



ASSEMENT STUDENT



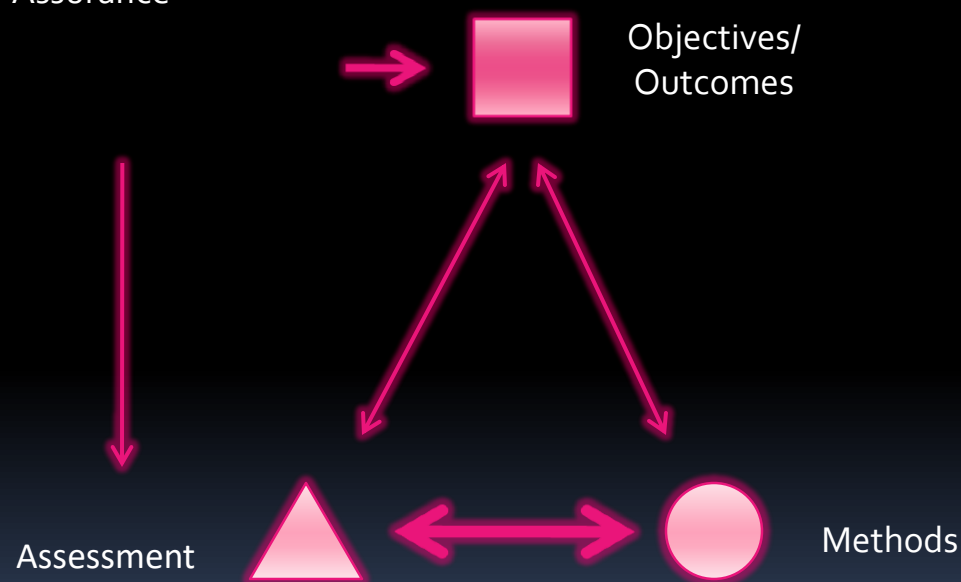
**Dr. Hj. Ermawati., SpOG(K)
Bagian Obstetri dan Ginekologi
FK UNAND / RSUP Dr. M. DJAMIL
PADANG**

DEFENISI

- Penilaian berbasis tempat kerja: menilai kompetensi klinis peserta didik dengan pasien nyata / lingkungan klinis kerja.
 - Penilaian autentik untuk penilaian kinerja-pasien berbasis kehidupan nyata
 - Interaksi yang diamati dengan pasien
 - Rekaman interaksi dengan pasien
 - Umpan balik perkembangan

Membingkai Elemen: Paradigma Pendidikan

Evaluation/Quality
Assurance



Penilaian berbasis
tempat kerja

- Hubungkan pengajaran, pembelajaran, dan penilaian
- Dapat menilai dimensi 'apakah' dari piramida Miller

PROSES BERKELANJUTAN





PENILAIAN PENDORONG PEMBELAJARAN

TUJUAN

Kurangnya umpan balik adalah kesenjangan besar dalam pendidikan kedokteran secara umum dan;

Umpan balik yang konstruktif diperlukan untuk meningkatkan pembelajaran peserta dan kinerja klinis

Penilaian berbasis tempat kerja berfokus pada atribut-atribut penting daripada apa yang paling mudah untuk dinilai

Atribut profesional yang rumit sulit untuk dinilai menggunakan metode penilaian standar (ujian tertulis atau OSCE) — lebih baik dinilai dalam situasi tempat kerja

Belajar yang paling kuat saat itu adalah 'otentik'

KERANGKA KERJA



PIRAMIDA MILLER



Based on work by Miller GE. The assessment of clinical skills/competence/performance: Acad.Med. 1990;65(9); 63-67
Adapted by Drs. R. Mehay & R. Burns. UK (jan 2009)

Metode Penilaian

- Latihan Evaluasi Klinik Mini (mini-CEX)
- Observasi Langsung Keterampilan Prosedural (DOPS)
- Umpan Balik Multi-Sumber (MSF)
- Diskusi Berbasis Kasus (CbD)



KAPAN PENILAIAN DILAKUKAN?

Penilaian formatif

Harus bebas dari ancaman, karena tujuannya adalah untuk membuat siswa mengungkapkan kekuatan dan kelemahan mereka daripada menyamarkannya - memetakan kemajuan siswa melalui program pembelajaran tertentu - tidak lulus / gagal

Formatif adalah umpan balik → peluang untuk mendapatkan umpan balik pada pengetahuan atau kinerja selalu dihargai oleh siswa

Umpan balik berpusat pada siswa dan tidak fokus pada peringkat siswa dalam grup tertentu

Penilaian sumatif

mengukur apakah tingkat kinerja tertentu telah tercapai pada suatu titik waktu tertentu (misalnya, penyelesaian rotasi yang berhasil).

Harus diinformasikan sepenuhnya

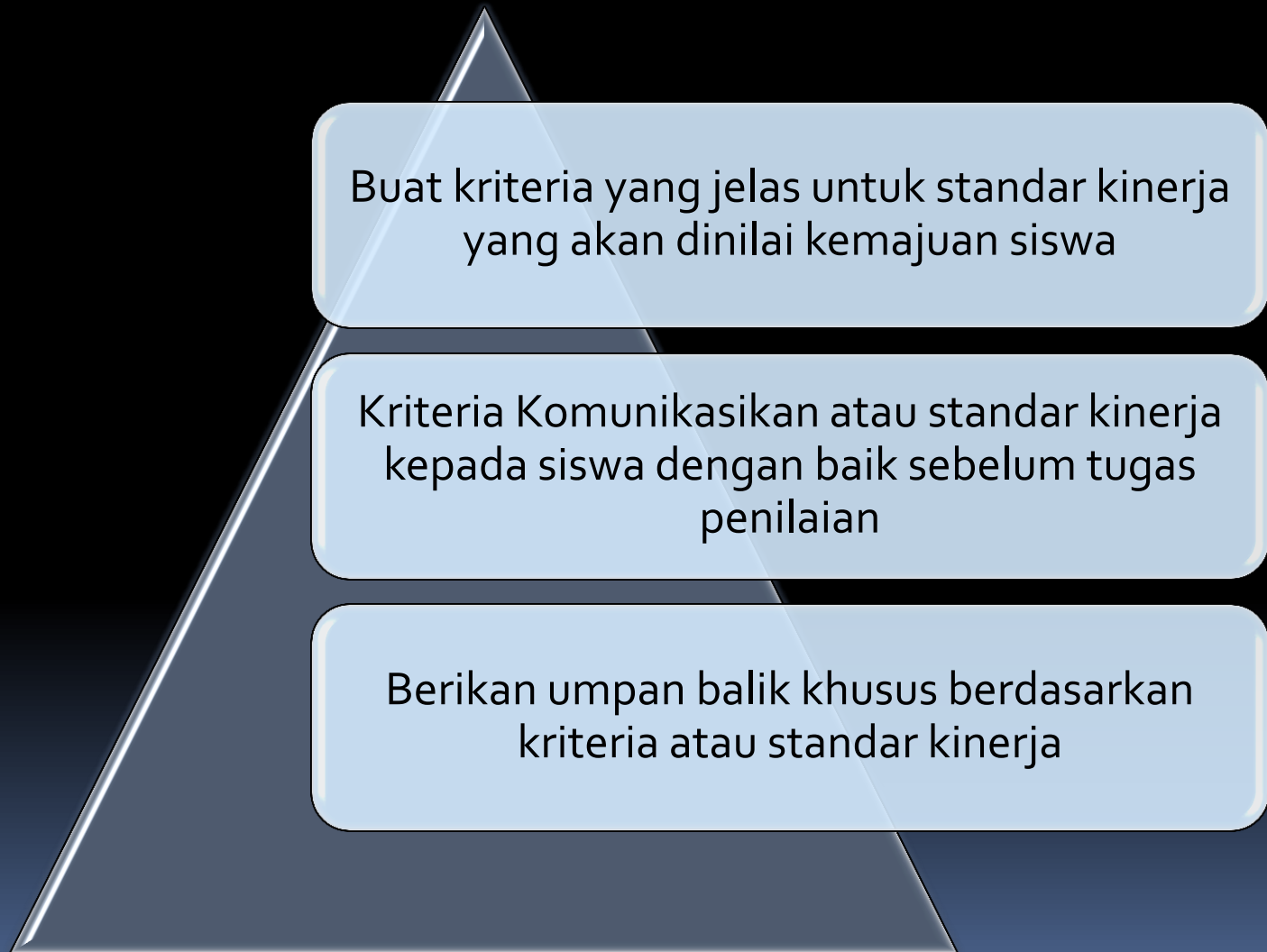
- **Adil**
- **Berdasarkan Kriteria**
- **Metode Penilaian**
- **Diberikan Bobot Pada Masing-Masing Komponen**

mengumpulkan informasi dari semua sumber yang relevan dan untuk menentukan apakah tujuan kursus telah terpenuhi secara memadai



BAGAIMANA CARANYA ?

Hal yang perlu dipertimbangkan



Norma dan Kriteria Referensi

Menghubungkan kinerja mentah siswa dengan standar sehingga perbandingan atau peringkat dapat ditarik.

- Berbasis norma: tanda lulus ditentukan oleh kinerja peserta pelatihan dibandingkan dengan yang lain (mis: 45% teratas siswa = lulus → berguna misalnya ketika ada tempat terbatas pada program dan oleh karena itu perlu didiskriminasikan untuk menemukan peserta yang terbaik.
- Kriteria berdasarkan: tanda lulus ditentukan terhadap pencapaian standar yang telah ditentukan (mis: skor > 65 = lulus)

Rumus Utilitas Van der Vleuten

$$U = aV \times bR \times cL \times dA \times eC$$

- V= validity
- R= reliability
- L=impact on learning
 - A=acceptance
 - C=cost

Lynch et al., 2004 – 88 different types of assessment

PENILAIAN

KEANDALAN

- Apakah tes secara konsisten mengukur apa yang seharusnya diukur?
- Jenis-jenis keandalan:
 - Inter-rater (konsistensi over-penilai)
 - Uji-tes ulang (konsistensi dari waktu ke waktu)
 - Konsistensi internal (atas berbagai item / bentuk)

VALIDITY

- Apakah kita mengukur apa yang seharusnya kita ukur
- Gunakan instrumen yang sesuai untuk pengetahuan, keterampilan, atau sikap yang Anda uji
- Jenis utama validitas
 - Menghadapi
 - Konten
 - membangun

Faktor-Faktor Yang Mempengaruhi Reliability

Test length

Tes yang lebih lama memberikan skor yang lebih reliability

Group homogeneity

Semakin heterogen kelompok, semakin tinggi reliability

Objectivity of scoring

Semakin banyak tujuan penilaian, semakin tinggi reliability

Reliability of an oral examination

(Swanson, 1987)

Testing Time in Hours	Number of Cases	Same Examiner for All Case	Same Examiner for Each Case	Two New Examiners for Each Case
1	2	0.31	0.50	0.61
2	4	0.47	0.69	0.76
4	8	0.47	0.82	0.86
8	12	0.48	0.90	0.93

Reliability of a number of measures

Testi ng time in hours	MCQ ¹	Case Base d Short Essay ²	PMP ¹	Oral Exam ³	Long Case ⁴	OSCE ⁵	Mini CEX ⁶	Practi ce Video Asses sment ⁷	Incog nito SPs ⁸
1	0.62	0.68	0.36	0.50	0.60	0.47	0.73	0.62	0.61
2	0.76	0.73	0.53	0.69	0.75	0.64	0.84	0.76	0.76
4	0.93	0.84	0.69	0.82	0.86	0.78	0.92	0.93	0.92
8	0.93	0.82	0.82	0.90	0.90	0.88	0.96	0.93	0.93

1. Swanson, 1987
2. Norcini et al, 1985
3. Stalenhoef-Halling et al., 1990

4. Wass et al., 2001
Petrusa, 2002
Norcini et al 1999

7. Ram et al., 1999
8. Gorter, 2002

Table 1. Perbandingan Alat Penilaian Yang Umum Digunakan

Tool	Reliability	Validity	Feasibility	Strengths	Limitations
MCQ	+	+	=	Universal familiarity Easy to score Extensive research	Random guessing Difficult to write well Trivial content
Essay	-	+/-	+/-	Assess problem solving Easy to compose	Difficult to score Inefficient
SAQ	+	+	+	Easy to write	Difficult to score
Oral	+/-	+	+/-	Assess higher order reasoning Trainee can clarify answers	Time intensive for faculty Examiner cueing Standardization issues
OSCE	+	+	-	Fidelity Scheduled Assess behaviour in critical events	Deconstruct patient-physician dynamic Potential scenarios limited Expensive to run
ITER	-	+	+	Assess 'soft domains Flexible Longitudinal assessment	Student's characteristic influence assessment Poor discrimination between domains Reliability issues
MSF	+/-	+	-	Assess soft domains Interprofessional Inter raters	Time intensive May affect patient- physician dynamic Observer training
Portfolios	-	+/-	-	Self reflection Promotes remediation Longitudinal assessment	Time intensive Intentional misrepresentation Time consuming
Logs	NA	+/-	+	Increases exposure to technical skills tracking	Assumes repetition equals competence Time consuming



SEBERAPA SERING ?



Mini-CEX (Clinical Examination Exercise)

Penilaian

- Sekitar 4 pertemuan cukup untuk mencapai 95% CI <1 dan 12-14 diperlukan untuk mencapai koefisien reabilitas 0,8 (Norcini et al 1995, 2003, Homboe et al 2003).
- Davies et al (2009) → interval kepercayaan 95% untuk sejumlah DOPS yang diberikan adalah sebagai berikut: 4 prosedur 0,59; 6 (0,48); 8 (0,42); 12 (0,34)

Validity

- Wilkinson et al (2008) → 91% peserta pelatihan merasa bahwa DOPS adalah metode yang adil untuk menilai keterampilan prosedural
- Davies et al (2009) → Skor untuk DOPS meningkat antara paruh pertama dan kedua tahun ini, menunjukkan validitas



CbD (Case based Discussion)

Penilaian

- Davies et al (2009) → interval pada skala enam poin untuk sejumlah CbD yang diberikan adalah sebagai berikut: 4 kasus 0,55; 6 kasus (0,45); 8 kasus (0,39); 12 kasus (0,32)

Validity

- Davies et al (2009) → Skor untuk CbD meningkat antara paruh pertama dan kedua tahun ini, menunjukkan validitas

PERSYARATAN MINIMUM

	Frekuensi Penilaian	Jumlah Penilaian	Jumlah Penilaian Per Penilai	Persyaratan Waktu
Mini CEX	4 per year	2	2	4 hours per year per SpR
DOPS	6 over 4 year	2	3	1 hour per year per SpR
MSF	2 in 4 years	12-20 raters 1 collater	1 or 2	One hour per year per SpR



PENILAIAN

- Kemajuan melalui masing-masing bidang kompetensi dapat dicatat sebagai salah satu dari yang berikut :
 - (I) Insufficient evidence
 - (N) Needs further development
 - (C) Competence
 - (E) Excellent
- 



KESIMPULAN

- Tidak mudah untuk menetapkan standar
 - Diperlukan pertimbangan profesional ahli
 - Pengambilan sampel luas dari berbagai sumber: berbagai alat, konteks, kasus, dan penilai
 - Umpan balik kepada peserta pelatihan adalah wajib
- 



TERIMAKASIH