Towards a new paradigme for assessment ?

Based on a paper by:

Cees van der Vleuten
School of Health Professions Education
Maastricht University
The Netherlands

Overview of presentation

2012; 34: 205–214 MEDICAL TEACHER

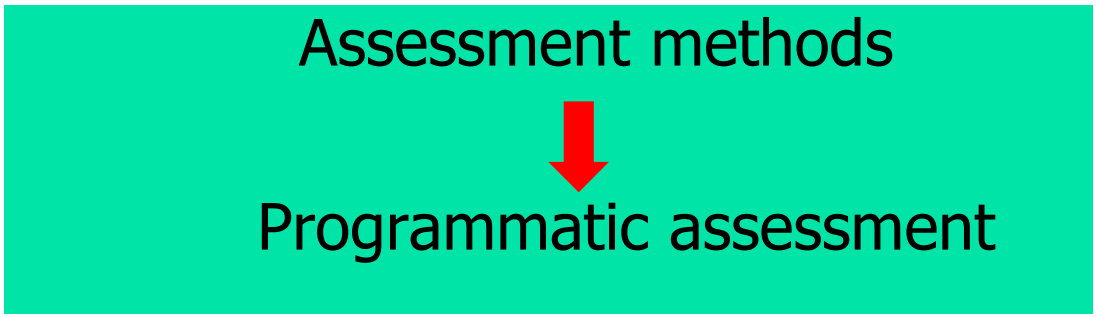
A model for programmatic assessment fit for purpose

C. P. M. VAN DER VLEUTEN[1], L. W. T. SCHUWIRTH[2], E. W. DRIESSEN[1], J. DIJKSTRA[1], D. TIGELAAR[3], L. K. J. BAARTMAN[4] & J. VAN TARTWIJK[5]

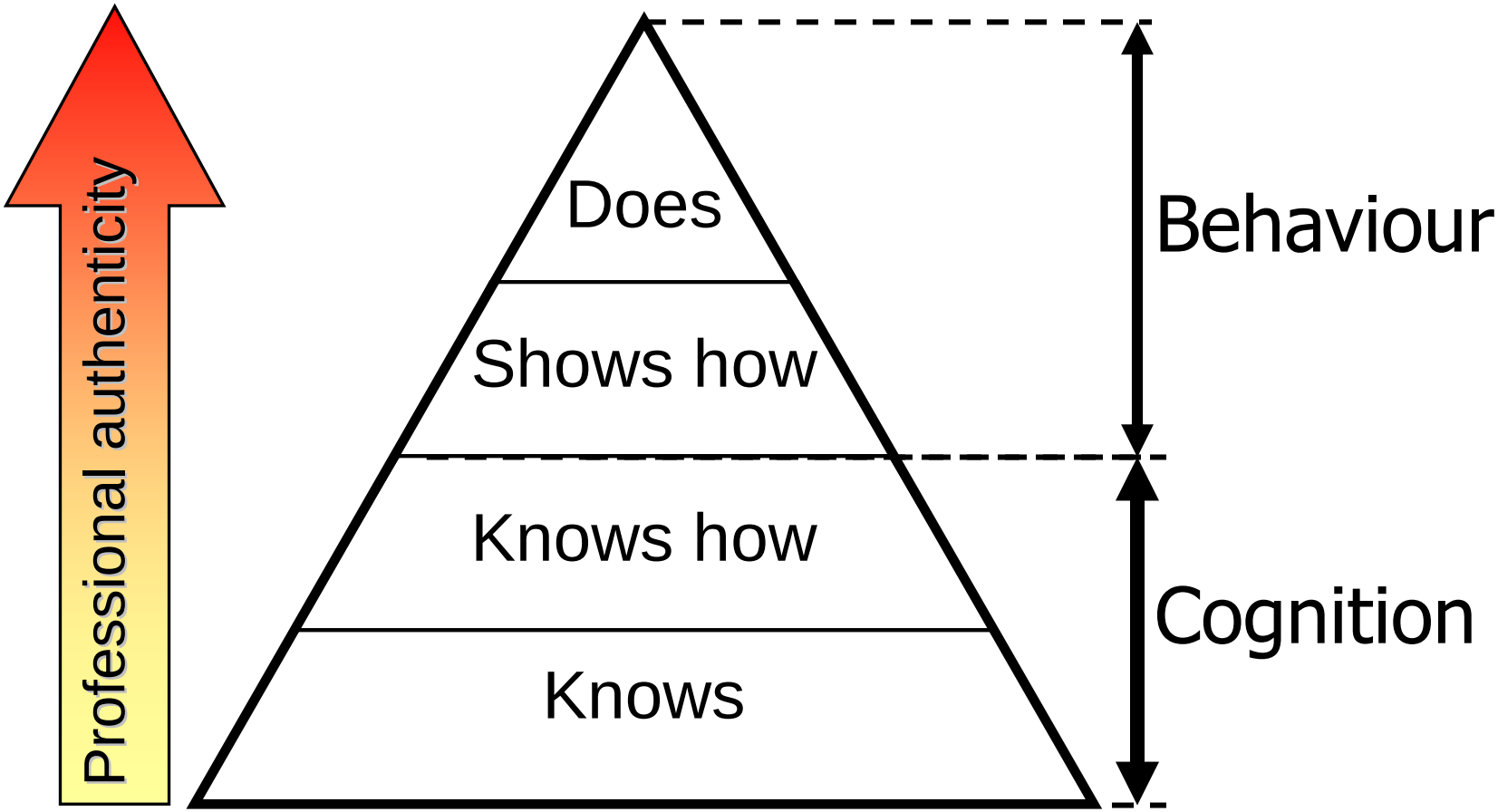
[1]Maastricht University, The Netherlands, [2]Flinders Medical School, Australia, [3]Leiden University Graduate School of Teaching, The Netherlands, [4]Utrecht University of Applied Sciences, The Netherlands, [5]Utrecht University, The Netherlands

Background

- « climbing » the pyramid
- A shift from an psychometric to an edumetric framework

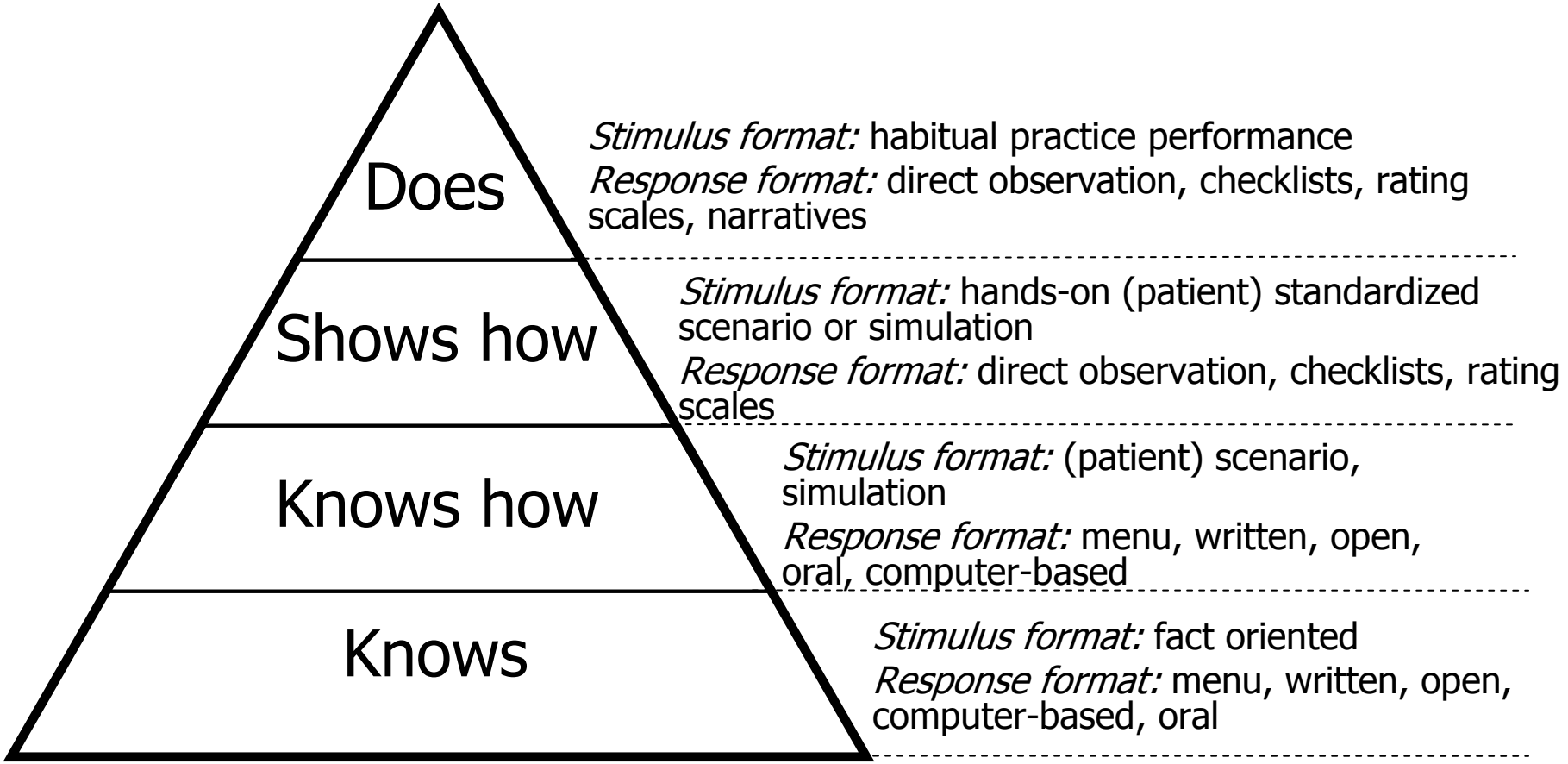


Simple competence model



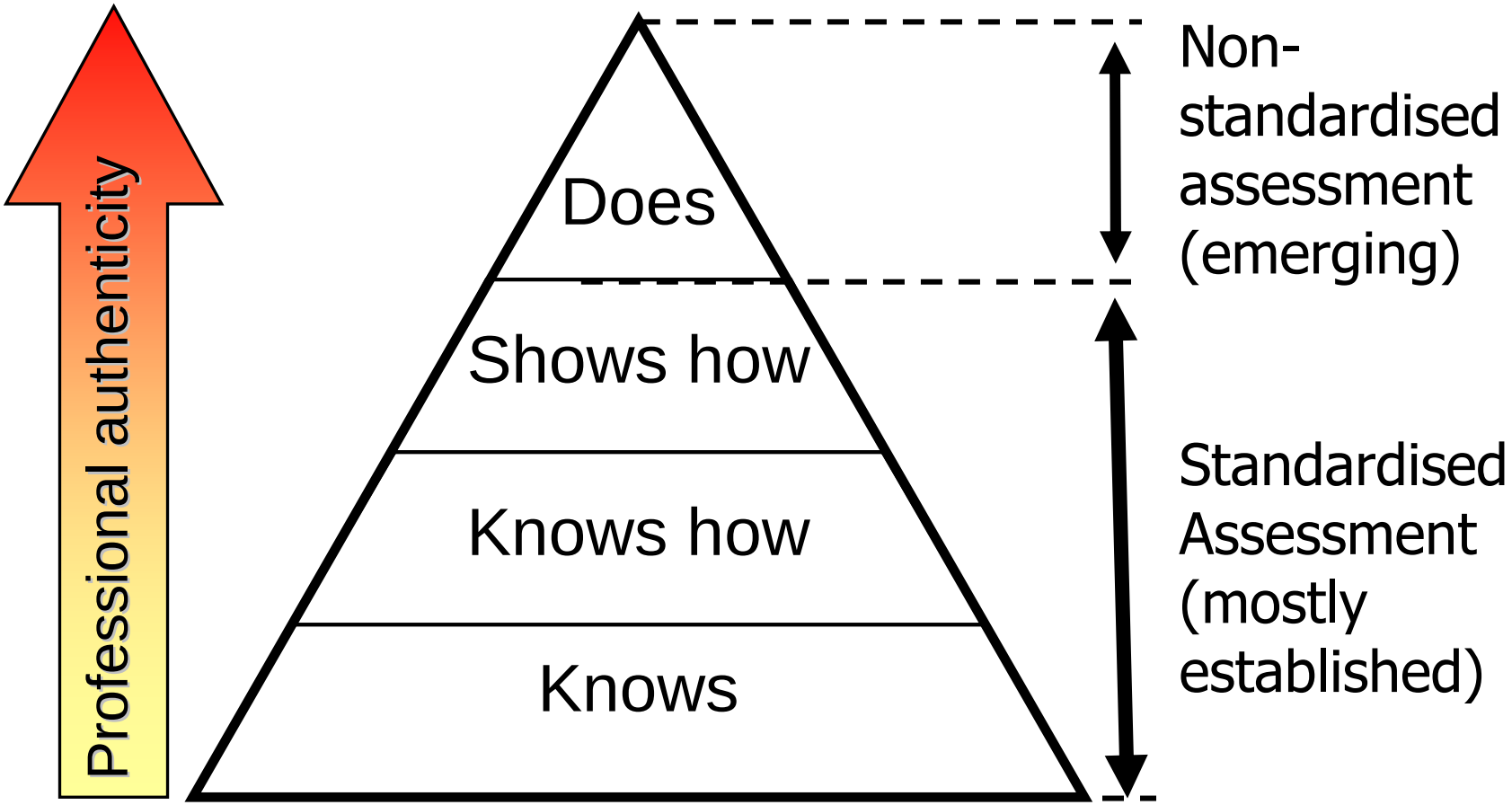
Miller GE. The assessment of clinical skills/competence/performance. Academic Medicine (Supplement) 1990; 65: S63-S7.

Assessment formats used



- Stimulus vs Response Format

Miller's pyramid



Miller GE. The assessment of clinical skills/competence/performance. Academic Medicine (Supplement) 1990; 65: S63-S7.

Assessing Knows, knows how and shows how

- Assessment principles
  1. Competence is specific, not generic
  2. Objectivity is not the same as reliability
  3. What is being measured is more determined by the stimulus format than by the response format
  4. Validity can be 'built-in'
  5. Assessment drives learning
  6. No single method can do it all

Assessment principle 1

- Competence is specific, not generic

Competence is not generic

Testing Time in Hours	MCQ[1]	Case-Based Short Essay[2]	PMP[1]	Oral Exam[3]	Long Case[4]	OSCE[5]	Mini CEX[6]	Practice Video Assess-ment[7]	In-cognito 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]Norcini et al., 1985
[2]Stalenhoef-Halling et al., 1990
[3]Swanson, 1987
[4]Wass et al., 2001
[5]Petrusa, 2002
[6]Norcini et al., 1999
[7]Ram et al., 1999
[8]Gorter, 2002

Practical implications

- Competence is specific, not generic
  - One measure is no measure
  - Increase sampling (across content, examiners, patients…) within measures
  - Combine information across measures and across time
  - Be aware of (sizable) false positive and negative decisions
  - Build safeguards in examination regulations

Assessment principle 2

- Objectivity is not the same as reliability

Objectivity is not the same as reliability

Testing Time in Hours	MCQ[1]	Case-Based Short Essay[2]	PMP[1]	Oral Exam[3]	Long Case[4]	OSCE[5]	Mini CEX[6]	Practice Video Assess-ment[7]	In-cognito 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

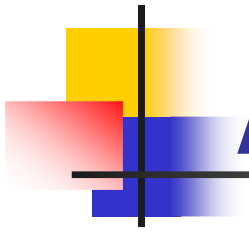
[1]Norcini et al., 1985
[2]Stalenhoef-Halling et al., 1990
[3]Swanson, 1987
[4]Wass et al., 2001
[5]Petrusa, 2002
[6]Norcini et al., 1999
[7]Ram et al., 1999
[8]Gorter, 2002

Reliability oral examination (Swanson, 1987)

Testing Time in Hours	Number of Cases	Same Examiner for All Cases	New 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

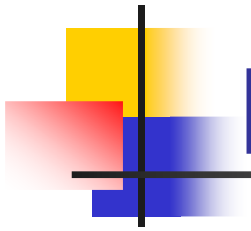
Practical implications

- Objectivity is not the same as reliability
  - Don’t trivialize the assessment (and compromise on validity) with unnecessary objectification and standardization
  - Don’t be afraid of holistic judgment
  - Sample widely across sources of subjective influences (raters, examiners, patients)



Assessment principle 3

- What is being measured is more determined by the stimulus format than by the response format



Empirical findings

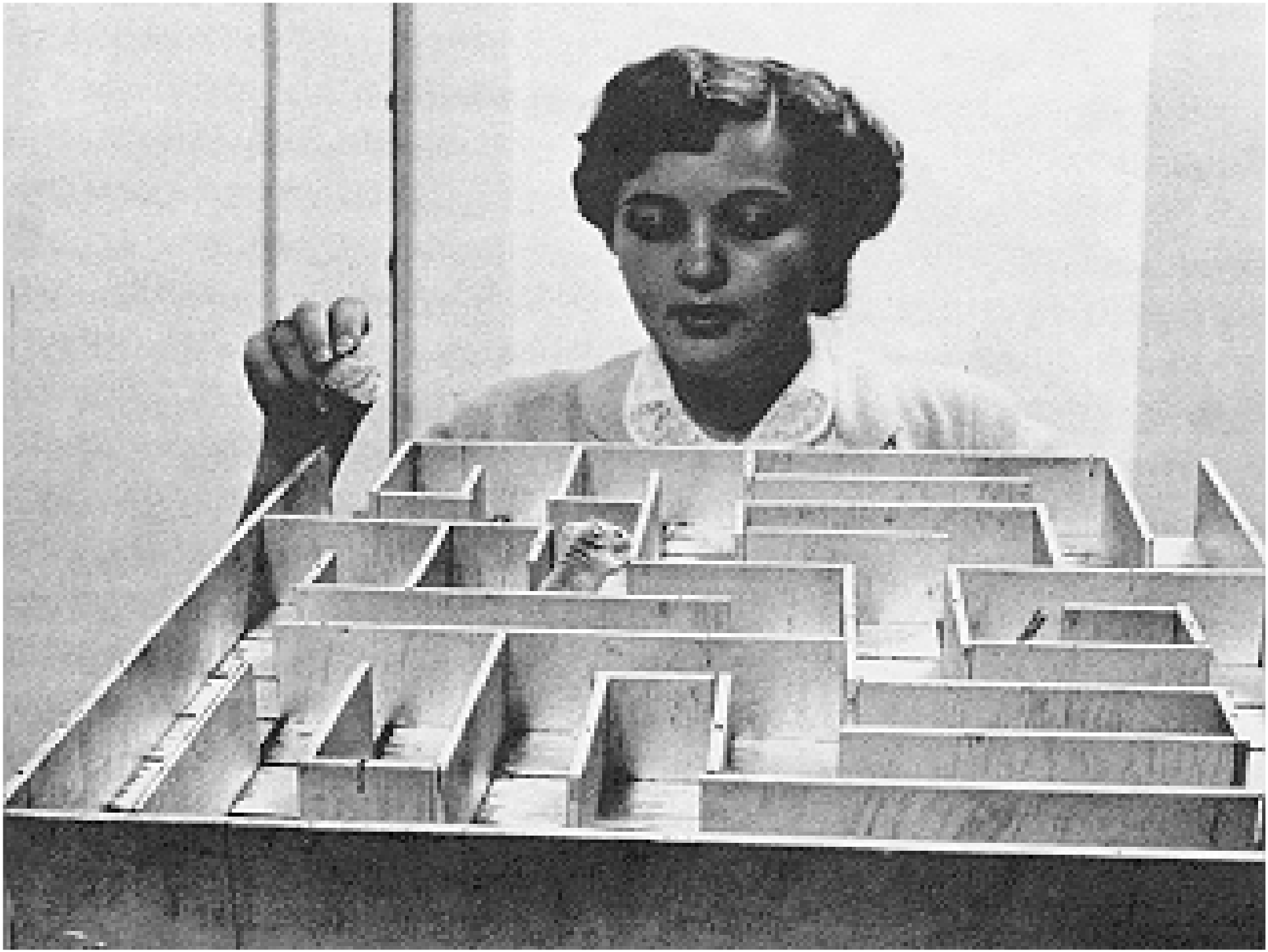
- Once reliable (meaning sufficient sampling) correlations across formats are huge
- Cognitive activities follow the task you pose in the stimulus format

Practical implications

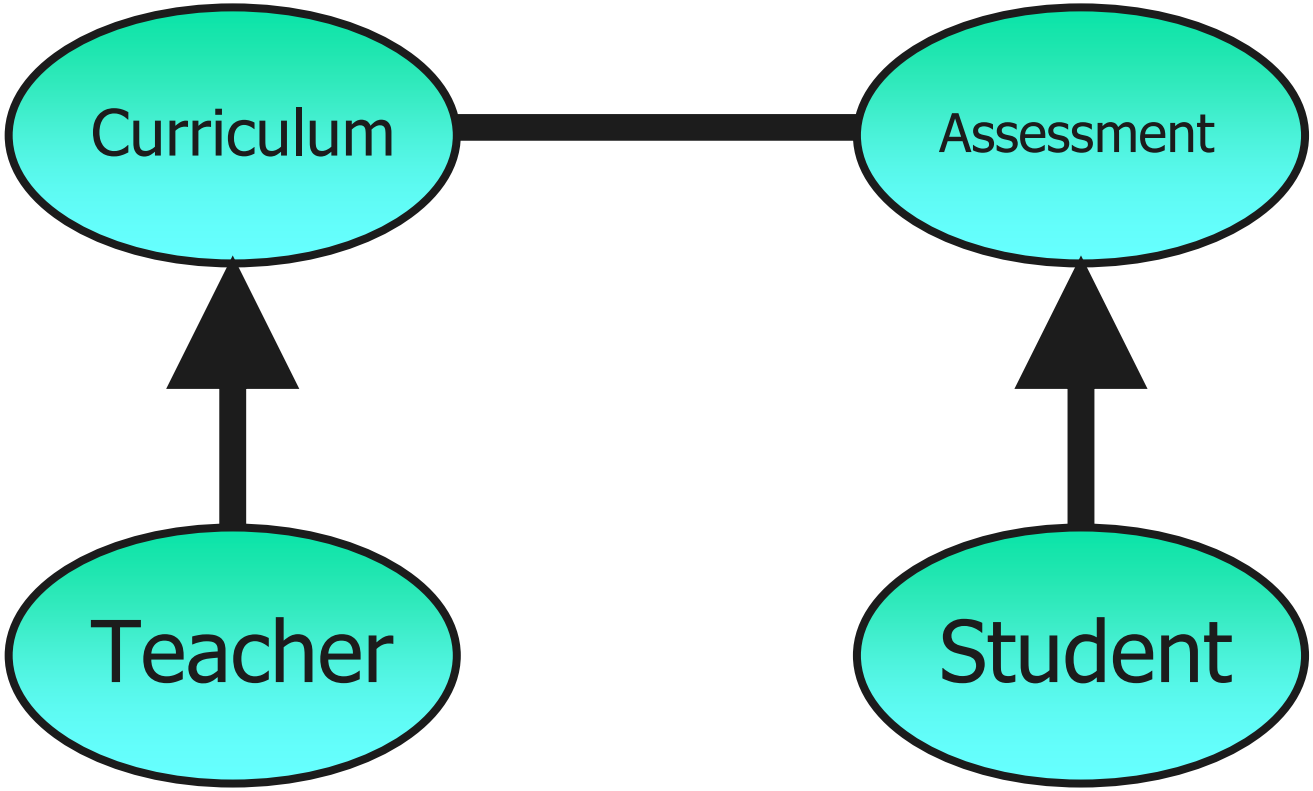
- What is being measured is more determined by the stimulus format than by the response format
  - Don’t be married to a format (e.g. essays)
  - Worry about improving the stimulus format
  - Make the stimulus as (clinically) authentic as possible (e.g. in MCQs, OSCEs)

Assessment principle 5

- Assessment drives learning



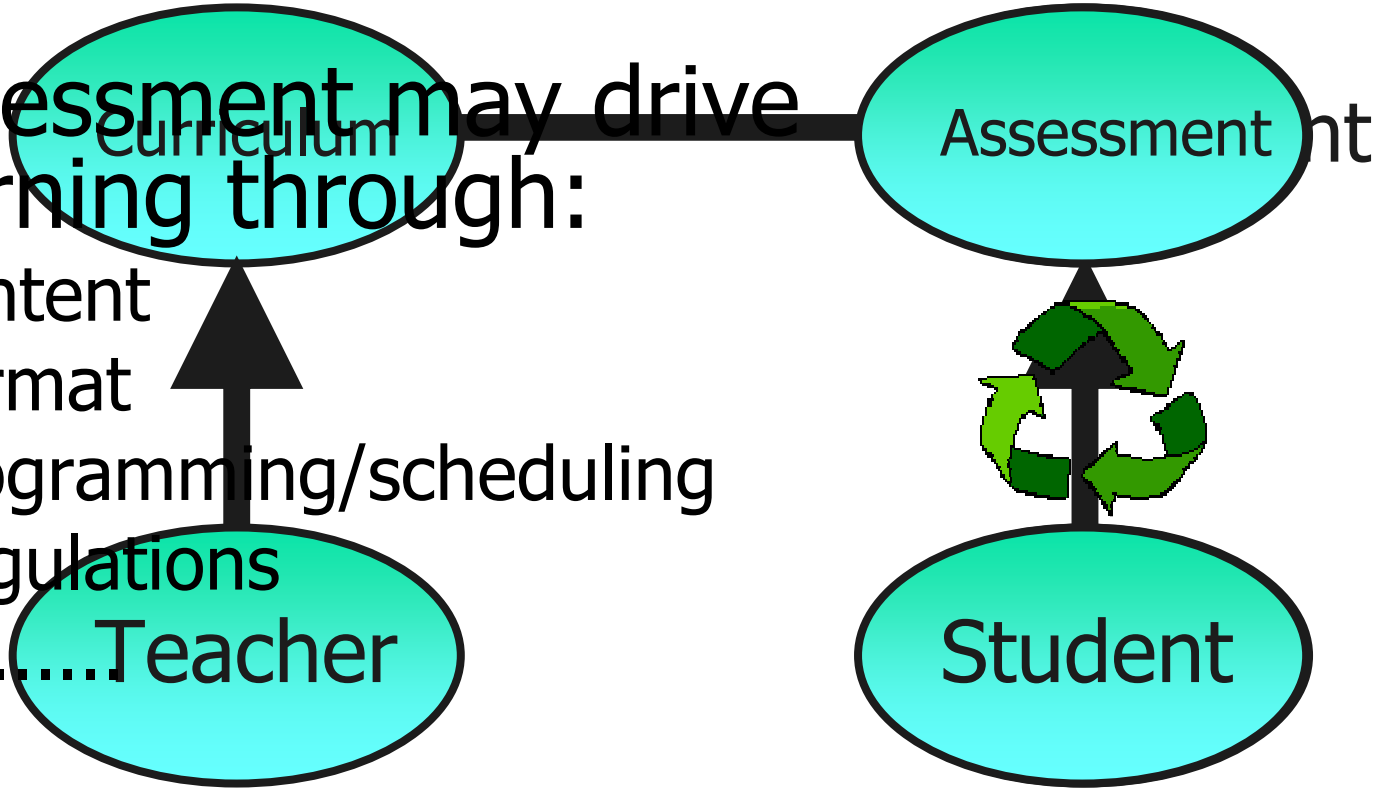
An alternative view



An alternative view

Assessment may drive learning through:

- Content
- Format
- Programming/scheduling
- Regulations
- ……

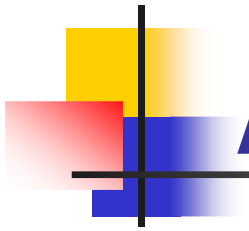


Empirical findings

- The relationship between assessment and learning is complex
- Summative assessment systems often drive in a negative way
- Formative feedback has dramatic impact on learning
- Learners want feedback (more than grades), but not getting it

Practical implications

- Assessment drives learning
  - For every evaluative action there is an educational reaction
  - Verify and monitor the impact of assessment (evaluate the evaluation); many intended effects are not actually effective -> hidden curriculum
  - No assessment without feedback!
  - Embed the assessment within the learning programme (cf. Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. Applied Measurement in Education, 13(2), 181-208.)
  - Use the assessment strategically to reinforce desirable learning behaviours



Assessment principle 6

- No single method can do it all

Empirical findings

- One measure is no measure
- All methods have limitations (no single superior method exists)
- Different methods may serve a different function
- In combination, information from various methods provide a richer picture and combines formative and summative functions

Practical implications

- No single method can do it all
  - Use a cocktail of methods across the competency pyramid
  - Arrange methods in a programme of assessment
  - Any method may have utility (including the 'old' assessment methods depending on its utility within the programme)
  - Compare assessment design with curriculum design
    - Responsible people/committee(s)
    - Use an overarching structure
    - Involve your stakeholders
    - Implement, monitor and change (assessment programmes 'wear out')

Assessing Does

- Assessment principles
  1. A feasible sample is required to achieve reliable inferences
  2. Bias is an inherent characteristic of expert judgment
  3. The validity lies in the users of the instruments, more than in the instruments
  4. Formative and summative functions are typically combined
  5. Qualitative, narrative information carries a lot of weight
  6. Summative decisions can be rigorous by using non-psychometric qualitative research procedures