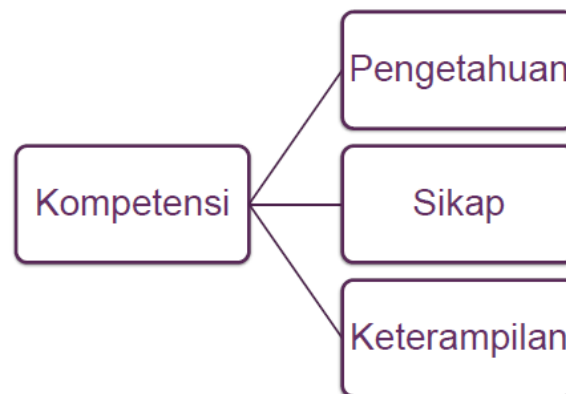
Menggunakan standar yang lain bukannya tidak diperbolehkan, tetapi berarti pendidik memiliki beban untuk harus mencari jalan/*path* lain untuk siswa. Contohnya sekolah yang menggunakan kurikulum Cambridge atau International Baccalaureate (IB).

Indonesia memang sedang bergeser dari *curriculum approach* ke *competency approach* yang menggunakan indikator pencapaian kompetensi (Kompetensi Inti dan Kompetensi Dasar). Penggunaan kriteria dalam penilaian berbasis kinerja diserahkan kepada guru, selama siswa dapat mencapai kompetensi yang telah ditetapkan untuk tingkat pendidikan siswa tersebut.

Hand Out Materials



CAPAIAN PEMBELAJARAN

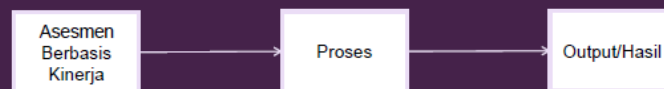


TAKSONOMI KETERAMPILAN

Dyers	Dave	Sympson
Mengamati	Imitasi	Persepsi/Kesiapan/Meniru
Menanya	Manipulasi	Membiasakan
Mencoba	Presisi	Mahir
Menalar	Artikulasi	Alami
Menyaji	Naturalisasi	Orisinil
Mencipta		

- ▷ Tidak dapat diukur melalui "*paper and pencil test*"
- ▷ Mengetahui bagaimana melakukan sesuatu **tidak sama** dengan **bisa** atau **dapat** melakukannya

➔ Penggunaan **asesmen berbasis kinerja yang didisain dengan baik** akan memungkinkan guru-guru untuk secara **obyektif dan menyeluruh** menilai pengetahuan, keterampilan, dan kemampuan siswa



ELEMEN ASESMEN BERBASIS KINERJA



Tugas Berstruktur

TUGAS BERSTRUKTUR



Khattri, Reeve dan Kane (1998)

TUGAS YANG DIBERIKAN



Tugas dapat ditujukan untuk mengukur:

- ▷ Pengetahuan terkait dengan cara mengumpulkan informasi dan penemuan konsep-konsep
- ▷ pembelajaran tingkah laku (keterampilan) terkait dengan perkembangan kompetensi murid dalam menuntaskan suatu tugas, menyelesaikan permasalahan, serta mengekspresikan dirinya
- ▷ Pembelajaran afektif (sikap) yang bertujuan untuk mengukur dan mengklarifikasi perasaan dan minat para siswa

(Silberman, 1996)

WAKTU

- ▷ Dapat ditentukan secara spesifik
- ▷ Disesuaikan dengan kemampuan murid



PRODUK

Aspek yang dinilai dapat berupa:

- ▷ Produk
- ▷ Proses

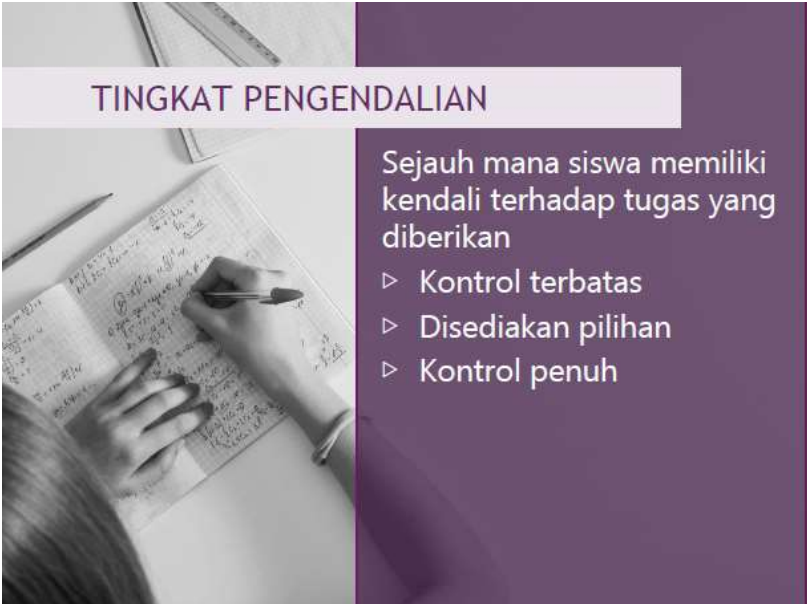




SUMBER DAYA

Sumber daya yang dibutuhkan

- ▷ Siswa dapat memperoleh dan membawanya
- ▷ Disediakan oleh sekolah



TINGKAT PENGENDALIAN

Sejauh mana siswa memiliki kendali terhadap tugas yang diberikan

- ▷ Kontrol terbatas
- ▷ Disediakan pilihan
- ▷ Kontrol penuh

Kriteria Kinerja

KRITERIA KINERJA

- ▷ Tingkah laku spesifik yang harus muncul dalam penyelesaian tugas
- ▷ Seluruh aspek, komponen, ataupun tugas-tugas yang akan dievaluasi
- ▷ Kriteria harus singkat, jelas dan masuk akal
- ▷ Airasian (2005):
 - Tentukan komponen tugas yang akan dinilai
 - Jumlah komponen tidak terlalu banyak
 - Diskusi dengan guru lain
 - Kriteria diekspresikan dalam bentuk perilaku atau produk yang memungkinkan untuk diamati
 - Urutkan sesuai dengan keurutan pemunculannya
 - *Don't start from scratch*

CONTOH: TUGAS

Bayangkan bahwa Anda menjadi saksi dari suatu kecelakaan lalu lintas dan Anda diminta untuk memberikan keterangan tertulis kepada polisi dengan menggunakan bahasa Inggris.

Keterangan Anda harus berjumlah 300 kata yang didalamnya terdapat:

- Waktu dan tempat kejadian kecelakaan
- Rincian spesifik terkait kejadian yang anda saksikan
- Penyebab dari kecelakaan berdasarkan pengamatan anda

Purwono & Randall, 2010

CONTOH: KRITERIA

Kriteria sebelumnya	Kriteria yang sudah direvisi
Pengekspresian gagasan	Gagasan diungkapkan secara komprehensif dan dilengkapi dengan detail-detail yang diperlukan
Keteraturan	Kalimat yang digunakan teratur, logis dan bermakna
Tata bahasa	Tidak membuat kesalahan "grammar" lebih dari 3
Tulisan tangan	Tulisan tangan mudah dibaca
Kalimat lengkap	Menggunakan kalimat lengkap
Pengejaan	Tidak melakukan kesalahan pengejaan lebih dari 4 kali
Tenses	Menggunakan kata kerja yang tepat dalam mendeskripsikan cerita

Purwono & Randall, 2010

Penilaian/ *Scoring*

Perangkat *Scoring*

Checklist

Rating

Rubrik

CHECKLIST

➔ Merupakan daftar dari kriteria performa

Contoh

Checklist untuk Presentasi (Airasian, 2005)

Ekspresi fisik

- ___ A. berdiri tegak dengan postur dan wajah menghadap kepada peserta/audience
- ___ B. Mengubah ekspresi muka dan intonasi suara untuk memberikan penekanaan-penekanan pada saat presentasi
- ___ C. mempertahankan kontak mata dengan hadirin

RATING SCALE

Rating Scales untuk Presentasi Lisan (Airasian, 2005)

Petunjuk : Nyatakan seberapa sering siswa menunjukan setiap tingkah laku dibawah ini pada saat mereka melakukan presentasi secara lisan. **Lingkari 4** apabila siswa selalu menampilkan tingkah laku tersebut. **Lingkari 3** apabila siswa sering menampilkan tingkah laku tersebut. **Lingkari 2** apabila siswa jarang menampilkan tingkah laku tersebut. Dan **lingkari 1** apabila siswa tidak pernah menunjukan tingkah laku tersebut.

- I. Ekspresi Tubuh
 - 4 3 2 1 A. Berdiri tegak dan menghadap ke hadirin
 - 4 3 2 1 B. Merubah ekspresi wajah yang mengikuti dengan perubahan nada bicara pada saat presentasi
 - 4 3 2 1 C. Menjaga kontak mata dengan para hadirin

RUBRIK

	Analistik	Holistik
Generik	1	2
Task Spesifik	3	4

RUBRIK


Tipe Scoring Rubrics	Definisi	Kelebihan	Kelemahan
Holistik atau Analytis : Satu atau Lebih Penilaian?			
Analistik	Setiap kriteria penilaian (dimensi, trait) dievaluasi secara terpisah	<ul style="list-style-type: none"> Memberikan gambaran diagnostik kepada guru Menyediakan feedback formatif bagi siswa Lebih mudah dihubungkan dengan pengajaran dibandingkan rubrik holistik Baik untuk tes formatif dan bisa diadaptasikan untuk asesmen sumatif dan dapat dikombinasikan jika dibutuhkan 	<ul style="list-style-type: none"> Membutuhkan waktu penskoran yang lebih lama dibandingkan rubrik holistik Membutuhkan waktu lebih lama untuk mencapai inter-rater reliability.
Holistik	Semua kriteria (dimensi dan trait) dievaluasi secara bersamaan	<ul style="list-style-type: none"> Lebih cepat dalam penskoran daripada rubrik analitis Membutuhkan waktu yang lebih singkat untuk mencapai inter-rater reliability Baik untuk asesmen sumatif 	<ul style="list-style-type: none"> Nilai keseluruhan tidak dapat memberikan informasi yang cukup untuk melakukan peningkatan Tidak baik untuk asesmen formatif

RUBRIK

Deskripsi Hasil Belajar : Generik atau Task-Spesifik?			
Generik	Deskripsi tugas mencakup karakteristik yang diperlukan untuk melakukan tugas tersebut (misalnya, menulis, pemecahan masalah)	<ul style="list-style-type: none"> Bisa didiskusikan dengan siswa secara jelas untuk menghubungkan tes dengan pengajaran Rubrik yang sama dapat digunakan untuk beberapa tugas Memacu belajar siswa dengan mengarahkan siswa melihat bahwa belajar dengan baik lebih penting daripada nilai satu tugas saja. Mendukung self evaluation siswa Siswa dapat membantu dalam pembuatan rubrik generik 	<ul style="list-style-type: none"> Reliabilitas yang lebih rendah dibandingkan rubrik task-spesifik

RUBRIK

Deskripsi Hasil Belajar : Generik atau <i>Task-Specific</i> ?			
<i>Task-Specific</i>	Deskripsi tugas mengarah pada aspek spesifik dari tugas (misalnya memberi jawaban, menspesifikasi kesimpulan)	<ul style="list-style-type: none"> Guru terkadang mengatakan bahwa metode ini membuat penilaian menjadi lebih mudah Mebutuhkan waktu lebih sedikit untuk mencapai inter-rater reliability 	<ul style="list-style-type: none"> Mebutuhkan latihan untuk dapat diaplikasikan dengan baik



Contoh-Contoh Rubrik

Kegunaan rubrik

- ▷ *Fairness*
- ▷ Reliabilitas



Terima kasih

Pertanyaan?

www.facebook.com/acer.edu.au.indonesia