# Assessment as an optimization problem: Work-based assessment and programmatic assessment

UCSF, Developing Medical Educators of the 21st Century

February 25-27, San Francisco, USA

Cees van der Vleuten

Maastricht University, The Netherlands

www.ceesvandervleuten.com

# Utility formula (Van der Vleuten, 1996)

R = Reliability
V = Validity
E = Educational impact
A = Acceptability
C = Cost

U = Utility
W = weight

$$U = R_{W_r} \times V_{W_v} \times E_{W_e} \times A_{W_a} \times C_{W_c}$$


U=VxRxEIxPxAxCE


ROYAL COLLEGE
COLLÈGE ROYAL

# Utility formula consequences

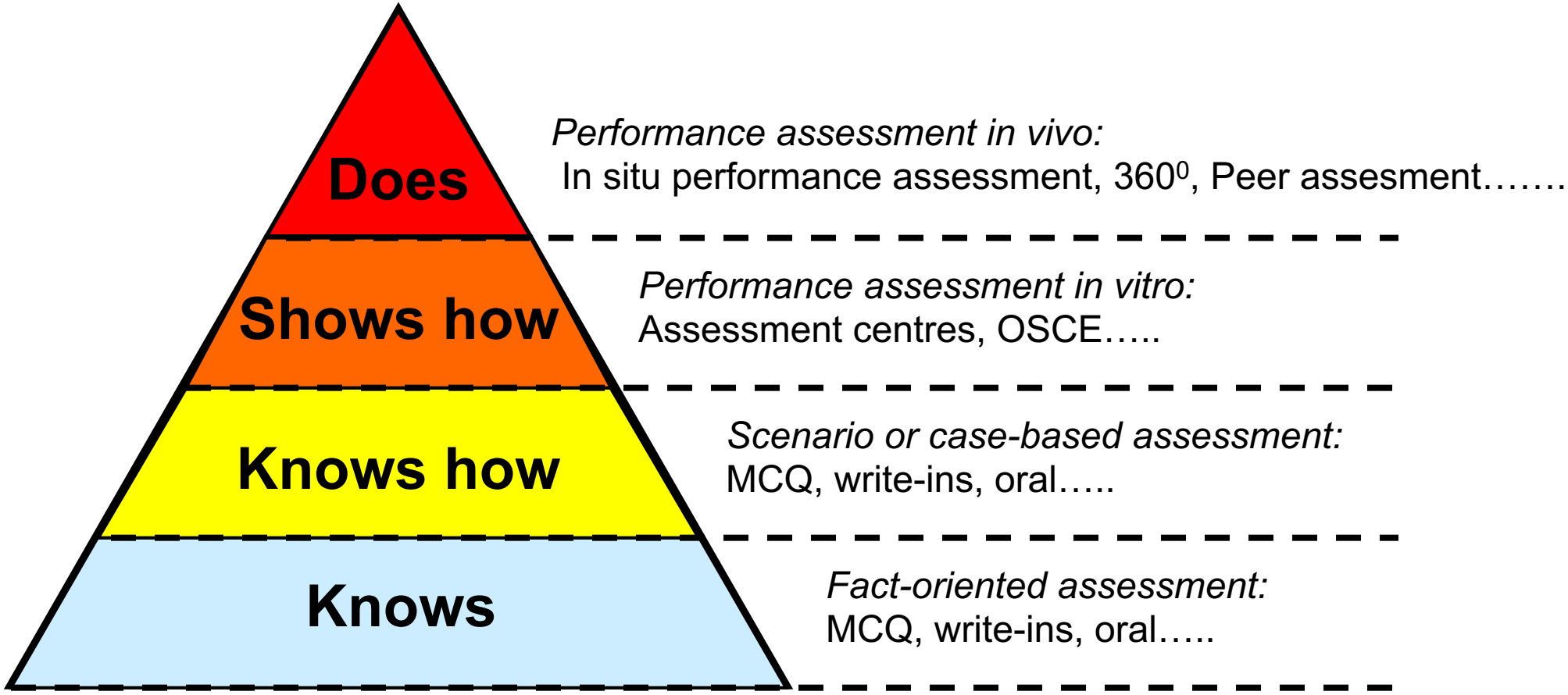
- Any assessment is a compromise
- You can’t have it all in one method
- Assessment is really an optimization problem!

# Method reliability as a function of testing time[1]

| Testing Time in Hours | MCQ | Case-Based Short Essay | PMP | Oral Exam | Long Case | OSCE | Mini CEX | Practice Video Assess-ment | In-cognito SPs |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.62 | 0.68 | 0.36 | 0.50 | 0.60 | 0.54 | 0.73 | 0.62 | 0.61 |
| 2 | 0.77 | 0.81 | 0.53 | 0.67 | 0.75 | 0.70 | 0.84 | 0.77 | 0.76 |
| 4 | 0.87 | 0.89 | 0.69 | 0.80 | 0.86 | 0.82 | 0.92 | 0.87 | 0.86 |
| 8 | 0.93 | 0.94 | 0.82 | 0.89 | 0.92 | 0.90 | 0.96 | 0.93 | 0.93 |

[1]Based on table 1 in: Van Der Vleuten, C. P., & Schuwirth, L. W. (2005). Assessing professional competence: from methods to programmes. *Medical Education*, *39*(3), 309-317.

# The way we climbed……



# Competency-frameworks

## CanMeds

- Medical expert
- Communicator
- Collaborator
- Manager
- Health advocate
- Scholar
- Professional

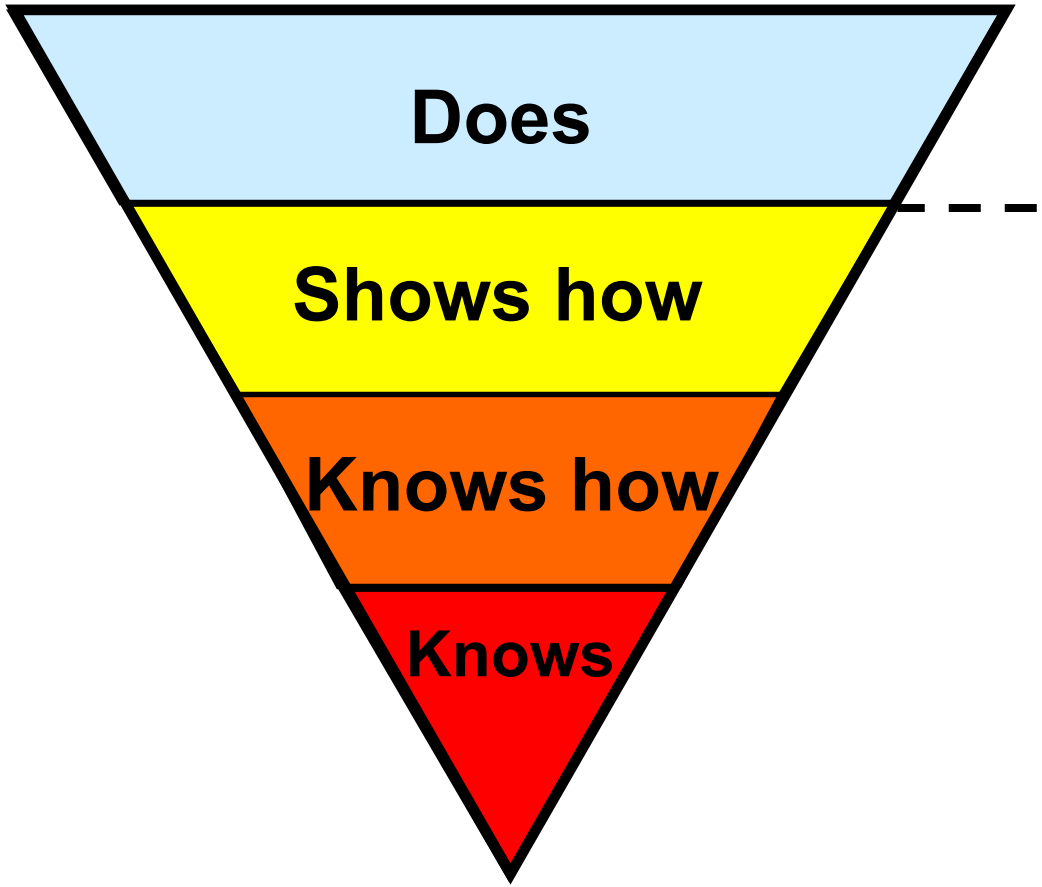
## ACGME

- Medical knowledge
- Patient care
- Practice-based learning & improvement
- Interpersonal and communication skills
- Professionalism
- Systems-based practice

## GMC

- Good clinical care
- Relationships with patients and families
- Working with colleagues
- Managing the workplace
- Social responsibility and accountability
- Professionalism

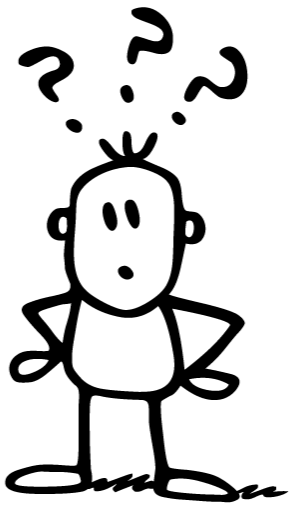
# Assessing complex behavioural skills




Unstandardized assessment

Standardized assessment

Professional Judgment through Observation and Interpretation



*Medical Teacher, Vol. 26, No. 4, 2004, pp. 305–312*

Taylor&Francis healthsciences

# Effectiveness of clinical rotations as a learning environment for achieving competences

H.E.M. DAELMANS[1], R.J.I. HOOGENBOOM[2], A.J.M. DONKER[3], A.J.J.A. SCHERPBIER[4], C.D.A. STEHOUWER[3] & C.P.M. VAN DER VLEUTEN[2]

*[1]Vrije Universiteit Medical Centre, Skills Training Department, Amsterdam, The Netherlands; [2]Department of Educational Development and Research, Maastricht University, The Netherlands; [3]Vrije Universiteit Medical Centre, Department of Internal Medicine, Amsterdam, The Netherlands; [4]Faculty of Medicine, Institute for Medical Education, Maastricht University, The Netherlands*

SUMMARY *Competences are becoming more and more prominent in undergraduate medical education. Workplace learning is regarded as crucial in competence learning. Assuming that effective learning depends on adequate supervision, feedback and assessment, the authors studied the occurrence of these three variables in relation to a set of clinical competences. They surveyed students at the end of their rotation in surgery, internal medicine or paediatrics asking them to indicate for each competence how often they had received observed and unobserved supervision, the seniority of the person who provided most of their feedback, and whether the competence was addressed in formal assessments. Supervision was found to be scarce and mostly unobserved. Senior staff did not provide much feedback, and assessment mostly targeted patient-related competences. For all variables, the variation between students exceeded that between disciplines. We conclude that conditions for adequate workplace learning are poorly met and that*

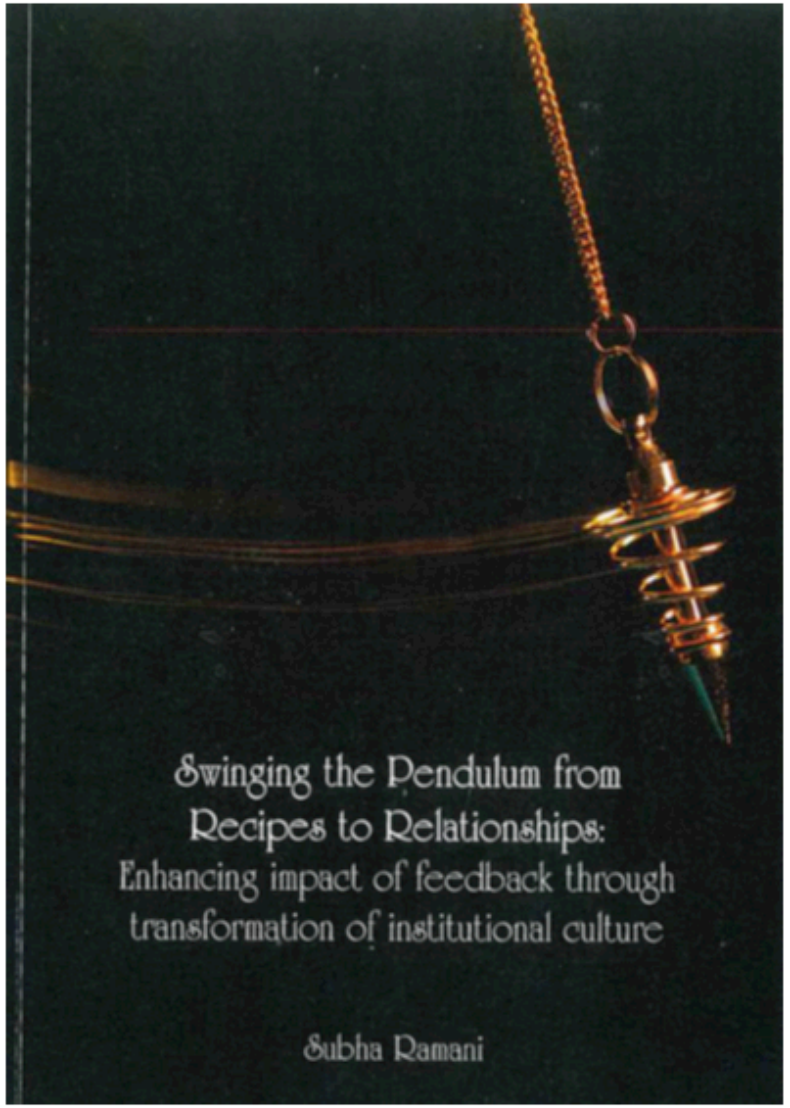
Irby, 1995; Remmen *et al.*, 2000, Van Der Hem-Stokroos *et al.*, 2001). Kilminster *et al.* (2002) found supervision to be an essential factor and others have pointed out that feedback and assessment play a major role in workplace learning (Newble & Jaeger, 1983; Stillman *et al.*, 1991; Irby, 1994). Unfortunately, these three factors occur rather infrequently, with feedback often being provided by professionals who are not fully qualified and assessment lacking sufficient congruence with the intended objectives (Remmen *et al.*, 1998; Kassebaum & Eaglen, 1999). Most of the studies referred to focus on separate procedural and clinical skills and did not address competences as learning goals. We wanted to investigate the effectiveness of the workplace as a learning environment for achieving competences. Our investigation focused on clerkship learning, because clinical rotations are the prominent form of workplace learning in

# Research in work-based learning points to:

- Limited direct observation, feedback and monitoring
- Limited reflection
- Too many tasks with little learning value
- Limitations of observational learning
- Transitions hampering learning
- Near adverse events (particularly postgraduate)
- Learning climate variability
- Limited attention to generic competencies.

# Classes of work-based assessment tools

1. Direct observation: Single encounter methods
   - Mini-CEX
   - DOPS, OSATS
   - P-MEX
   - …….
2. Global performance measures
   - Multi-Source Feedback (MSF or 360)
   - In-training Evaluation Reports (ITER)
3. Aggregation and reflection measures
   - Logbook
   - Portfolio


Swinging the Pendulum from
Recipes to Relationships:
Enhancing impact of feedback through
transformation of institutional culture
Subha Ramani

# What did we learn from work-based assessment?

- The users of instruments are more important than the instruments themselves
- Feedback is a dialogue (with follow-up)
- (Longitudinal) Relationships are very important
- Narrative feedback is much more powerful than numbers or grades
- Raters can hardly be calibrated
- A very summative approach doesn’t work; learning should be promoted.


# HIDDEN IN PLAIN SIGHT

*The untapped potential of written assessment comments*

Shiphra Ginsburg

| Quantitative/ Psychometric approach | | Qualitative/ Interpretivist approach |
|---|---|---|
| Scores/grades | ⟷ | Words/narratives |
| Statistical computation | ⟷ | Professional judgment |
| Cut-off scores | ⟷ | Performance standards, EPAs, milestones |
| Algorithmic | ⟷ | Judgmental/triangulation of information |
| Bias | ⟷ | Perspective |
| True score | ⟷ | Multiple perspectives |
| Reliability | ⟷ | Saturation of information |
| Validity | ⟷ | Trustworthiness/credibility |

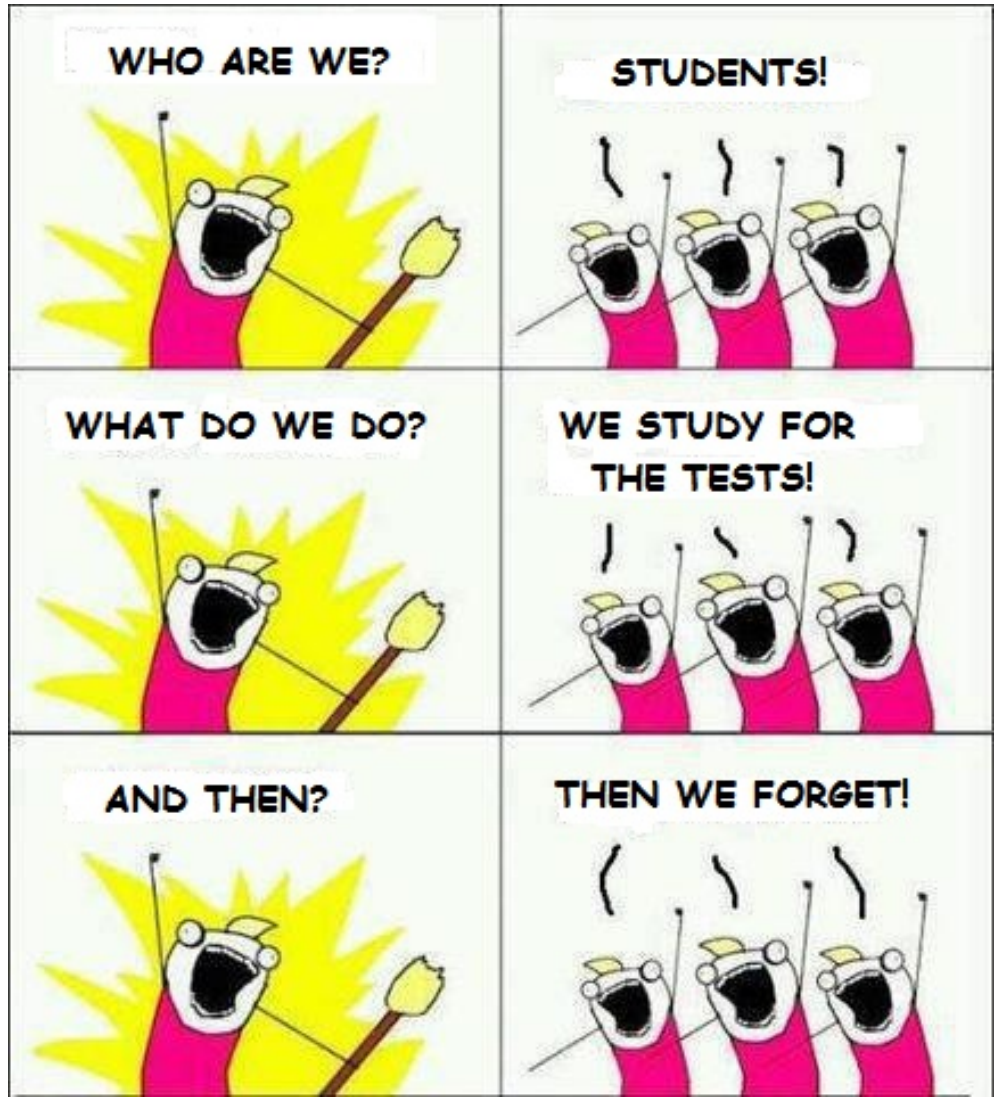
# THE PRE-ASSESSMENT LEARNING EFFECTS OF CONSEQUENTIAL ASSESSMENT

## Modelling how the Examination Game is Played

François J Cilliers

# Assessment driving learning …….often bad news again!

- Impact on learning is often very negative (Cilliers et al, 2011; 2012; Al-Kadri et al, 2012)
  - Poor learning styles (cram & dump)
  - Grade culture (grade hunting, competitiveness)
  - Grade inflation (e.g. in the workplace)
- A lot of REDUCTIONISM!
  - Little feedback (grade is poorest form of feedback one can get; Shute 2008)
  - Non-alignment with curricular goals
  - Non-meaningful aggregation of assessment information
  - Few longitudinal elements
  - Tick-box exercises (OSCEs, logbooks, work-based assessment).


WHO ARE WE?
STUDENTS!
WHAT DO WE DO?
WE STUDY FOR THE TESTS!
AND THEN?
THEN WE FORGET!


feedback has been unclear. This study demonstrates the benefits of moving away from a behaviouristic approach to assessment, based on punishment and rewards. It reveals the potential benefits of applying three constructivist principles to assessment: authenticity, empowering students with a more active role and gradual descaffolding to enable transformation towards a learning orientation.

PLOS | ONE

RESEARCH ARTICLE

**Table 2. The five phases and two overall conditions in GP residents' learning process of communication skills.**

| Cognitive and Emotional space | | | | |
|---|---|---|---|---|
| **Phase 1** | **Phase 2** | **Phase 3** | **Phase 4** | **Phase 5** |
| Confrontation with the effect of a behavior | Becoming conscious of own behavior | Searching and receiving alternative behavior | Personalization of new behavior | Internalization and clinical integration |
| **Safety** | | | | |

doi:10.1371/journal.pone.0125958.t002



☯ These authors contributed equally to this work.
‡These authors also contributed equally to this work.
* v.vdeertwegh@maastrichtuniversity.nl

CrossMark
click for updates

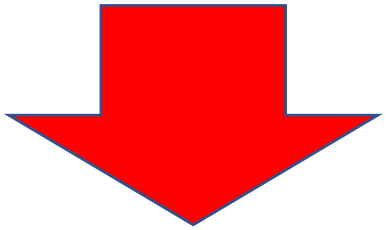
# Abstract

# My conclusions so far

- Individual assessments  are not really fit for (pass/fail) decision-making
- Our summative culture hinders a learning culture
- Our classic approach to assessment belongs to an outdated model of learning
- We need a radical change in which we need to rethink when to optimize what!

# New pathway suggestions

- Stop optimizing everything in a single assessment
- Focus on feedback, reflection and mentoring
- Make high stake decisions only when you have sufficient data.

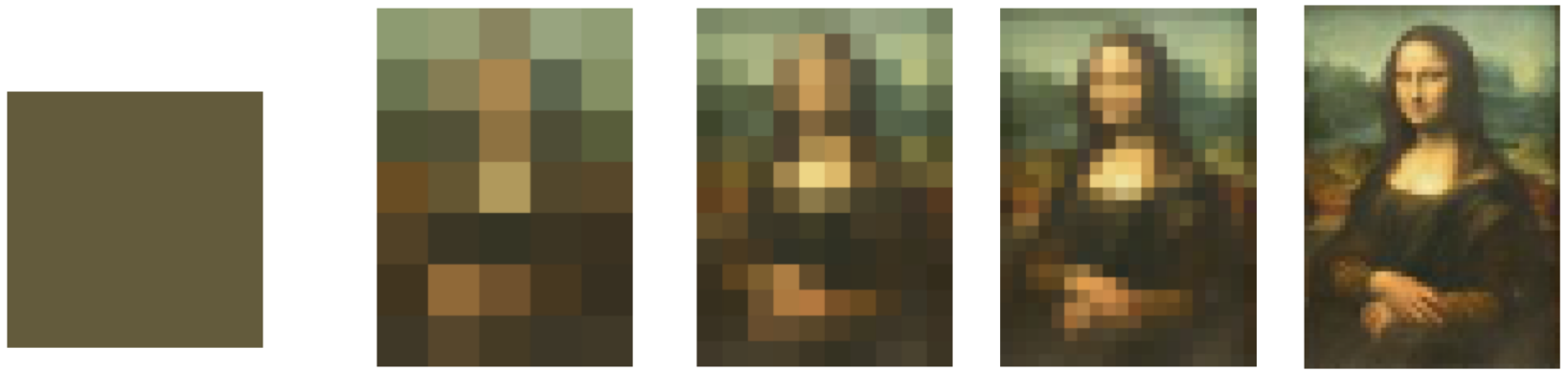


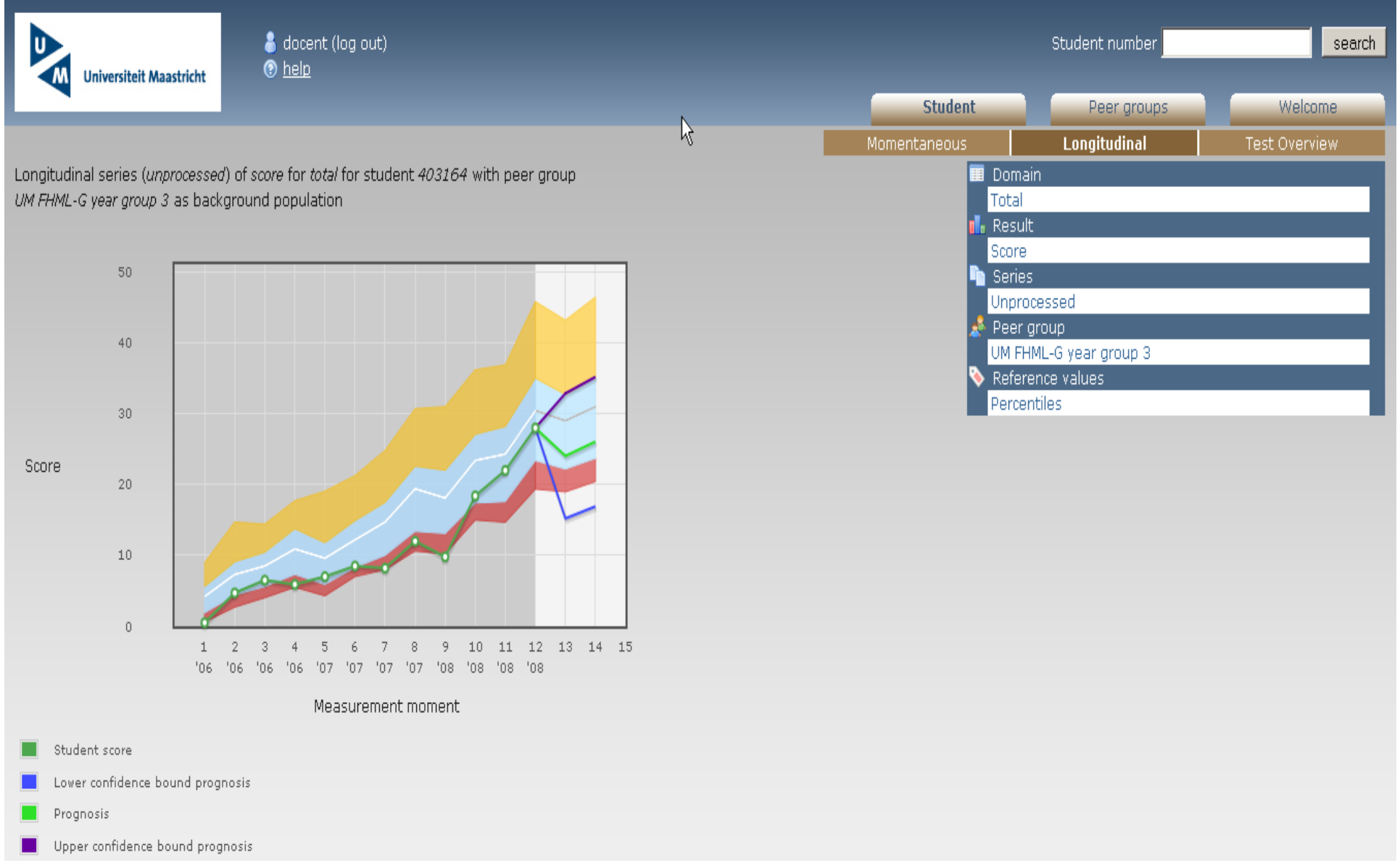
# Programmatic assessment

# Ground rules in programmatic assessment

- No pass/fail decision on a single data point (single assessment), but feedback
- There is mix of methods of assessment
- The number of data points is proportionally related to the stakes of a decision
- To promote feedback use and self-directed learning learners are coached/mentored
- High stake decisions are based on professional judgment of a group of experts or committee.

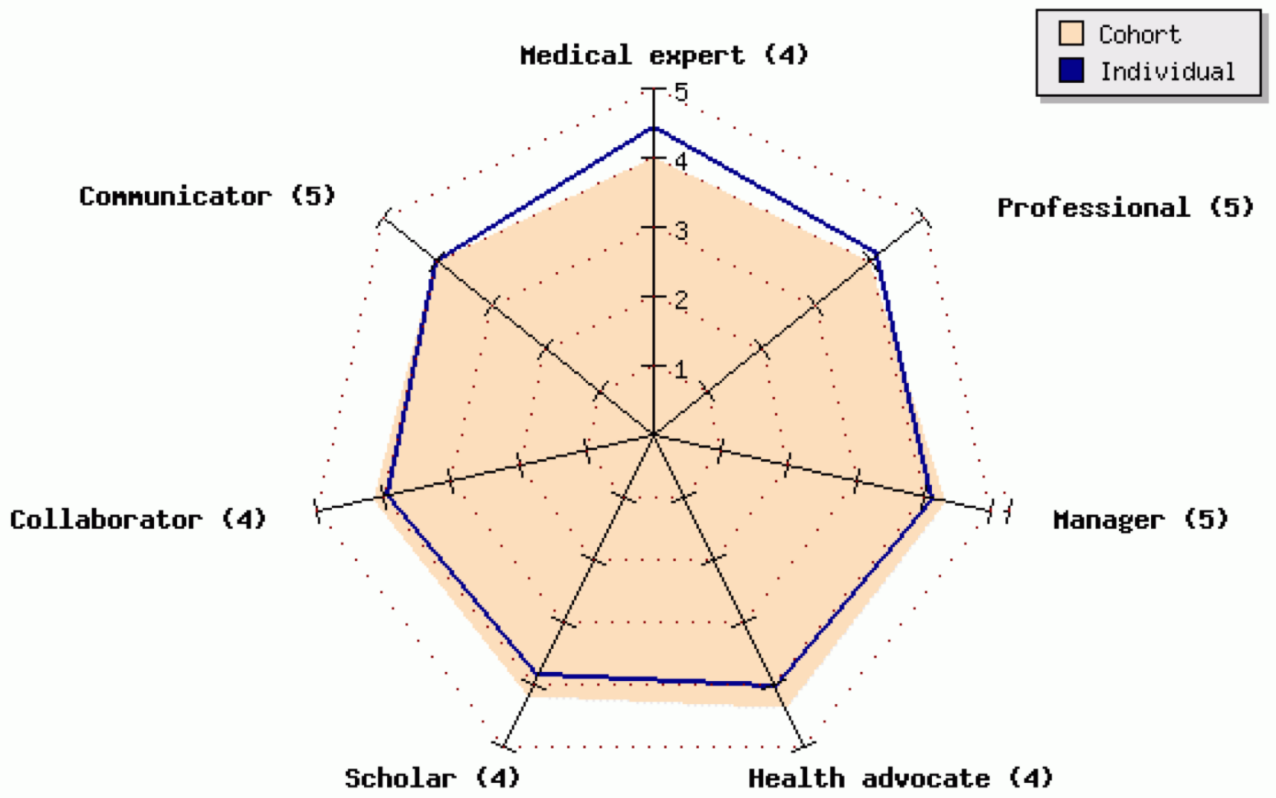
# Assessment information as pixels





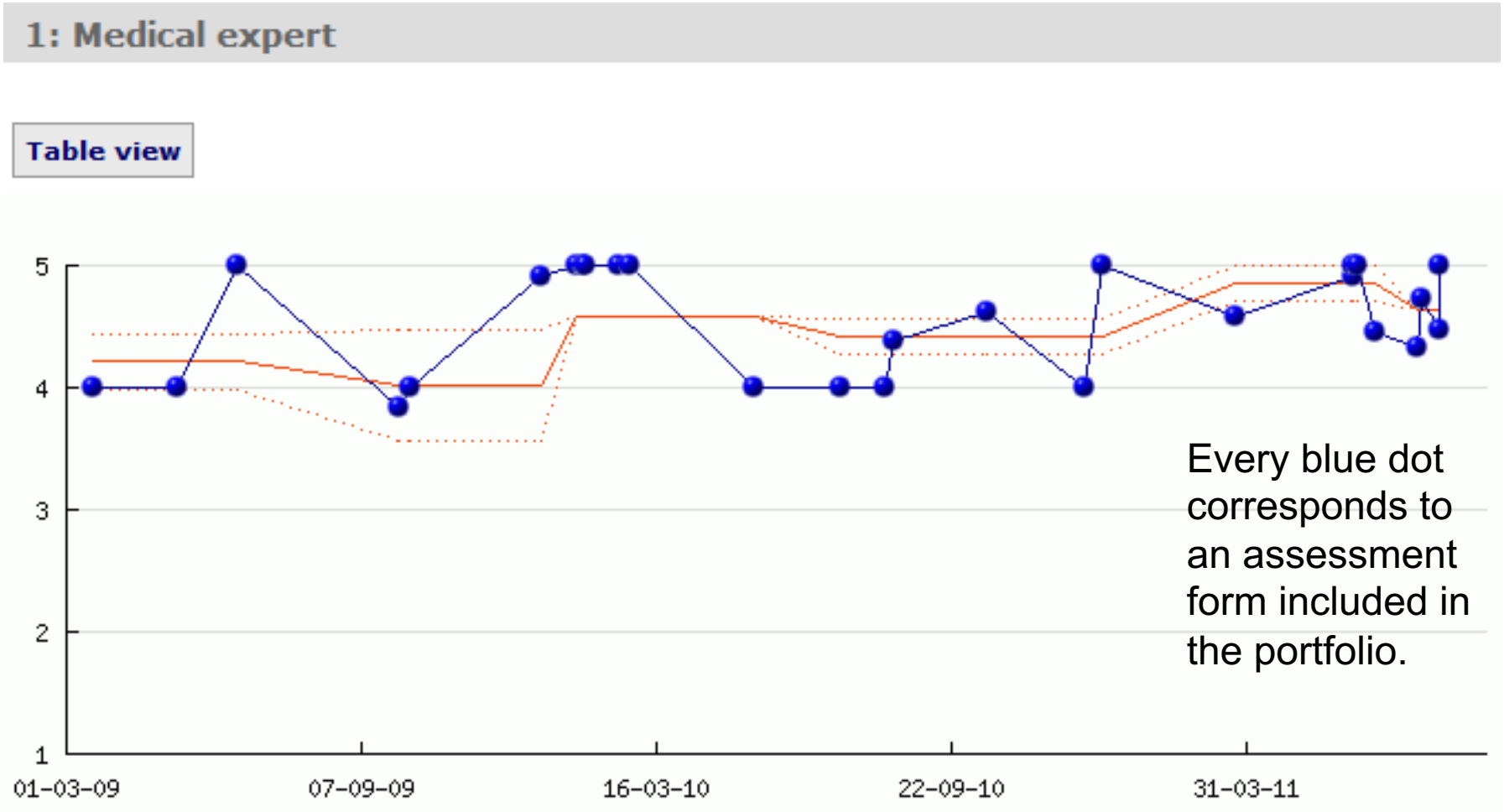

Longitudinal total test scores across 12 measurement moments and predicted future performance

# Maastricht Electronic portfolio (ePass)




Comparison between the score of the student and the average score of his/her peers.

# Maastricht Electronic portfolio (ePass)



**Narrative feedback**

Feedbacktype: all

Competency: all

| Date | Feedbacktype | Competency | Narrative feedback | Form |
|---|---|---|---|---|
| 06-11-2013 | Improvement | General | don't repeat too much, no irrelevant details<br>Conclusion: antenatal care in pregnancy may be done by a midwife and delivery will be done by a gynecologist, I revise | Mini-CEX-N |
| 06-11-2013 | Strength | General | included all information. | Mini-CEX-N |
| 06-11-2013 | Improvement | General | don't repeat too much, no irrelevant details.<br>Conclusion: antenatal care in pregnancy may be done by a midwife, delivery will be done by a gynecologist, I revise. | Mini-CEX-N |
| 06-11-2013 | Strength | General | included all info. | Mini-CEX-N |
| 18-10-2013 | Improvement | General | more communication with the patient, in this case difficult because of language barrier<br>more communication with supervisor | OSATS |

# Coaching by counselors

- Coaching is essential for successful use of reflective learning skills
- Counselor gives advice/comments (whether asked or not)
- He/she counsels if choices have to be made
- He/she guards and discusses study progress and development of competencies

# Decision-making by committee

- Committee of counselors and externals
- Decision is based on portfolio information & counselor recommendation, competency standards
- Deliberation is proportional to clarity of information
- Decisions are justified when needed; remediation recommendation may be provided

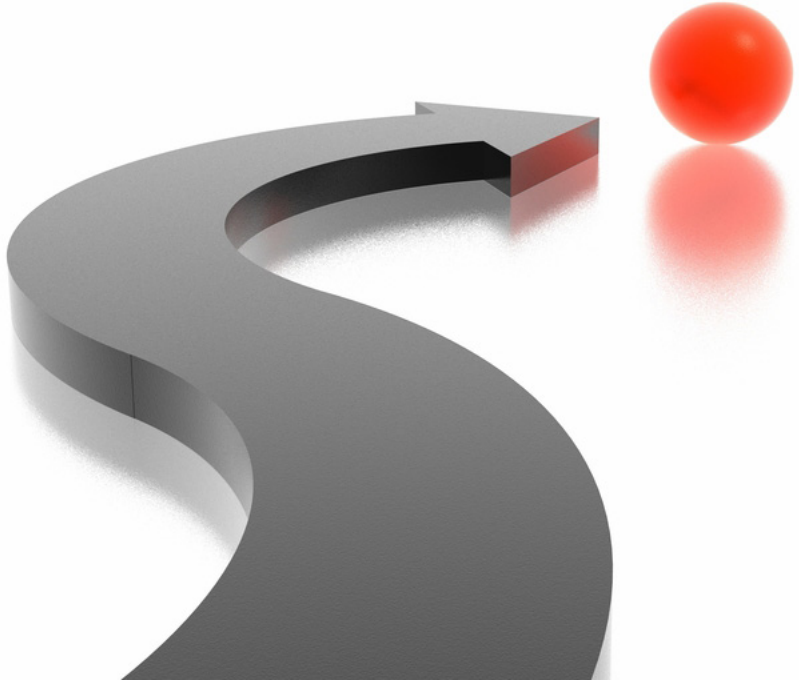

www.ceesvandervleuten.com

# Findings on programmatic assessment so far

- The quality of the implementation defines the success (Harrison et al., 2018)
- Getting high quality feedback is a challenge (Bok et al., 2013)
- Leaners may perceive low stake assessments as high stake, all depending on the learning culture created (Schut et al., 2018)
- Coaching and mentoring is key to the success (Heeneman & Grave, 2017)
- High stake decision-making in competence committees work really well (Oudkerk-Pool et al., 2017, De Jong et al, in preparation).

# Conclusions 1: The way forward

- We should embrace the “subjective” in single encounter work-based assessment
- We need to assess from a “growth” mindset using trusted relationships
- Subjectivity is dealt with through sampling and procedural bias reduction methods (not with standardization or objectification)
- Assessing is knowing when to optimize what; optimizing everything in a single assessment is a dead end..



# Conclusions 2: The way forward

- The programmatic approach to assessment optimizes:
  - The learning function (through information richness and dialogues)
  - The pass/fail decision function (through the combination of rich information)

- Learning should drive assessment!

# Creative Commons License

**Attribution-NonCommercial-Share Alike 3.0 Unported**

**You are free:**

- to Share — to copy, distribute and transmit the work
- to Remix — to adapt the work

**Under the following conditions:**

- **Attribution.** You must give the original authors credit (but not in any way that suggests that they endorse you or your use of the work).
- **Noncommercial**. You may not use this work for commercial purposes.
- **Share Alike**. If you alter, transform, or build upon this work, you may distribute the resulting work only under a license identical to this one.

See http://creativecommons.org/licenses/by-nc-sa/3.0/ for