Assessment and evaluation of learning

TED, Ankara, Turkey
October, 23, 2015 – Third Day of Workshop for Academics
Dr. Anne Nevgi
Centre for Research and Development of Higher Education
University of Helsinki

Today's programme

- 9.30-11.00 Assessment of learning, basic constructs, rubrics
- 11.00-12.00 Developing rubric for own class based on criteria developed on first workshop – how to assess learning outcomes?
- 12.00-13.00 Lunch
- 13.00-15.00 Formative assessment, self-assessment, peer assessment, peer tutoring, oral assessment, poster assessment - group work and discussion
- 15.00-15.30 Break, coffee, tea
- 15.30-17.30 How to assess learning outcomes at own course?; group work. Ending the three-day workshop, feedback and general discussion

A change in the assessment culture

(Biggs 2003; Falchikov 2005; Gibbs, 2006)

Traditional assessment

- Measures the amount of knowledge
- Assessment is seen as a separate part of teaching and learning , often at the end of the course (summative assessment)
- Guides students to pass courses (certification)
- Norm-based assessment (students' learning outcomes are compared with each other)
- Teacher takes full responsibility of assessment

Developmental assessment

- Quality instead of quantity
- Assessment is a part of the teaching-learning process and occurs during the course (formative assessment)
- Guides students to understand the content and rehearse skills needed in the working life (learning experience)
- Criterion-based assessment (learning outcomes are assessed through assessment criteria)
- Students are involved in the assessment

The changing demands of the working life

In addition to subject specific knowledge, working life requires a number of different skills:

- Interactivity
- Critical thinking
- Ability to analyse, synthesize and make interpretations
- Information literacy reading skills
- Problem solving skills
- Skills to evaluate one's own competencies

Assessment of student learning

Central questions regarding assessment:

- *What* is assessed? (subject knowledge, skills…)
- *When* to assess? (before, during, after)
- *Who* is assessing? (teacher, peer, student him/herself)

Multiple dimensions of assessment – Assessment Cubic

What level of understanding/thinking?

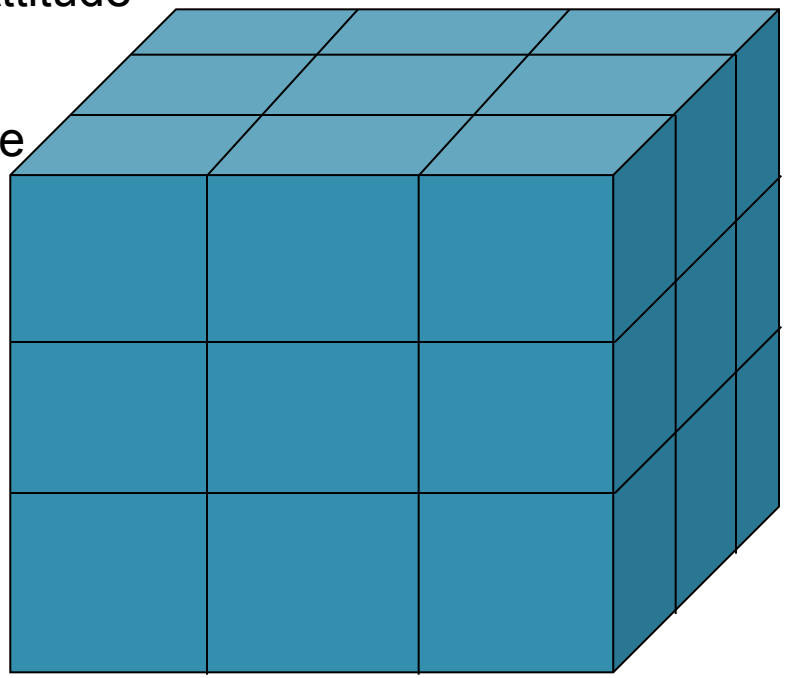
Declarative Procedural Metacognitive
Factual Evaluation Theory
Conceptual Application Creation

What is being assessed?

Attitude
Skill
Knowledge

Purpose of assessment?

Diagnostic
Formative
Summative



Self Peer Teacher

Who assesses?

What Criteria?

SOLO Taxonomy

Bloom's Revised Taxonomy

Holistic or analytic assessment

(Assessment Cubic: Auli Toom, 11.2.2008)

When to assess?

- **Summative assessment**
  - At the end of the learning process
  - The aim is to assess, how well the students' have learned what they were supposed to learn
  - The assessment of the learning outcomes
- **Formative assessment**
  - During the learning process
  - The focus is on giving/receiving feedback during the learning process
    - Feedback for both the student and the teacher
    - Developmental assessment
- **Diagnostic assessment**
  - Before the learning process
  - Provides information to the teacher anout the students' prior knowledge
  - Helps the student to activate prior knowledge and link it to the content of the course (e.g., Brown et al, 1997)

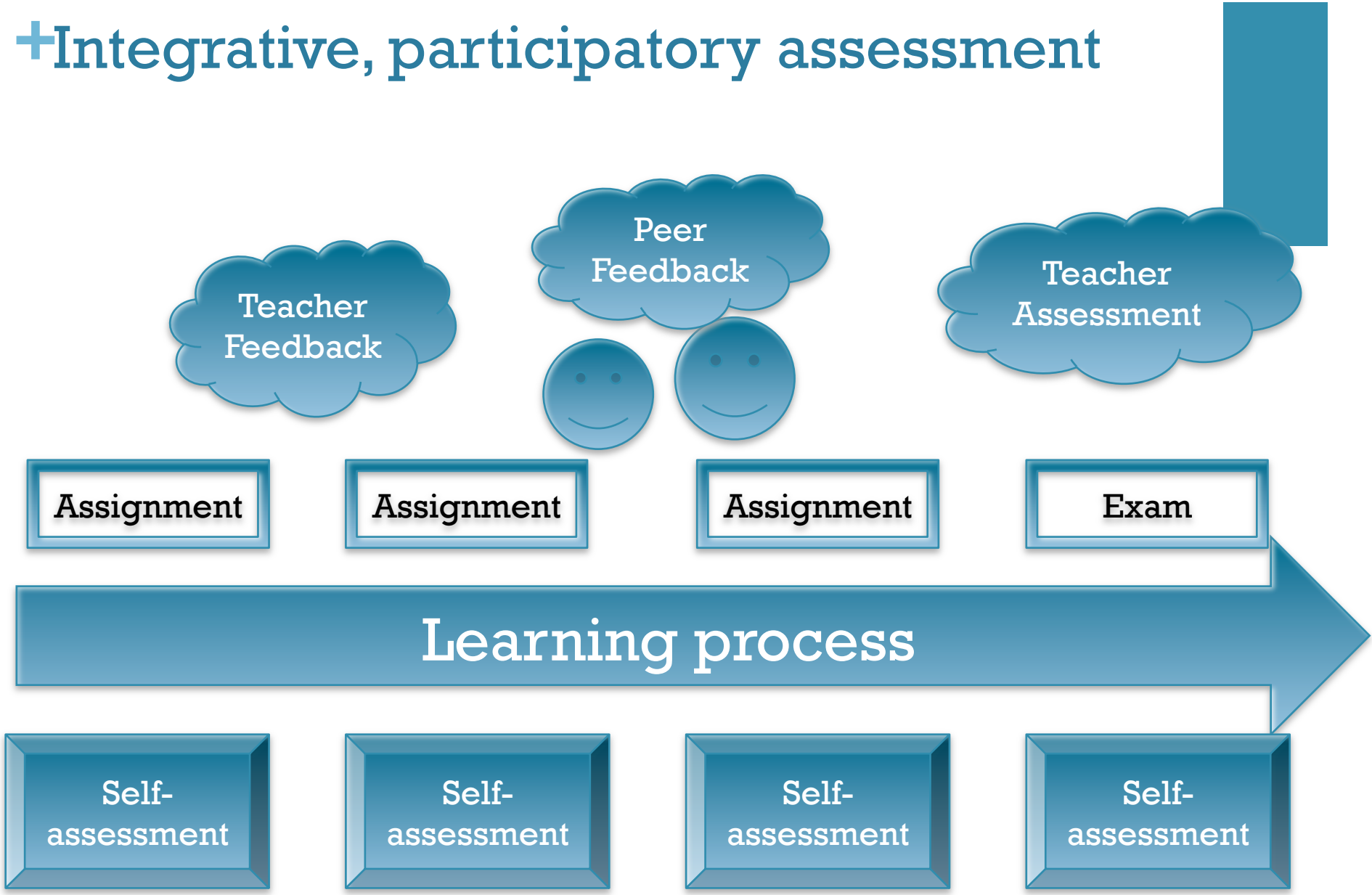
Peer- and self-assessment

(e.g. Davey 2011, 2012; Sluijsman, 2006; Struyven et al., 2005)

- Activates students and develops generic and reflective skills
- Combines the learning and assessment processes together
- Supports higher quality learning outcomes
- Gives feedback to students about their own learning
- Requires clear assessment criteria
- Reduces the teachers' workload

Challenges in peer- and self-assessment?

Integrative, participatory assessment



Discussion

- *What kind of challenges have you faced when assessing your students' learning?*

Discuss with your group and write down 1-3 challenges!

Grading schemes /Grading scales

UNIVERSITY OF HELSINKI

- 5 = Excellent
- 4 = Very Good
- 3 = **Good**
- 2 = Satisfactory
- 1 = Passable
- 0 = Failed

- Laudatur
- Eximia cum laude approbatur
- Magna cum laude approbatur
- **Cum laude approbatur**
- Non sine cum laude approbatur
- Lubenter approbatur
- Approbatur
- Improbatur

- Pass excellence
- **Pass**
- Failed

- **Pass**
- Failed

Grading scales / schemes in USA, Turkey

USA (typical)

- A = 92 – 100 %
- B = 83 – 91 %
- C = 74 – 82 %
- D = 65 – 73 %
- E/F = 64 % or less

TURKEY

5 (85–100%) = Very Good (*Pekiyi*)

4 (70–84%) = Good (*İyi*)

3 (60–69%) = Average (*Orta*)

2 (50–59%) = Pass (*Geçer*)

1 (0–49%) = Fail (*Zayıf*)

European Credit Transfer and Accumulation System Grading scale – relative scale

Grade	Definition	Cumulative %
A	Outstanding performance with only minor errors	Best10 %
B	Above the average standard but with some errors	Next 25 %
C	Generally sound work with a number of notable errors	Next 30 %
D	Fair but with significant shortcomings	Next 25 %
E	Performance meets the minimum criteria	Next 10 %
FX	Fail – some more work required before the credit can be awarded	Fail (almost passing)
F	Fail – considerable further work is required	Fail

http://en.wikipedia.org/wiki/ECTS_grading_scale

Discussion in groups

- How and on what based we should to define the level for passable level of intended learning outcome?
- How to define what is the lowest acceptable level for a student to pass a exam?
- Quality of learning or quantity of learning?

SOLO-taksonomy **S**tructure of the **O**bserved **L**earning **O**utcome

- Prestructural
- Unistructural
- Multistructural
- Relational
- Extended abstract

SOLO-taxonomy

CRITICAL THINKING

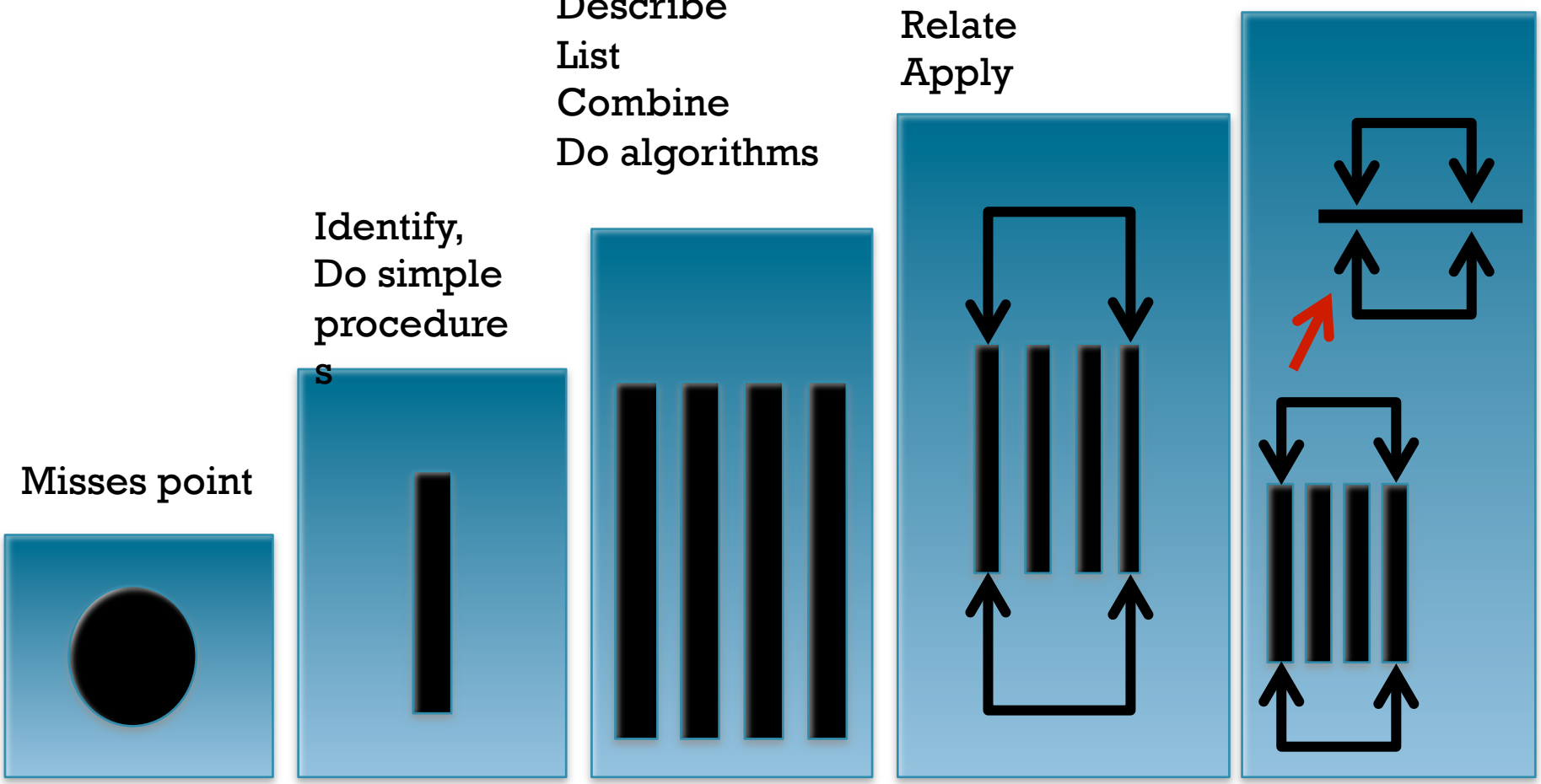
Misses point

Identify,
Do simple procedures

Enumerate
Describe
List
Combine
Do algorithms

Compare /contrast
Explain causes
Analyse
Relate
Apply

Theorise
Generalise
Hypothesise
Reflect



Prestructural | Unistructural | Multistructural | Relational | Extended abstrac

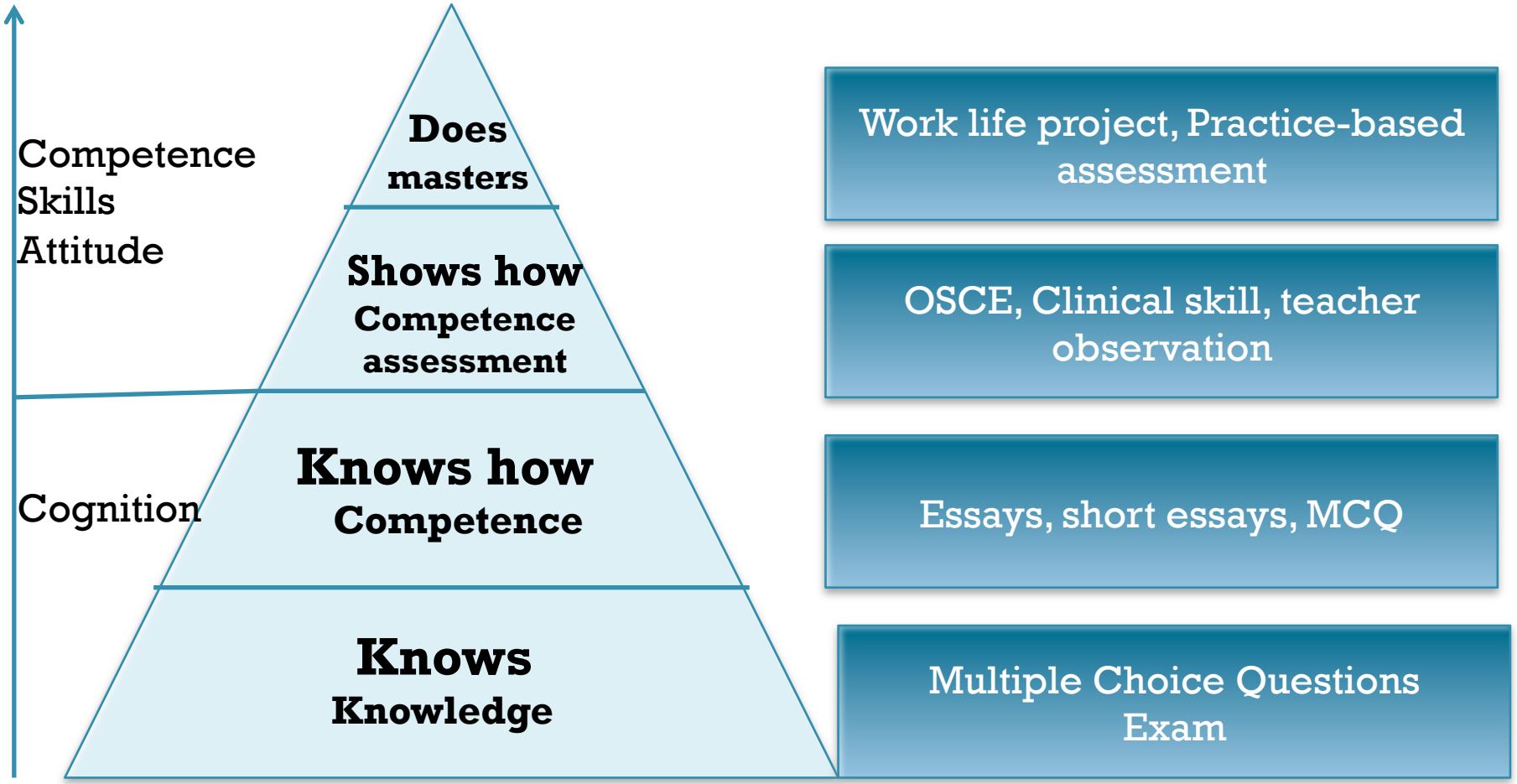


Figure. Miller's pyramid of competence in medical sector

Three essential features of Rubric

- Assessment (evaluation) criteria (leftmost column)
  - the factors that an assessor considers when determining the quality of a student's work
  - the criteria reflect the processes and content judged to be important
- Quality definitions (2-5 column)
  - provide a detailed explanation of what a student must do to demonstrate a skill, proficiency or criterion in order to attain a particular level of achievement
    - E.g. poor, fair, good or excellent
- Scoring strategy

Assessment criteria

(Biggs, 2003; Sadler, 2005)

- Assessment criteria indicate what students should know/be able to do in order to get a certain grade
- Should be clear to the students from the beginning of the course – *Transparency!*
- In *norm-referenced assessment* the students performance is compared to other students performance in the course and students grades are formed in relation to other students: (*grading on the curve)*
  - The problem: Does not measure what and how well the students have learned and reached the learning goals
- In *criterion-referenced assessment* the grades are formed on the basis of students' performance and are not dependent on other students' performance
  - Students performance is compared in relation to assessment criteria
  - Increases the reliability and fairness of assessment
  - Student learning can be influenced through assessment criteria!

Rubric and grading

Criteria	Excellent	Very good	Good	Satisfactory	Poor
Criterion A	Hypothesizes generalizes, presents new model	Compares, analyzes, applies existent model	Compare, describes, Defines, combines,	Describes, lists,	Lists
Criterion B	Presents how facts are related to general theoretical model	Describes, explains and analyzes how facts are related	Describes and explains how facts are related	Describes how facts are related	Facts, but mostly irrelevant

Students' perspective on rubric use

- Clarifies the target what they are expected to learn
- Allows them to regulate their progress
- Make marks and grades fair and transparent
- Help them to focus essential issues
- Reduces anxiety and fear
- Rubrics should be given to students at the beginning of course, before they prepare for an exam
- Negative experiences if rubric is given after an exam

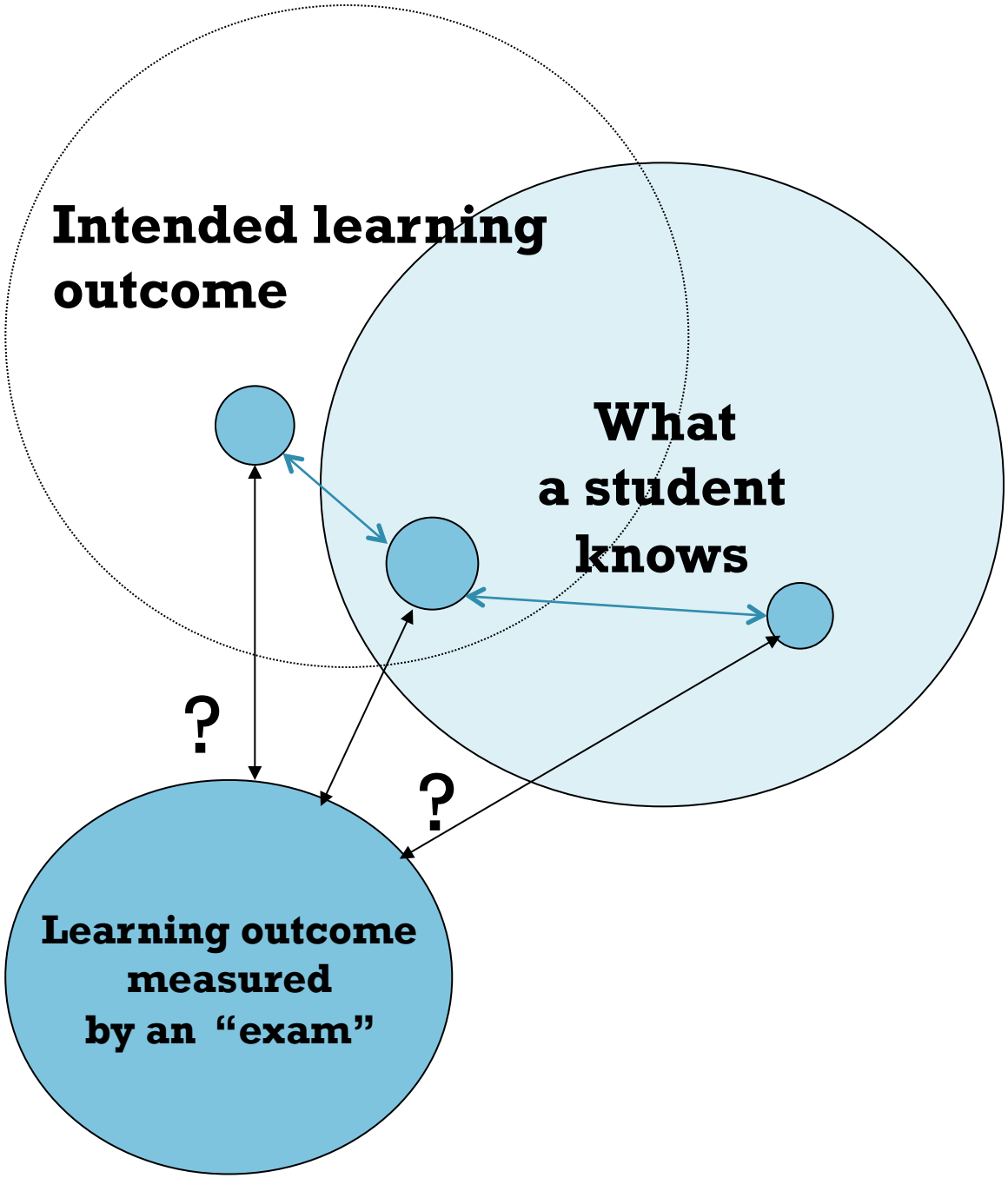
Teachers' perspective on rubric use

- Both for and resistance to use rubrics
- Positive: helps to mark and grade consistently, reliably, and efficiently, increases transparency in assessment
- Negative: resistance due to narrow concept of rubric use in assessment

Group work: Practicing to create a rubric

- Define and describe a learning outcome using concrete verbs (SOLO, Bloom)
- Create a rubric
- Define a criterion (what is a criterion for intended learning outcome?)
- Define the levels of student achievement in relation to criterion. Describe the levels by using concrete verbs.

VISITING OTHER GROUP: One member (host) of group remains in the table, other members move to another group table. Visiting group comments criterion and levels, how clear, instructive, informative they are. Returning back to own table.


Intended learning outcome
What a student knows
?
?
Learning outcome measured by an "exam"

Validity and reliability of an exam / assessment

Validity – does an exam measure that learning outcome that it is expected to measure?

Reliability – an exam is reliable if it is valid

Content validity: an exam is valid ft the questions/tasks are relevant, appropriate and representative of the construct that is examined and/or cognitive processes they aim to test.

- **Face validity**: an exam/test looks like an appropriate test in terms of readability, clarity, ease of administration.

Construct validity: an exam is valid if the question/tasks measure the domain of knowledge that is examined.

The reliability of peer-assessment

An example from a Bioscience course
(Rytkönen, Virtanen & Postareff, under review)

- 84 students assessed each others exam answers; also the teacher assessed all the answers
  - In 58% of the cases the grades given by the teacher and the student were exactly the same
  - In 35% of the cases the grades differed from each other by one point
  - In 7% of the cases the grades differed from each other by two (or three)

Unclear assessment criteria influences the reliability and validity of assessment

(Hailikari, Postareff, Räisänen, Tuononen & Lindblom-Ylänne, 2014)

- Assessment criteria are not always clear to the teachers themselves; this is reflected in the ***reliability*** of assessment

  Teacher: *"If you think about the situation in this question, there are huge differences in the points even though the answers were principally of the same quality."*

  "That is actually quite scary, the evaluation of the answers can vary two points depending on the time of the day, the mood or the order I´ve assessed the answers."(Teacher)

- The exams do not always measure what is intended to measure; this is reflected in the ***validity*** of assessment

  Teacher: *"You just have to remember things. For God's sake [surprised], you don't have to apply and integrate knowledge in any of these tasks. The fourth task is the only one but actually it's enough that you just remember what has been talked about."* (Teacher)

The grades do not always reflect the quality of learning

- The grades and students' descriptions of their own learning and understanding might be contradictory (Asikainen, Virtanen, Parpala & Lindblom-Ylänne, 2012)
- The grades and level of understanding were not always in relation with each other (Hailikari et al., 2014)

Teacher: *"The grades should reflect it (learning outcomes) but unfortunately they aren´t reflecting it in this exam because if you answer these tasks, you don´t have to know anything else but what has been told in lectures and you can achieve a good grade."* (Teacher)

Which assessment practices could serve the new purposes?

For example, the use of *self assessment, peer assessment and formative assessment,* which are oriented towards facilitating students' learning and enabling students to judge their own achievements.

Boud & Falchikov, 2006; Bartram, 2004; De Corte, 1996

The influence of assessment varies depending on the individual
(Lindblom-Ylänne & Lonka 2001)

- Assessment guides learning especially for students with a reproduction orientation
  - These students study in order to be successful in the exam
- Students having an ”understanding orientation” are immune to the learning environment
  - They study in order to *learn* for themselves
  - However, they are a minority

The benefits of group exam

- The exam itself is a learning situation, the aim is to enhance learning
- Group exam resembles real-life situations and teaches working-life skills
  - Problem-solving, collaboration and communication skills
- Reduces teachers' workload
- Reduces exam-anxiety

Formative assessment and feedback

- Safe atmosphere and trust.
- Be positive, try to see student's perspective.
- Good feedback is concrete and focuses on student's needs to improve her learning, understanding, competence, skills. Be specific what you say.
- It is timely, just-in-time, immediately after and/or during the performance
- It is critical but not mean. Start from the positive and say honestly what needs to be corrected.

Good feedback (Nicol-Macfarlane-dick, 2006)

1. helps clarify what good performance is (goals, criteria, expected standards);

2. facilitates the development of self-assessment (reflection) in learning;

3. delivers high quality information to students about their learning;

4. encourages teacher and peer dialogue around learning;

5. encourages positive motivational beliefs and self-esteem;

6. provides opportunities to close the gap between current and desired performance;

7. provides information to teachers that can be used to help shape teaching.

How to carry out an oral group exam: a concrete example

- Form groups of 6-10 students:
  - Each group is divided into two subgroups (A and B)
- Each group has a tutor
- Altogether three exam questions (given on paper):
  - First group A answers, group B listens, group B has an opportunity to fill in the answer
  - Then vice versa(second exam question): group B answers, group A competes
  - Third question is for the whole group together
- Teacher's/ tutor's role is active
  - Poses clarifying questions if necessary
  - Makes notes for the assessment during the exam
- (The participants can use their own notes)
- Model answers and feedback for the group

Example of the schedule of an oral group exam (3 h)

EXAMINATION

- QUESTION 1:
  - Preparation Group A (5 min.)
  - Group A answers (35 min.)
  - Group B fills in for 10 min.
- QUESTION 2:
  - Preparation Group B(5 min.)
  - Group B answers (35 min.)
  - Group A fills in for 10 min.
- QUESTION 3:
  - Preparation both groups A & B together (5 min.)
  - Both groups answer together (35 min.)

FEEDBACK AND DISCUSSION

30 min.:

- clarifying questions, experiences,
- short feedback
- clarifying misunderstandings
- self and peer assessment (form)

What is assessed in an oral group exam?

- The content: knowledge and level of understanding (listing or relating/ applying)
- The structure of the answers:
  → the definitions of concepts, literacy, consistency, understanding the relations
- Participation and interaction skills
- Critical thinking and argumentation

Grading of an oral group exam

- The group gets a mutual grade but individual grades are possible
- Short feedback right after the exam
- Written feedback for the whole group: what was the purpose of the question, what was good/bad in the answer, clarifying misconceptions
- The problem of free-riders:
  - Written peer-assessment regarding the exam and the group work during the course

SELF AND PEERASSESSMENT FORM

	Own participation in group work *How did I participate in the group? What was my contribution to the group work?*	**Own content knowledge** *How well did I master the contents? How well did I communicate it to the group?*
SELF ASSESSMENT ***NAME:***	**WRITTEN ASSESSMENT:** **GRADE 1-5:**	**WRITTEN ASSESSMENT:** **GRADE 1-5:**
	Group work *How well did the group work? Did everybody participate equally? If not, whose participation was divergent and how?*	**Groups' content knwoledge** *How well and equally the group members were familiar with the contents? How did the members of the group share their knowledge? Were there exceptions(who, how?) How well did the group handle the learned contents (different perspectives, innovativeness)?*
RYHMÄN ARVIOINTI ***RYHMÄN NIMI:***	**WRITTEN ASSESSMENT:** **GRADE 1-5:**	**WRITTEN ASSESSMENT:** **GRADE 1-5:**

Poster exam

- So called authentic exam (Karjalainen 2001): A scientific preseantion at the conference
- Teachers give the topic for the poster in advance
- Students prepare and present a poster as she would for a conference
  - Everything that will be presented has to be in the poster
- The focus is on concepts and their relations, the use of literature and informativeness of the poster
- 20 minutes to present; Teacher and other student groups provide feedback both orally and in writing
  - Presentations are open for other teachers and students at the dept.

What is being assessed in the poster exam?

- The content: How structured and well-organised the content is (in the poster)
- Argumentation skills, reflective skills
- Participation in presentation

An example of assessment criteria for the poster exam

Grade 5/5:

- The poster is informative and clear
- The poster includes all the relevant concepts and their relations
- The literature has been used to support own thinking
- The relationships between phenomena are analysed in depth
- Different phenomena are analysed from different perspectives
- All the group members master the materials and have participated in preparing the poster
- All the group members participate in the presentation

Developing assessment of learning for own course

- Describe the intended learning outcome / learning objective
  - What is being assessed – knowledge, skills, attitudes

- Design how diagnostic, formative and summative assessment are aligned to support students to become aware of their prior knowledge, to get feedback of learning process, to get the grade (summative) based on transparent and fair assessment.

Closing the workshop

- Sharing learning experience:
  - Write 3-5 ideas you have got during workshops how to develop your teaching and assessment practices. (3-5 min.)
  - Discuss in groups
  - Share with other groups: Each group reports first the most common idea, then second common…
- General discussion and farewell

Lähteitä

- Benett, Y. (1993). The Validity and Reliability of Assessments and Self-assessments of Work-based Learning. Assessment & Evaluation in Higher Education, 18 (2).
- Birenbaum, M., & Dochy, F.J.R.C. (Eds.)(1996). Alternatives in assessment of achievements, learning processes, and prior knowledge. Dordrecht: Kluwer.
- Hailikari, T., Nevgi, A., & Lindblom-Ylänne, S. (2007). Exploring alternative ways of assessing prior knowledge, its components and their relation to learning outcomes. A mathematics-based case study. *Studies in Educational Evaluation*, 33 (3–4), 320–337.
- Knight, Peter (2002). Summative Assessment in Higher Education: practices in disarray. *Studies in Higher Education, 27*(3), 275–286.
- Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65:S63–S67.
- Reddy, Y. M., & Andrade, H. (2010). Review of the rubric use in higher education. *Assessment & Evaluation in Higher Education, 35*(4), 435-448.
- Sambell, K., McDowell, L., & Brown, S. (1997). "But is it fair?" : an exploratory study of the consequential validity of assessment. *Studies in Educational Evaluation, 23*(4), 349-371.
- Van der Vleuten, C. (2000). Validity of final examinations in undergraduate medical training. BMJ, 321, 1217-1219.