

A close-up photograph of several interlocking white plastic gears. The gears are arranged in a way that shows their teeth meshing together. The lighting is soft, creating subtle shadows and highlights on the surfaces of the gears.

It is time to let learning drive assessment in PBL

Assessment Seminar, UM

April 13, 2018

Cees van der Vleuten

Department of Educational Development and
Research

www.ceesvandervleuten.com



Overview

- From practice to research
- From research to theory
- From theory to practice
- From practice to research
- Conclusions

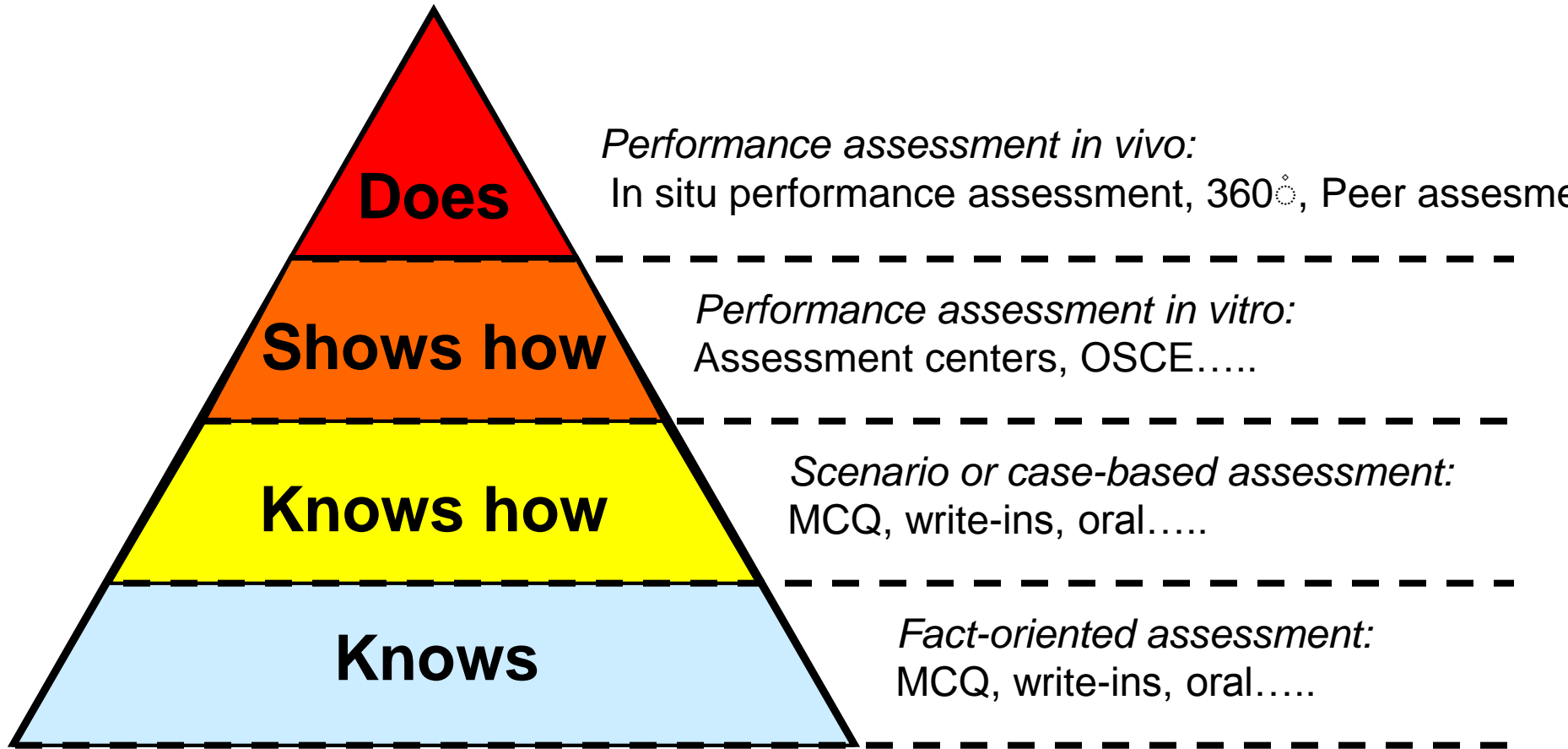


The Toolbox

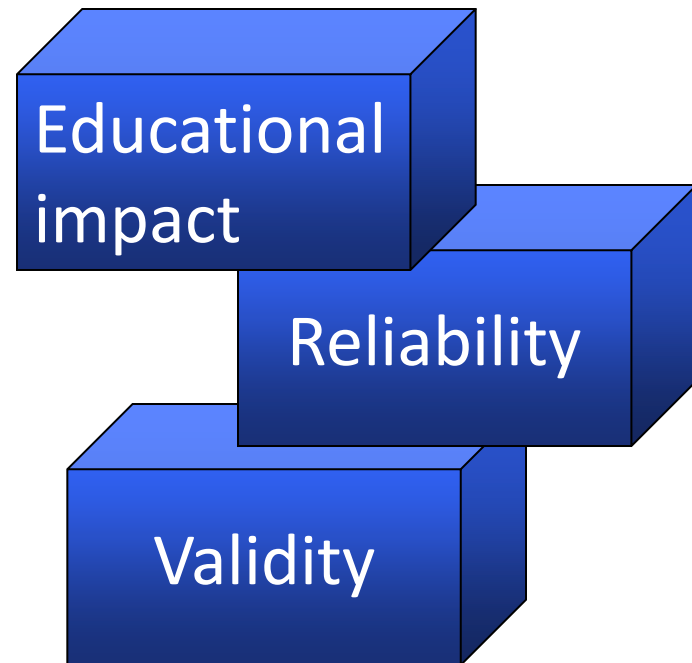
- MCQ, MEQ, OEQ, SIMP, Write-ins, Key Feature, Progress test, PMP, SCT, Viva, Long case, Short case, OSCE, OSPE, DOCEE, SP-based test, Video assessment, MSF, Mini-CEX, DOPS, assessment center, self-assessment, peer assessment, incognito SPs, portfolio.....



The way we climbed.....



Characteristics of instruments



Validity: what are we assessing?

- Curricula have changed from an input orientation to an output orientation
- We went from haphazard learning to integrated learning objectives, to end objectives, and now to (generic) competencies
- We went from teacher oriented programs to learning oriented, self-directed programs

Competency-frameworks



CanMeds

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



ACGME

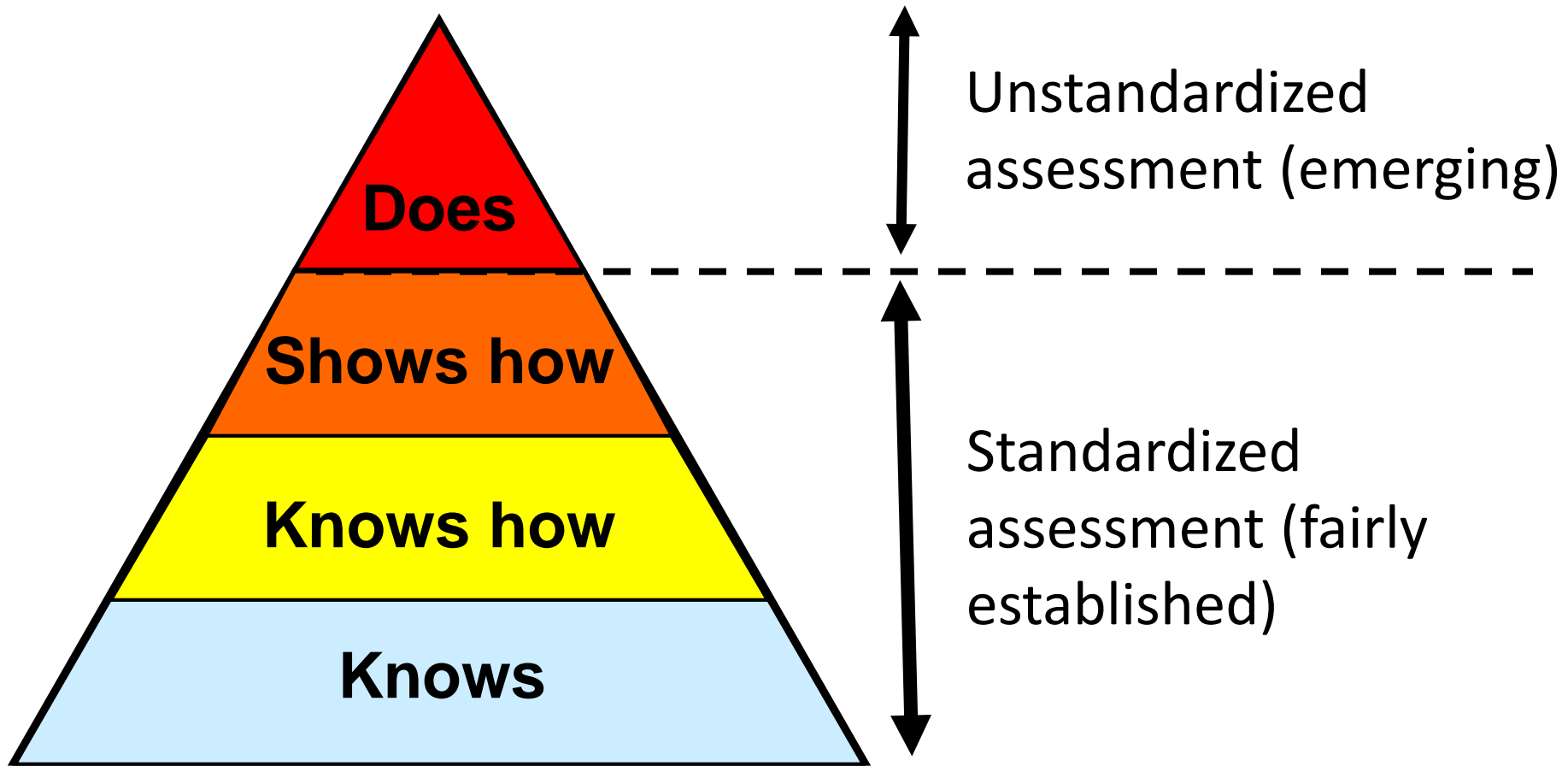
- Medical knowledge
- Patient care
- Practice-based learning & improvement
- Interpersonal and communication skills
- Professionalism
- Systems-based practice



GMC

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

Validity: what are we assessing?



Messages from validity research

- There is no magic bullet; we need a mixture of methods to cover the competency pyramid
- We need BOTH standardized and non-standardized assessment methods
- For standardized assessment quality control around test development and administration is vital
- For unstandardized assessment the users (the people) are vital.



Method reliability as a function of testing time

Testing Time in Hours	MCQ	Case-Based Short Essay	Problem solving Test	Oral Exam	Case-based Oral Exam	Lab Simulation	Observation in Practice	Practice Video Assessment	In-cognito clients
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

Source: Van der Vleuten, C. P., & Schuwirth, L. W. (2005). Assessing professional competence: from methods to programmes. *Medical education*, 39(3), 309-317.

Messages from reliability research

- Acceptable reliability is only achieved with large samples of test elements (contexts, cases) and assessors
- No method is inherently better than any other (that includes the new ones!)
- Objectivity is NOT equal to reliability
- Many subjective judgments are pretty reproducible/reliable.



Educational impact: How does assessment drive learning?

- Relationship is complex (cf. Cilliers, 2011, 2012)
- But impact is often very negative
 - Poor learning styles
 - 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)
 - 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!



Messages learning impact research

- No assessment without (meaningful) feedback
- Narrative feedback has a lot more impact on complex skills than scores
- Provision of feedback is not enough (feedback is a dialogue)
- Longitudinal assessment is needed.



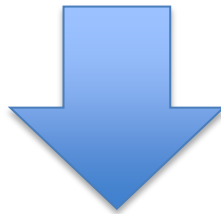
Overview

- From practice to research
- **From research to theory**
- From theory to practice
- From practice to research
- Conclusions



Implications

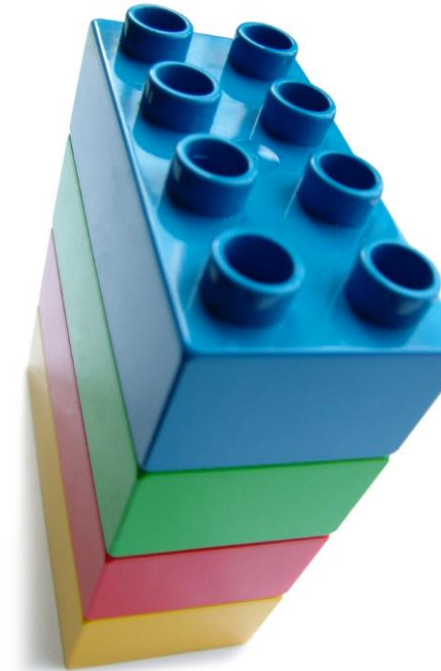
- **Validity:** a multitude of methods needed
- **Reliability:** a lot of (combined) information is needed
- **Learning impact:** assessment should provide (longitudinal) meaningful information for learning



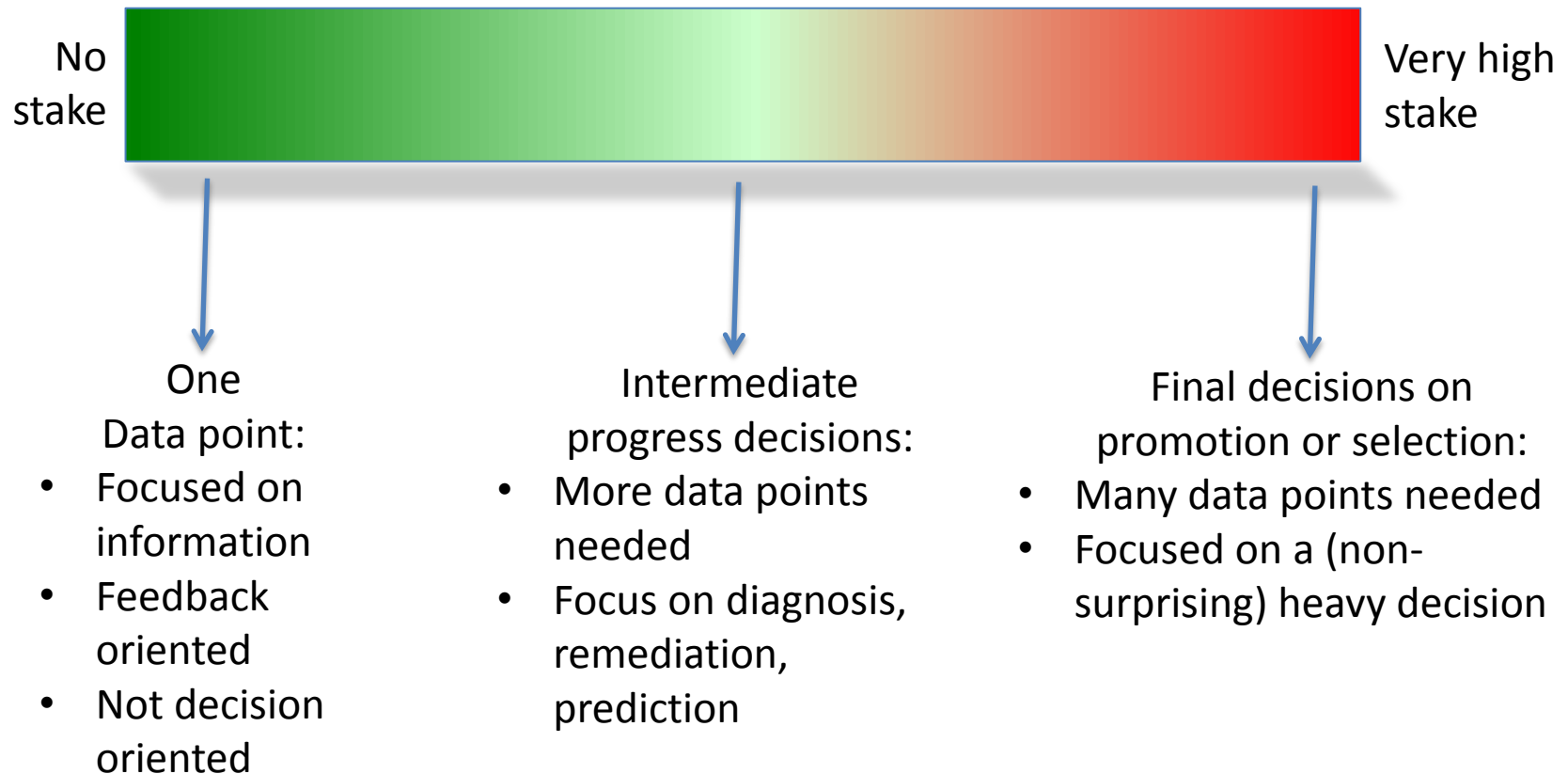
Programmatic assessment

Building blocks for programmatic assessment 1

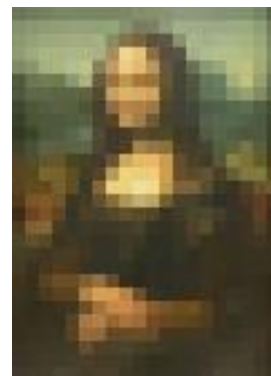
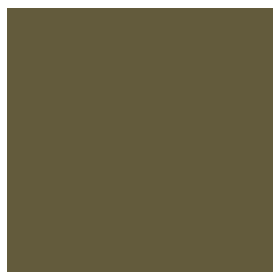
- Every assessment is but one data point (Δ)
- Every data point is optimized for learning
 - Information rich (quantitative, qualitative)
 - Meaningful
 - Variation in format
- Summative versus formative is replaced by continuum of stakes (stakes)
- N data points are proportionally related to the stakes of the decision to be taken.




Continuum of stakes, number of data point and their function



Assessment information as pixels



Classical approach to aggregation

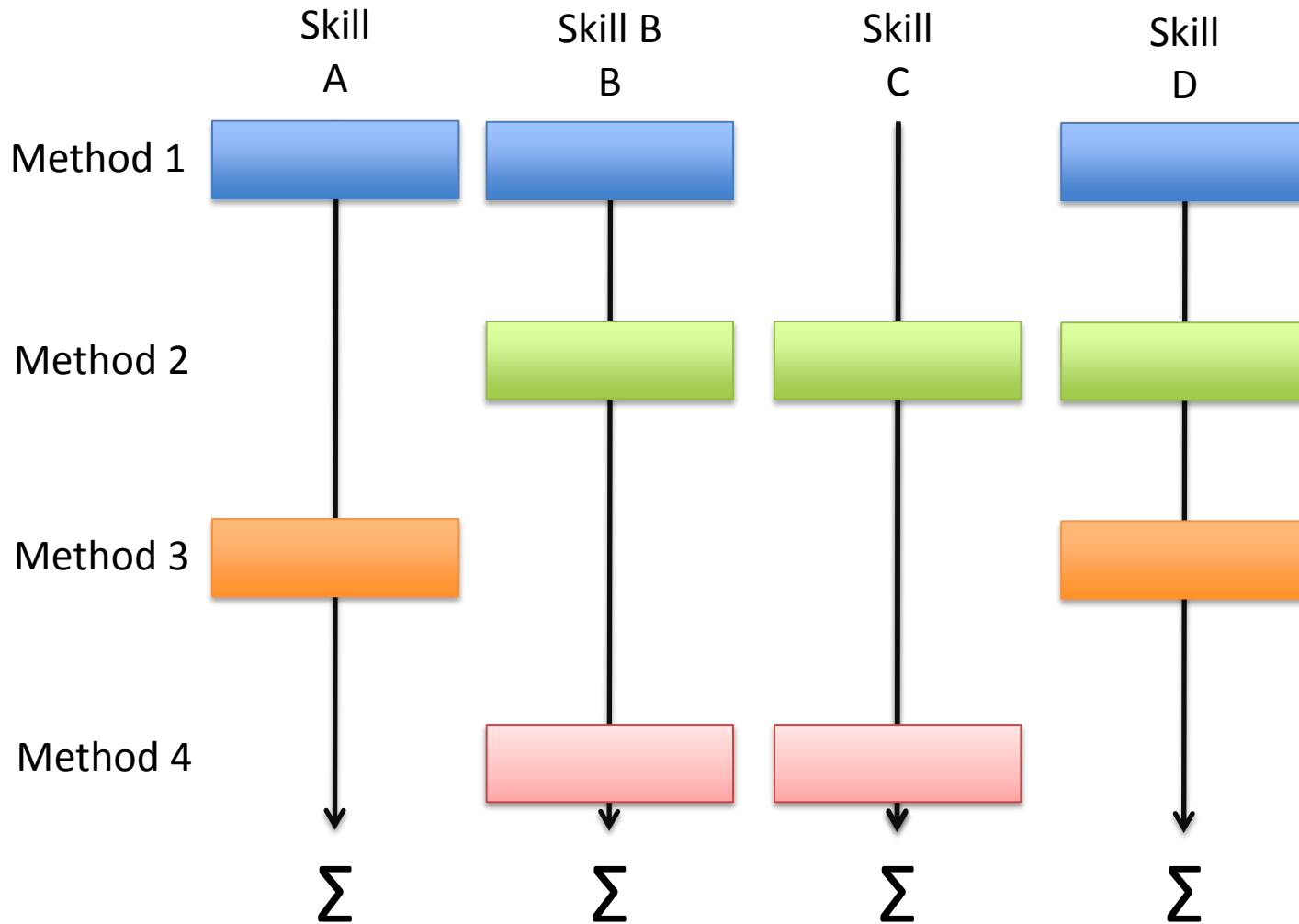
Method 1 to assess skill A  Σ

Method 2 to assess skill B  Σ

Method 3 to assess skill C  Σ

Method 4 to assess skill C  Σ

More meaningful aggregation





Overview

- From practice to research
- From research to theory
- **From theory to practice**
- From practice to research
- Conclusions

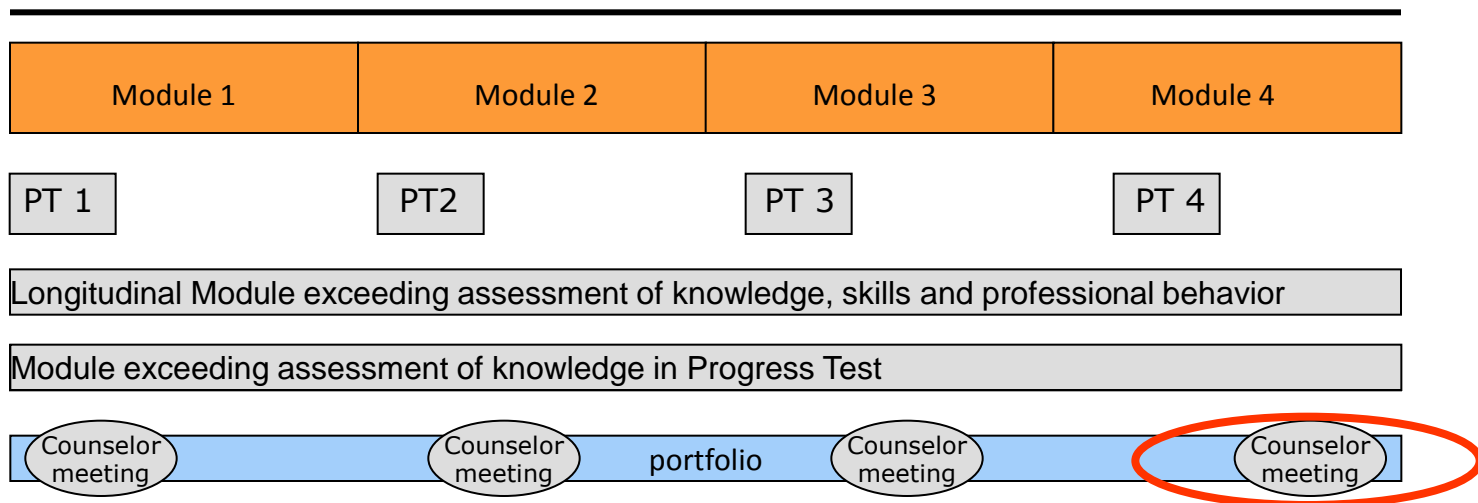


Physician-clinical investigator program

- 4 year graduate entry program
- Competency-based (Canmeds) with emphasis on research
- PBL program
 - Year 1: classic PBL
 - Year 2: real patient PBL
 - Year 3: clerkship rotations
 - Year 4: participation in research and health care
- High expectations of students: in terms of motivation, promotion of excellence, self-directedness

The assessment program

- Assessment in Modules: assignments, presentations, end-examination, etc.
- Longitudinal assessment: assignments, reviews, projects, progress tests, evaluation of professional behavior, etc.
- All assessment is informative and low stake formative
- The portfolio is central instrument



Student

Peer groups

Welcome

Momentaneous

Longitudinal

Test Overview

Domain	
Total	
Result	
Score	
Series	
Unprocessed	
Peer group	
UM FHML-G year group 3	
Reference values	
Percentiles	

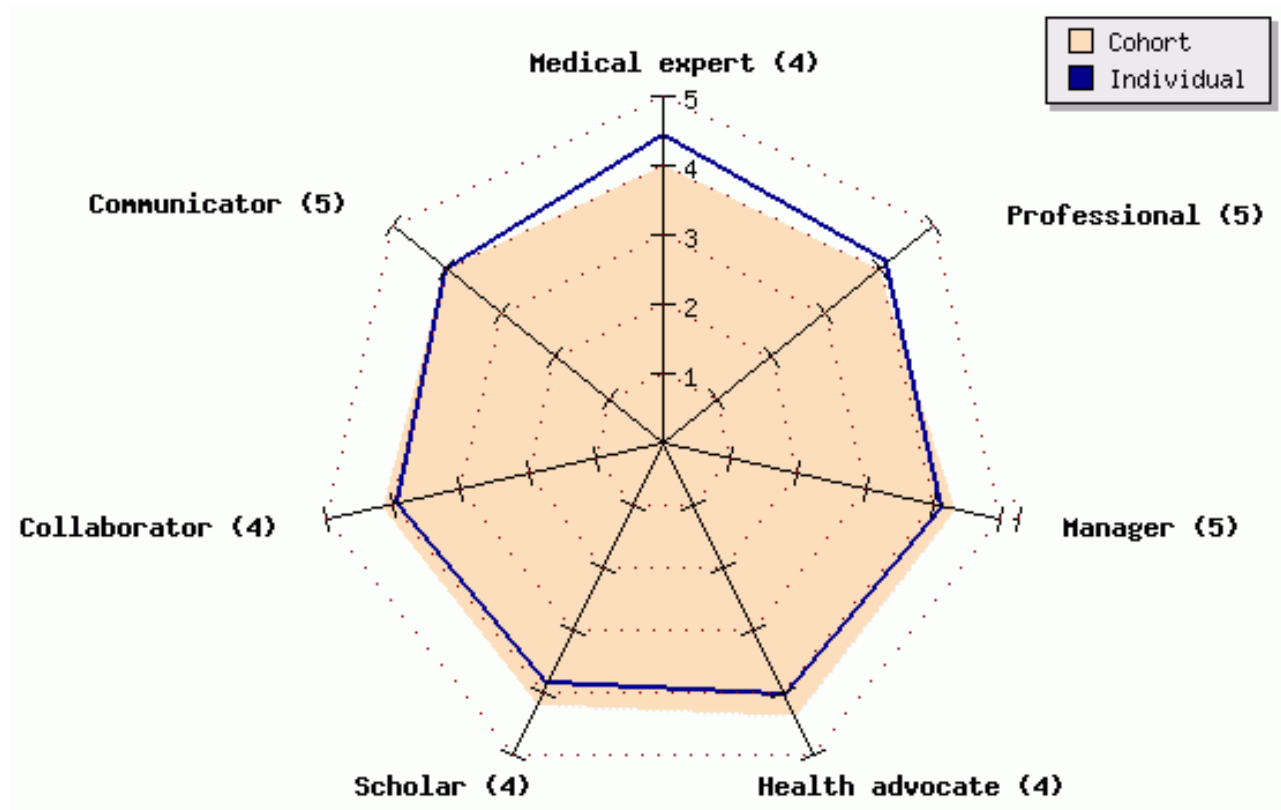
Longitudinal series (*unprocessed*) of score for *total* for student 403164 with peer group UM FHML-G year group 3 as background population



- Student score
- Lower confidence bound prognosis
- Prognosis
- Upper confidence bound prognosis

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

Maastricht Electronic portfolio (ePass)

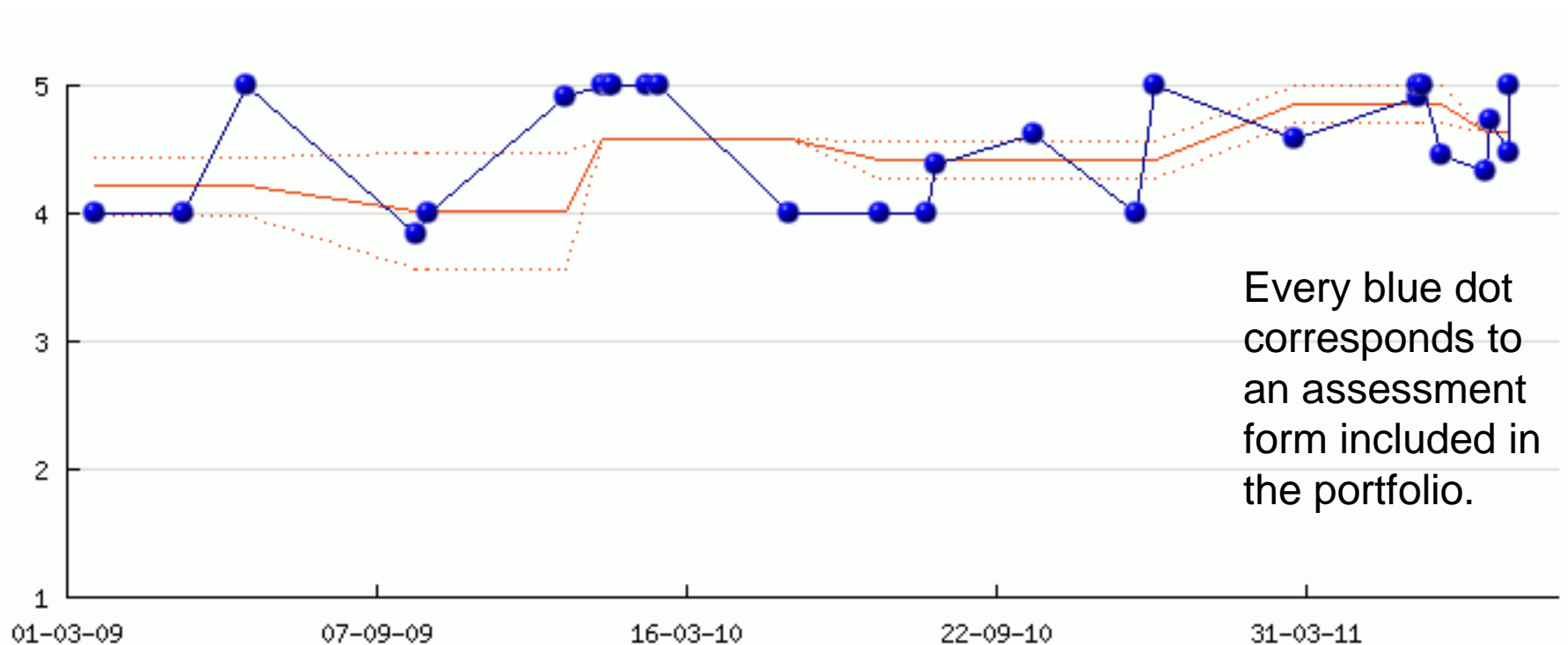


Comparison between the score of the student and the average score of his/her peers.

Maastricht Electronic portfolio (ePass)

1: Medical expert

Table view



Narrative feedback

Feedbacktype: Competency:

all

all

Date	Feedbacktype	Competency	Narrative feedback	Form
06-11-2013	Improvement	General	don't repeat too much, no irrelevant details 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. 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 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



"'LUNCH'? WELL, YES--BUT WHAT ARE YOUR LONG-TERM GOALS?"

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



Strategy to establish trustworthiness	Criteria	Potential Assessment Strategy (sample)
Credibility	Prolonged engagement	Training of examiners
	Triangulation	Tailored volume of expert judgment based on certainty of information
	Peer examination	Benchmarking examiners
	Member checking	Incorporate learner view
	Structural coherence	Scrutiny of committee inconsistencies
Transferability	Time sampling	Judgment based on broad sample of data points
	Thick description	Justify decisions
Dependability	Stepwise replication	Use multiple assessors who have credibility
Confirmability	Audit	Give learners the possibility to appeal to the assessment decision

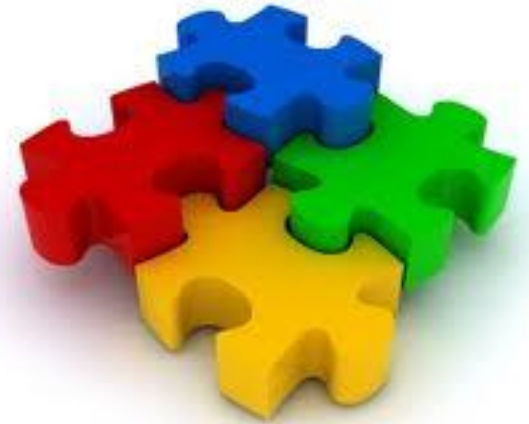
State of implementation in FHML

- Graduate entry medicine program
- Master of Medicine
- Partial implementation in Bachelor of medicine (full implementation in preparation)
- Partial implementation in Bachelor of life sciences.



Overview

- From practice to research
- From research to theory
- From theory to practice
- From practice to research
- Conclusions



Programmatic of learning t

LAMBERT W. T. SCHUW
Maastricht University, The

Abstract

In assessment a considerable amount of research implications for the types of construct human judgement and assessment programmes, how judgement is much needed.

A model for programmatic assessment fit for purpose

C. P. M. VAN DER VLEUTEN¹,
D. TIGELAAR³, L. K. J. BA
¹Maastricht University, The Netherlands,
²Utrecht University, The Netherlands,
³Utrecht University, The Netherlands

Abstract

We propose a model for programmatic assessment for decision making. The model is based on empirical research from intermediate and final assessment. The model is based on empirical research from intermediate and final assessment. The model is based on empirical research from intermediate and final assessment.

2011; 33: 478-485
MEDICAL
TEACHER

2012; 34: 205-214
MEDICAL
TEACHER

2015, 37: 641-646
MEDICAL
TEACHER

TWELVE TIPS

Twelve Tips for programmatic assessment

C.P.M. VAN DER VLEUTEN¹, L.W.T. SCHUWIRTH², E.W. DRIESSEN¹, M.J.B. GOVAERTS¹ &
S. HEENEMAN¹
¹Maastricht University, Maastricht, The Netherlands, ²Flinders University, Adelaide, Australia

Abstract

Programmatic assessment is an integral approach to the design of an assessment program with the intent to optimise its learning function, its decision-making function and its curriculum quality-assurance function. Individual methods of assessment, purposefully chosen for their alignment with the curriculum outcomes and their information value for the learner, the teacher and the organisation, are seen as individual data points. The information value of these individual data points is maximised by giving feedback to the learner. There is a decoupling of assessment moment and decision moment. Intermediate and high-stakes decisions are based on multiple data points after a meaningful aggregation of information and supported by rigorous organisational procedures to ensure their dependability. Self-regulation of learning, through analysis of the assessment information and the attainment of the ensuing learning goals, is scaffolded by a mentoring system. Programmatic assessment-for-learning can be applied to any part of the training continuum, provided that the underlying learning conception is constructivist. This paper

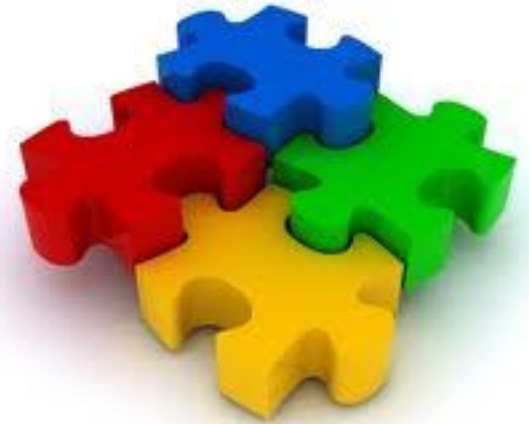
www.ceesvandervleuten.com

Findings on programmatic assessment so far

- The quality of the implementation defines the success
- Getting high quality feedback is a challenge
- Learners may perceive low stake assessments as high stake, all depending on the learning culture created
- Coaching and mentoring is key to the success
- High stake decision-making in competence committees work really well.

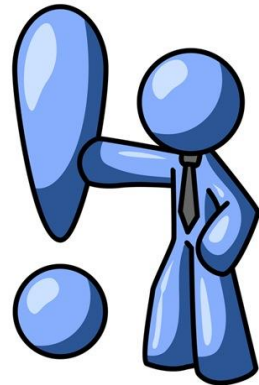
Overview

- From practice to research
- From research to theory
- From theory to practice
- From practice to research
- **Conclusions**



Conclusions 1: The way forward

- We have to stop thinking in terms of individual assessment methods
- A systematic and programmatic approach is needed, longitudinally oriented
- Every method of assessment may be functional (old and new; standardized and unstandardized)
- Professional judgment is imperative (similar to clinical practice)
- Subjectivity is dealt with through sampling and procedural bias reduction methods (not with standardization or objectification).



Conclusions 2: The way forward

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





Further reading:
www.ceesvandervleuten.com

12 tips

1. Develop a master plan
2. Adopt a robust system for collecting information
3. Develop examination regulations that promote feedback orientation
4. Assure that every low-stakes assessment provides meaningful feedback for learning
5. Provide mentoring to learners
6. Ensure trustworthy decision-making
7. Organise intermediate decision-making assessments
8. Encourage and facilitate personalised remediation
9. Monitor and evaluate the learning effect of the programme and adapt
10. Use the assessment process information for curriculum evaluation
11. Promote continuous interaction between the stakeholders
12. Develop a strategy for implementation