ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Department of Education

CHINLONE

Co-funded by the
Erasmus+ Programme
of the European Union

WP2 Teaching Quality Assurance

Assessment Methods

Yangon, Myanmar, 8-11 July 2019

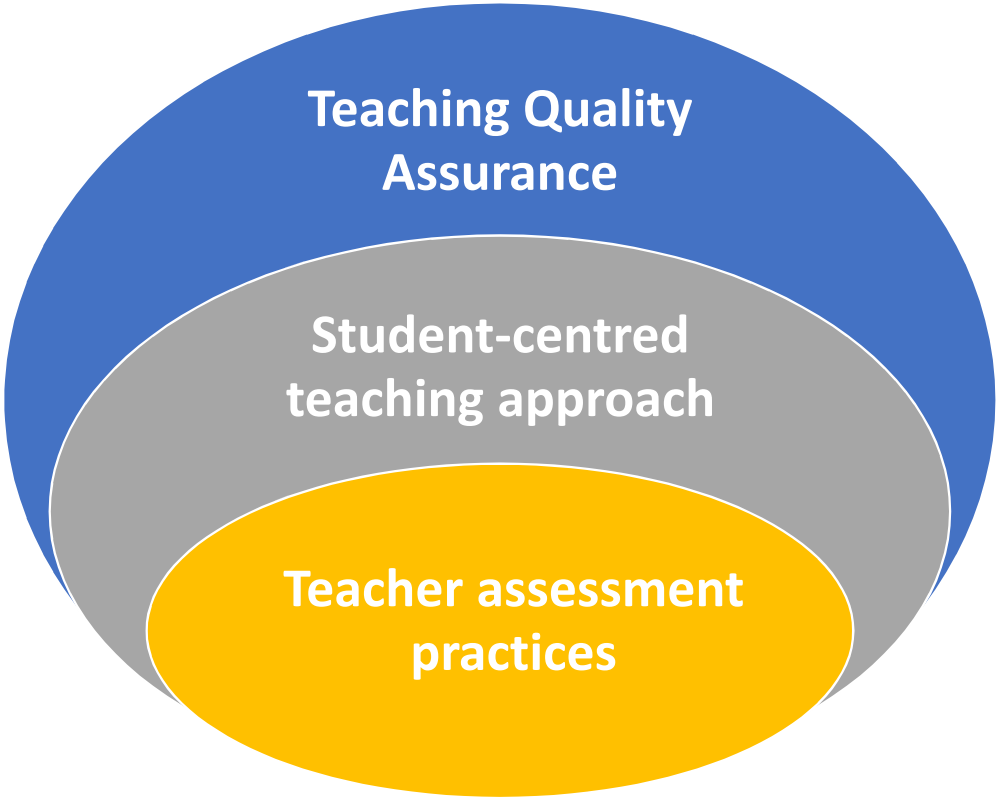
Andrea Ciani & Alessandra Rosa

Pilot Training Agenda

July 8th	Assessment: history of docimology, definitions, functions and common distortions
July 9th	Assessment procedures and tools: typologies, metrological characteristics, indications for planning, construction and analysis
July 10th	Formative assessment: strategies of assessment for learning and for quality teaching
July 11th	Authentic assessment: definition and construction of authentic tasks

Why assessment?

Why the need to analize and critically reflect on assessment in Higher Education?



Some problems in the assessment procedures in HE

- Lack of matching/alignment between assessment methods and learning objectives
- Assessment criteria not clearly expressed and shared with students
- Poor variety in the types of assessment tests (prevalence of written tests, multiple-choice questions)
- Inadequate or superficial feedback to students
- ……

William Thompson, 2010

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Department of Education

CHINLONE

Co-funded by the
Erasmus+ Programme
of the European Union

WP2 Teaching Quality Assurance
Assessment Methods

Assessment:
history of docimology, definitions, functions and common distortions

Learning assessment as a field of study and research: origins and developments of docimology

DOCIMOLOGY

Specific disciplinary field of the educational sciences focused on the issues and problems concerning the evaluation of students' learning

From the Greek **δοκιμ(άζω)** [dokimazo] «to examine, evaluate, estimate» and **λόγος** [logos] «thought, speech»

Andrea Ciani & Alessandra Rosa - University of Bologna

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Origins of docimology

CONTEXT

The docimological research has its roots between the late nineteenth and early twentieth century, when phenomena such as industrialization and the increase in schooling also at the level of secondary education highlight assessment issues that in conditions of more limited schooling had not emerged with the same urgency

PROBLEMS

Poor **validity** and **reliability** of the assessment procedures in the final exams
Interference of subjective elements in examiners' judgments
Need for more objective and transparent forms of assessment

Origins of docimology

From the need for a critical and rigorous analysis of the assessment methods in the final examinations of the secondary school, a specific sector of educational research originated which was called "docimology", initially understood as a discipline

«*which has as its object the systematic study of examinations, in particular the grading systems and the behavior of examiners and examinees*» (De Landsheere, 1971)

The term was introduced by the French scholar **Henri Piéron**, whose empirical studies on assessment in examinations are "milestones" in the field of docimological studies

H. Piéron and the empirical study on the *Baccalauréat* (1936)

Assessment of written examinations: differences between assessors

(*rating on twenty point scale*)

	Average gap	Most frequent gap	Highest gap
Essay in French	3,3	7	13
Latin translation	3,0	5	12
Essay in English	2,2	4	9
Maths' test	2,1	4	9
Philosophical dissertation	3,4	7	12
Physics test	1,9	4	8

Docimological research on school examinations and grades: principal findings

Lack of transparent and shared assessment criteria
Subjectivity of judgments

- There are significant discrepancies between the grades given to the same performance not only by different assessors, but also by the same assessor at different times
- Each assessor tends to give particular importance to certain aspects of student performance, while it tends to neglect others
- The ranges of grades used by different assessors are different (different interpretations of the same grading scale)
- Disagreement between assessors concerns not only the evaluation phase (attribution of marks), but also that of the correction (type and number of errors considered)

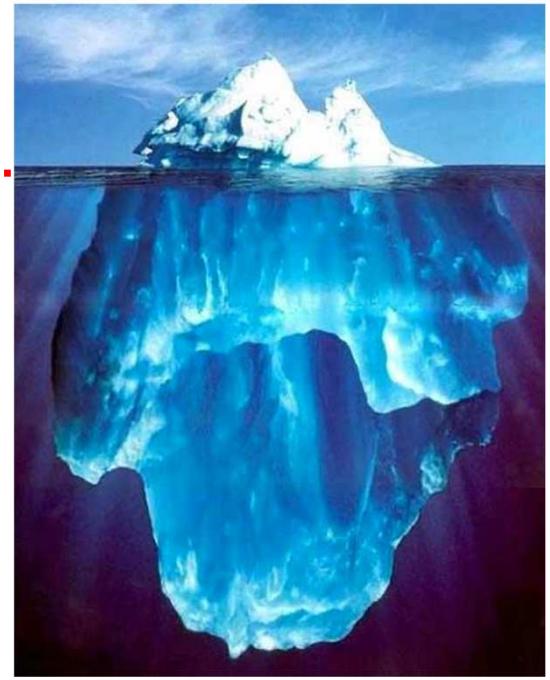
Subjectivity in the assessment: distortions or bias

Assessment distortions or bias are linked to teacher conceptions and beliefs that often implicitly affect teaching behavior, specifically assessment processes

The relationship between thought and action of the teacher: a model by Clark & Peterson (1986)

Visible part

Part not directly visible (submerged)



Teacher actions and their observable effects

Thought processes
Tacit and implicit conceptions and beliefs

Subjectivity in the assessment: main distortions or bias

- **Pygmalion effect**: influence of performance expectations (self-fulfilling prophecy) https://www.youtube.com/watch?v=4aN5TbGW5JA
- **Halo effect**: tendency for an impression created in one area to influence opinion in another area (e.g. irrelevant elements are decisive in the judgment)
- **Contagion effect**: influence of the judgment of others on the evaluation («What is your average grade?»)
- **Stereotypy effect**: strong influence of previous judgments (assessment fixity)
- **Succession/Contrast effect**: over or underestimate of a performance based on a comparison with another performance
- **Forced distribution effect**: forcing of individual differences, assimilation to normal distribution (bell curve)

Developments of docimology

From the **criticizing phase** to the **constructive phase**

Considering the problems identified, how the procedures and tools for assessing students' learning can be improved**?**

- emphasis on objective or structured tests
- indications to make traditional assessment procedures more valid and reliable
- proposal of new assessment strategies and tools, alternatives to both testing and traditional methods

Developments of docimology

Enlargement of the field:
rethinking assessment in a systemic and multidimensional perspective

Change in the definitions of docimology

From … ⇢ **To …**

"Systematic study of examinations, in particular of the grading systems and the behavior of examiners and examinees"

De Landsheere, 1971

"Study of assessment systems in education"

De Ketele, 1982

"Educational assessment science"

Vertecchi, 1993

Developments of docimology

Extension of the concept of assessment in multiple directions

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Department of Education

CHINLONE

Co-funded by the
Erasmus+ Programme
of the European Union

WP2 Teaching Quality Assurance

Assessment Methods

Assessment Procedures and Tools: typologies, metrological characteristics, indications for planning, construction and analysis

Design and implementation of correct assessment procedures: the operations or phases of the evaluation process

"The evaluation is articulated in three moments, which are to be identified in the ***choice of objectives to be assessed****, in the operations to be carried out to* ***ascertain the chosen objectives****, and in the* ***judgment to be given on the results*** *of the ascertainment made"*

Gattullo (1968)

Three operations or phases of the evaluation process

1. **Clarify what you intend to assess**: clear definition of the learning outcomes to be measured and evaluated

2. **Measure the identified learning outcomes**: intentional and planned collection of valid and reliable informations related to the identified learning outcomes allowing to discriminate different levels of performance

3. **Evaluate the learning outcomes you measured**: interpretation of the measurement results and expression of an overall and synthetic judgment

NO impressionistic, intuitive, and subjective evaluation (distorsions)
BUT judgment supported by pertinent and accurate informations!

1 Clarify what you intend to assess

Identify in a clear and unambiguous way the object of the assessment

The object of the assessment must be **formulated in operational terms**, that is in terms of **observable, measurable and verifiable indicators**

What must the student do in order to prove that he/she has achieved the objectives assessed?

What are the observable performances that demonstrate the achievement of the objectives?

1

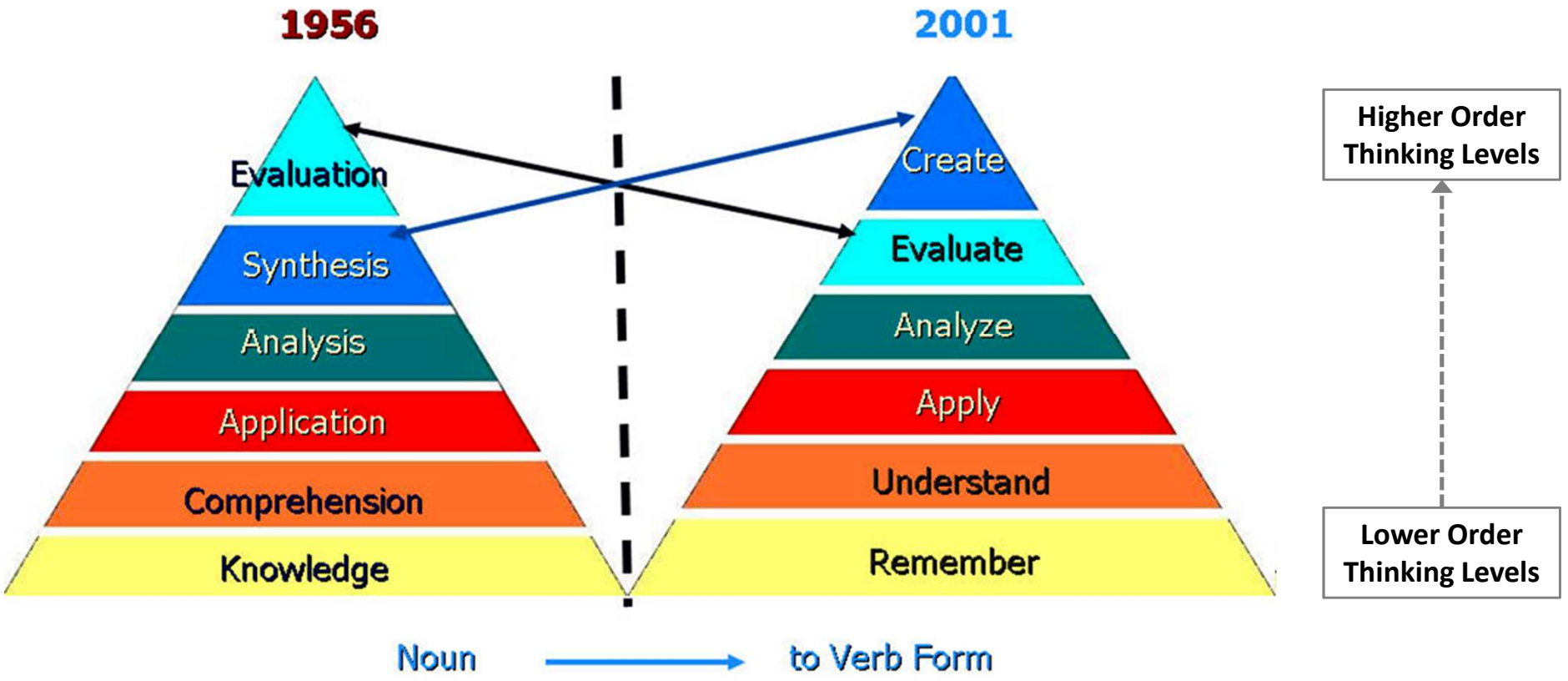
A guide to formulating goals: the TAXONOMIES of educational goals

Taxonomies are a tool that can help to define and operationalize educational goals

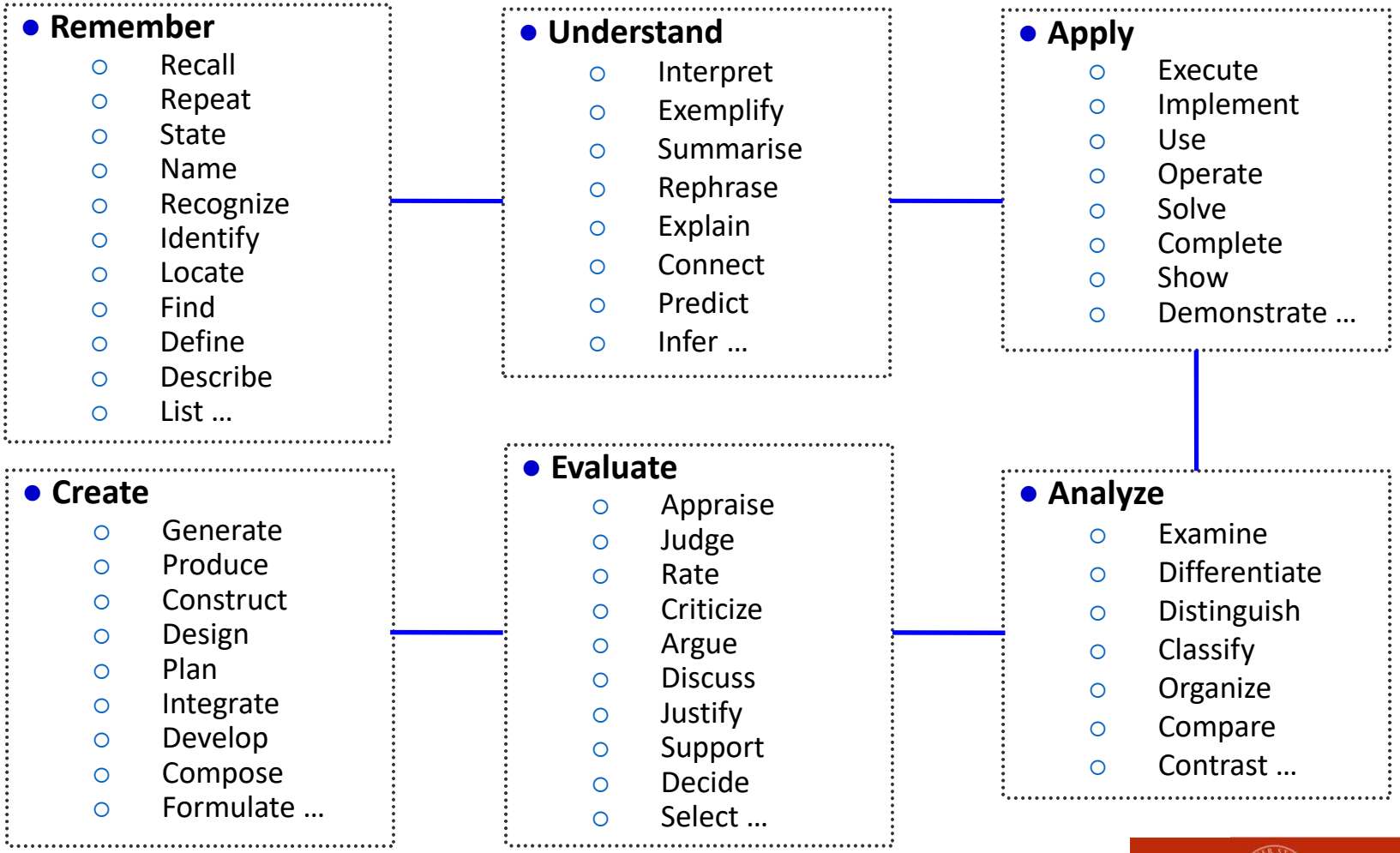
They distinguish different levels of educational objectives arranged hierarchically from the simplest to the most complex

1

The taxonomy of **Bloom (1956)** and the version revised by **Anderson & Krathwohl (2001)**



1 Some verbs to specify learning objectives at each level of the taxonomy



Our learning outcomes (Course objectives) – Lessons 1 & 2

1. Identifying the principal evaluation distorsions
2. Understanding the influence of teacher beliefs on evaluation practice
3. Simulating a real situation of evaluation distorsion
4. Identifying the three phases of evaluation process
5. Knowing the characteristics of different assessment tools
6. Using the measurement procedures to assess a semi-structured test, including the choice of the evaluation criteria (absolute and relative)
7. Creating a structured test aligned with specific LO according to A&K Taxonomy and indications for building test items

1

Constructive Alignment Model (Biggs, 1999)

Intended Learning Outcomes (ILOs)

What should the students know and be capable of doing after having completed this unit/course/program?

Assessment Methods

What kinds of assessment methods will allow ILOs to be tested and evaluated?
What kinds of assessment tasks will allow students to demonstrate the extent of ILOs achievement?

Teaching & Learning Activities

What kinds of TLAs will support students in achieving ILOs?

Examples of assessment tasks aligned with learning outcomes at different levels of the taxonomy

Remembering	Identify the correct definition of the term XXX Give a brief definition of the concept of XXX List the main features of .... Arrange the following events in chronological order Name the authors who have developed the following theories
Understanding	Give an example of ... Explain why there is an association between ... Summarize the main consequences of ... Identifies the rule that was applied in the following example
Applying	Solve the following problem / exercise Apply the correct procedure to execute ... Use the data provided to demonstrate ...
Analysing	Identify the similarities and differences between ... Organize the information given based on the following categories ... Distinguish the general concepts from those subordinate to them
Evaluating	Explain if in your opinion the following procedure is adequate for the purpose of ... Find the errors and inconsistencies in the following text Choose the best solution to the problem from those proposed and explain why you chose it Order the XXX (from the best to the worst) based on the degree to which they meet the following criteria
Creating	Formulate a hypothesis to explain the following event Plan a project to build ... Look at the following XXX and produce a better one that meets the same goals Compose a text / image to convey the following message

COMPETENCES TO DEMONSTRATE IN HIGHER EDUCATION STUDIES

1. Demonstrating knowledge and understanding
2. Accessing and managing information
3. Solving problems and developing plans
4. Performing procedures and demonstrating techniques
5. Communicating
6. Managing and developing oneself
7. Thinking critically and making judgements
8. Designing, creating, performing

From: Nightingale, P., Te Wiata, I.T., Toohey, S., Ryan, G., Hughes, C., Magin, D. (1996). *Assessing Learning in Universities*. Professional Development Centre, University of New South Wales, Australia

Selecting appropriate assessment methods according to learning outcomes

Broad categories of Learning Outcomes	Methods of Assessment
Demonstrating knowledge and understanding (Recalling, describing, reporting, recounting, recognising, identifying, relating & interrelating)	Written examination Oral examination Essay Report ….
Accessing and managing information (Researching, investigating, interpreting, organising information, reviewing and paraphrasing information, collecting data, searching and managing information sources, observing and interpreting)	Annotated bibliography Project Dissertation Applied problem ….
Solving problems and developing plans (Identifying problems, posing problems, defining problems, analysing data, reviewing, designing experiments, planning, applying information)	Group Work Work-based problem Draft a research plan Analyse a case ....
Performing procedures and demonstrating techniques (Computation, taking readings, using equipment, following laboratory procedures, following protocols, carrying out instructions)	Demonstration Role play Prepare and deliver a peer teaching session Produce a poster Lab report …

From: Nightingale, P., Te Wiata, I.T., Toohey, S., Ryan, G., Hughes, C., Magin, D. (1996). *Assessing Learning in Universities*. Professional Development Centre, University of New South Wales, Australia

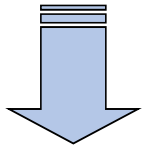
Selecting appropriate assessment methods according to learning outcomes

Broad categories of Learning Outcomes	Methods of Assessment
Communicating (One and two-way communication; communication within a group, verbal, written and non-verbal communication. Arguing, describing, advocating, interviewing, negotiating, presenting; using specific written forms)	Written presentation (essay, report, reflective paper etc.) Oral presentation Group work Discussion/debate/role play ….
Managing and developing oneself (Working co-operatively, working independently, learning independently, being self-directed, managing time, managing tasks, organising)	Journal Portfolio Group work Peer/self-assessment of contribution to group work ….
Thinking critically and making judgments (Developing arguments, reflecting, evaluating, assessing, judging)	Essay Report Journal Write a newspaper article Comment on an article's theoretical perspective ….
Designing, creating, performing (Imagining, visualising, designing, producing, creating, innovating, performing)	Portfolio Performance Presentation Projects ….

From: Nightingale, P., Te Wiata, I.T., Toohey, S., Ryan, G., Hughes, C., Magin, D. (1996). *Assessing Learning in Universities*. Professional Development Centre, University of New South Wales, Australia

2 Measure the identified learning outcomes

Collection of valid and reliable informations on the learning outcomes previously identified using appropriate tools and procedures



Selection and construction of the tests

Use of the tests (administration to students)

Correction and scoring of the tests ⟶ assigning score to each item/question

B) *READ THE PARAGRAPHS AND ANSWER THE QUESTIONS (13 points)*

LAURA'S DAY

My day begins at 7:30, when my mother wakes me up. I get up at 7:35 and I take a shower. I get dressed, then I have breakfast with my family. That's at about 7:45. I leave home at 8 o'clock and catch the bus. I get to school at about 8:20. Our lessons begin at 8:30, and I always arrive on time. At 10 o'clock we have a 20-minute break. After that, lessons continue till our lunch break at 11:50. We have about 30 minutes for lunch, and I eat in the school canteen. We finish school at 15:15 and I go back home.

I arrive home at about a quarter to four. When I get home, I do my homework. But before that, I drink something: usually it's a glass of milk. Before dinner I do my homework. We always have dinner at 7:30. I usually help my mother with the washing up. My mother is a teacher. After dinner, she's always busy - she corrects her students' work and prepares her lessons. Then I watch TV and go to bed.

1) What time does Laura have breakfast? *(2 points)*
   - **a)** at a quarter to eight
   - **b)** in quarter past eight
   - **c)** in fifteen to eight
   - **d)** on eight fifteen

2) How does Laura get to school? *(2 points)*
   - **a)** on bus **b)** with a bus
   - **c)** with bus **d)** by bus

3) Another way of saying 'continue' is: *(2 points)*
   - **a)** go on **b)** go past
   - **c)** go in **d)** go you

4) How long is her lunch break? *(2 points)*
   - **a)** half hour **b)** a hour half
   - **c)** half past **d)** half an hour

5) Another way of saying 'I arrive home' is: *(2 points)*
   - **a)** I go the home **b)** I go to the home
   - **c)** I get home **d)** I get to home

6) Fill each space with one of these words: **Finally, First, Then**. *(3 points)*

   …… , Laura has a glass of milk.

   After that, she does her homework.

   …… she has dinner.

   …… , she watches TV and goes to bed.

VALIDITY and RELIABILITY

The essential metrological characteristics to guarantee the quality of the measuring instruments and the information they allow to collect

Measurement quality

VALIDITY

The instrument actually detects what it is intended to detect

Is the information collected relevant and representative with respect to the objects of the assessment?

RELIABILITY

The instrument provides stable results at different times and with different assessors

Are the measurements obtained accurate and free of errors/distortions?

VALIDITY and RELIABILITY

- the **validity** of a test is linked to the way in which the learning objectives to be measured with the test have been defined (clarity and specificity)

- the **reliability** of a test is linked to the predetermination of unambiguous criteria for the attribution of scores and to the control of potential sources of error relating to the measurement process

The reliability of an assessment tools is obtained not only by having an agreement on the measurement scales (low level of interpretability and therefore of subjectivity), but also by controlling the phases of construction and use of the instrument

Benvenuto (2003)

3 Evaluate the learning outcomes you measured

Interpretation of the measurement results and expression of a judgment

- using a reference **CRITERION**, that is a parameter with which to compare the measurement results
- using a specific **LANGUAGE**, that is conventional and understandable symbols/terms (eg numerical grades, adjectives, letters ...)

The evaluation CRITERIA

ABSOLUTE CRITERION

(***criterion-referenced test* o *standard-based test***)

comparison with a predefined standard,
with a predefined level of mastery or performance

RELATIVE CRITERION

(***norm-referenced test***)

comparison with the performance of other students,
with the performance of a reference group

3

The acceptability or sufficiency threshold (*cut-off score*)

It marks the passage from a positive evaluation ("passed" test) to a negative evaluation ("not passed" test) → it makes it possible to distinguish between those who have achieved the objectives being assessed at a sufficient level and those who have not achieved them

Absolute criterion
(*criterion-referenced or standard-based test*)

*Threshold established **a priori***

eg. *60% of the total score*

Relative criterion
(*norm-referenced test*)

*Threshold established **a posteriori***

eg. *Average level of the reference group*

CIRCLE TIME

What kind of assessment methods/tools do you use in your teaching?

Why do you use that kind of assessment methods/tools ?

Types of learning assessment tools

To measure and evaluate learning it is possible and appropriate to use a variety of tools

A criterion that can be used to classify the different types of tests consists in their **level of structuring**, given by the type of provided stimulus and expected response

STIMULUS

The type of solicitation provided to the student

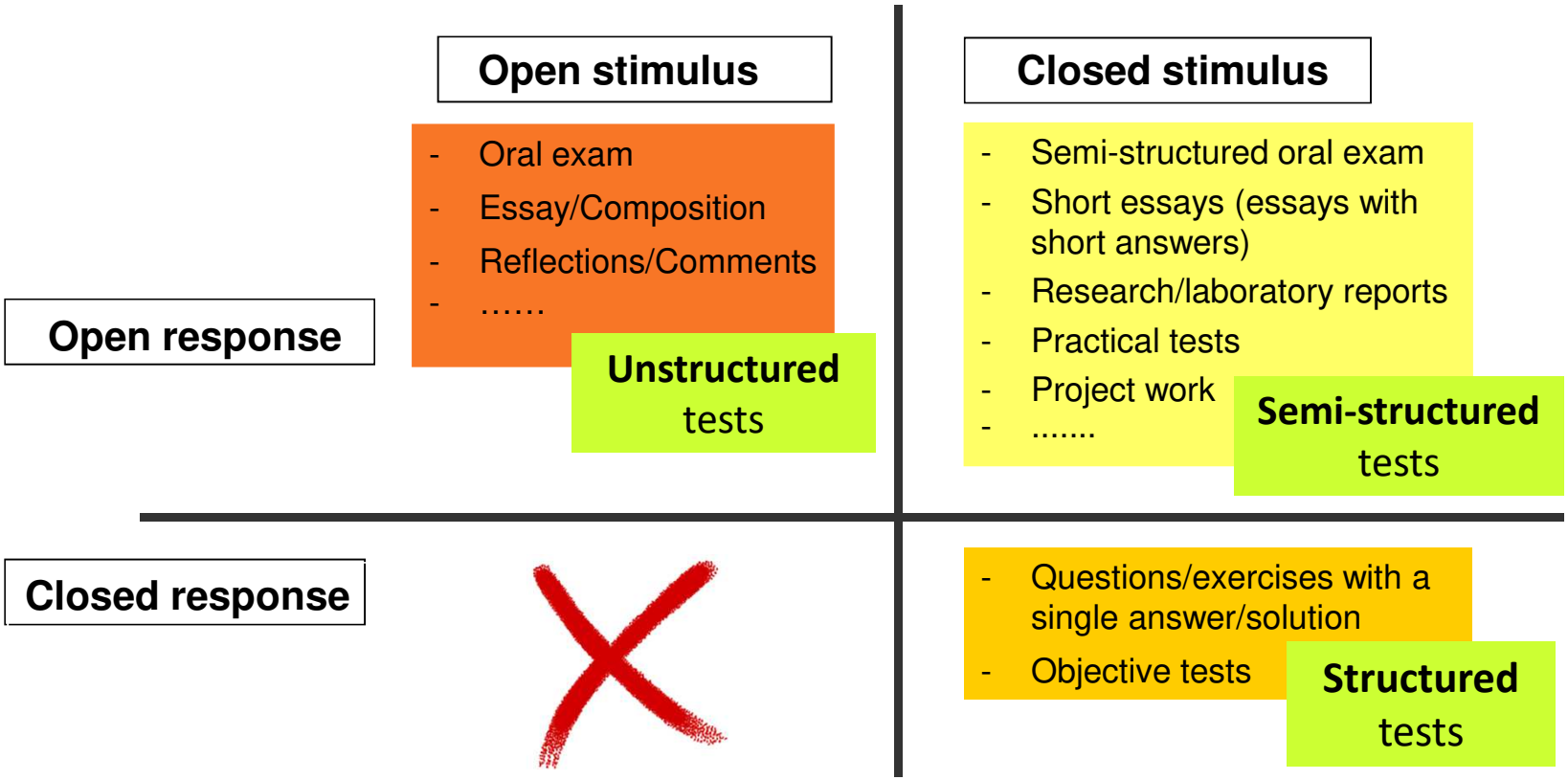
OPEN **CLOSED**

RESPONSE

The type of answer requested to the student

OPEN **CLOSED**

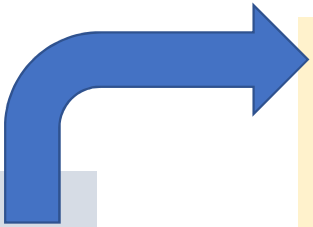
A classification of the types of tests



How to choose?

There are no better or worse instruments at all,
but only instruments more or less suitable for the purposes of assessment and the learning outcomes to be assessed

WHAT LEARNING OUTCOMES?



WHAT KIND OF TEST?

Learning assessment tools:

the objective or structured tests

Characteristics of the OBJECTIVE TESTS

The objective tests consist of a set of structured questions (items) with a predefined answer

The correct answers to the questions and the scoring criteria are predetermined

Therefore the problems connected to the influence of the subjectivity are eliminated:

anyone who corrects/evaluate the test should get the same results!

PLANNING of the test

Planning the test means taking a series of decisions that lead to defining and specifying different aspects

- ✓ **Test objectives**: identify and define in operational terms the learning outcomes to be measured
- ✓ **Test format**: define the type and number of structured questions to be used
- ✓ **Test scoring and test evaluation criteria**: define how to score the test items and how to interpret the scores (how to pass from scores to judgments/grades)

The different types of structured questions

Four main types

- **MULTIPLE CHOICE**
- **TRUE/FALSE**
- **COMPLETIONS (CLOZE TEST)**
- **MATCHES**

Es. Multiple choice

The VALIDITY of a test refers to	**General part of the item**
a) the degree of precision in the measurement**.** **b)** the degree to which it measure what you want to measure. **c)** the degree of agreement between different measures. **d)** the degree of objectivity in the measurement.	**Alternatives (correct answer and distractors)**

Es. True/False

A *norm-referenced test* is based on an absolute standard as a criterion for evaluating student results

V F

Es. Cloze test

Among the sciences of education, the docimology is a specific discipline that deals with the issues of the _____ . Early research addressed the problem of _____ of the judgments in the final school exams. Studies carried out in this field have helped to highlight various types of evaluation _____ , such as the halo, stereotypy or succession effects. Especially in a first phase, great emphasis was given to the _____ tests as they allow greater _____ in assessment: the correct answers, in fact, are predetermined.

1.instruction *2.structured* *3.authentic* *4.procedures* *5.objectivity*
6.subjectivity *7. evaluation* *8.easiness* *9.distorsions* *10.communication*

Es. Matches

Types of item

a) _____ true/false

b) _____ multiple choice

c) _____ matches

d) _____ completions

Modality of response

1. choice between different answers of which only one is right
2. connection between different lists of elements
3. choice between different answers on more lists of elements
4. identification of the elements eliminated in a sentence or text
5. completion of a list of answers
6. choice of the correct answer between two alternatives

The different ways of assigning scores

- Assignment of the same weight to each question → eg *1 point for each correct answer, 0 points for errors and omissions*

- Attribution of different weights to the different questions based on:
  - the complexity/difficulty of the items → eg *1 point for questions that measure mnemonic/reproductive abilities, 2 or more points for those that measure more complex abilities*
  - the probability of responding well randomly, "trying to guess" → eg *1 point for true/false items, 2 or more points for multiple-choice items*

- Penalization of errors → to discourage answers given at random, errors can be penalized (eg *-0.2 point*)

CONSTRUCTION of the test

QUESTIONS

Items must be relevant and representative with respect to the previously defined objectives of the test

They must also be formulated following some indications (emerging from the docimological research) aimed at increasing the validity and reliability of the measurement

FRONTISPIECE

It must contain clear and concise indications on:

- the objectives of the test
- how to respond and how to correct a given answer
- the time available to carry out the test
- how the scores are attributed

LAYOUT

Regarding the layout of the test, the following suggestions can be considered:

- avoid "breaking" the text of a question into two different pages
- leave adequate space between the various questions and present a clean, orderly and formally correct product

Some indications for the construction of multiple-choice questions

- Each item must propose, in the general part, only one "problem"
- Each item must be independent of the others
- Items must be concise and essential and the language clear and simple (avoid superfluous details, ambiguous terms, complex propositions)
- There must be logical, grammatical and syntactic continuity between the general part of the item and each alternative
- In the construction of items it is necessary to avoid, as far as possible, to use negative formulations (if used, underline them)
- All the elements common to the alternatives must be introduced, as far as possible, in the general part of the item
- The optimal number of answer alternatives is 4 or 5
- None or all alternatives must repeat concepts or terms of the general part
- Alternatives must be homogeneous in terms of construction, type of language and length
- Wrong alternatives (*distractors*) must be plausible
- The correct answers must be distributed among all the alternatives
- It is good to avoid using response alternatives such as "all previous" or "none of the above"
- Some orthographic conventions must be respected (eg. if the question is based on the "question-answer" scheme the alternatives must start with a capital letter, if it is based on the "completion" scheme with the lower case letter)

BRAINSTORMING

What are the strengths and weaknesses of the objective tests?

The debate on objective tests

ADVANTAGES or STRENGTHS

- Possibility to submit a large number of questions in a short time
- Uniformity of examination conditions among students
- Quick and easy correction of replies
- Univocal interpretation of results, without subjective distortions

The debate on objective tests

LIMITS or WEAKNESSES

- Planning and development of valid and reliable tests requires time and expertise
- Only certain aspects or dimensions of learning can be measured (not all types of learning outcomes)
- It is possible to respond well by chance, guessing

The debate on objective tests

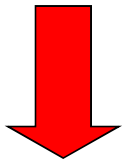
PREJUDICES

- They encourage mnemonic learning, based on superficial factual knowledge
- They only measure reproductive abilities
- They can only be useful in a summative evaluation perspective (focus on the product and not on the learning process)

Beyond the prejudices

Objectives or structured tests

- encourage mnemonic learning, based on superficial factual knowledge
- only measure reproductive abilities



It depends on how the tests are planned and built!

Well formulated questions - especially multiple-choice items - give the possibility to assess a wide range of abilities and can solicit very different operations, from those that require only to remember or recognize to the more complex ones that involve connections, comparisons, inferences, interpretations ...

Es. International survey OCSE-PISA

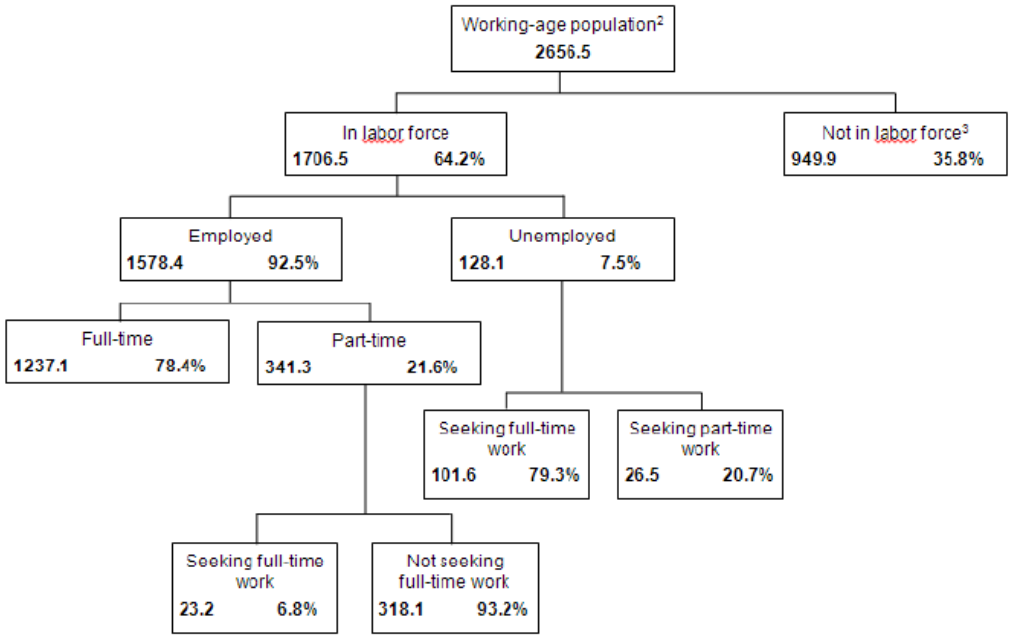
(*Programme for International Student Assessment*)

Reading literacy

The tree diagram below shows the structure of a country's labor force or "working-age population". The total population of the country in 1995 was about 3.4 million.

LABOR

The Labor Force Structure year ended 31 March 1995 (000s)[1]



Notes

1. Numbers of people are given in thousands (000s).
2. The working-age population is defined as people between the ages of 15 and 65.
3. People "Not in labor force" are those not actively seeking work and/or not available for work.

Source: D.Miller, *Form 6 Economics*, ESA Publications, Box 9453, Newmarker, Auckland, NZ, p.64.

QUESTION 3: LABOR *R088Q04*

In which part of the tree diagram, if any, would each of the people listed in the table below be included?

The first one has been done for you.

	"In labor force: employed"	"In labor force: unemployed"	"Not in labor force"	Not included in any category
A part-time waiter, aged 35	☒	☐	☐	☐
A business woman, aged 43, who works a sixty-hour week	☐	☐	☐	☐
A full-time student, aged 21	☐	☐	☐	☐
A man, aged 28, who recently sold his shop and is looking for work	☐	☐	☐	☐
A woman, aged 55, who has never worked or wanted to work outside the home	☐	☐	☐	☐
A grandmother, aged 80, who still works a few hours a day at the family's market stall	☐	☐	☐	☐

Es. International survey OCSE-PISA

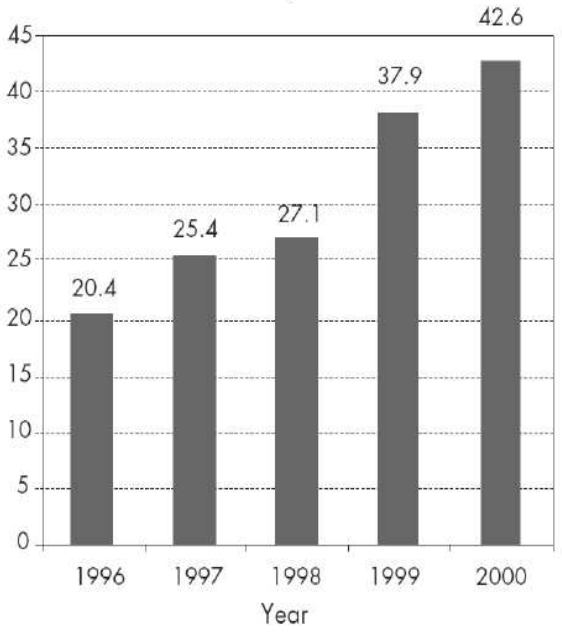
(*Programme for International Student Assessment*)

Mathematics literacy

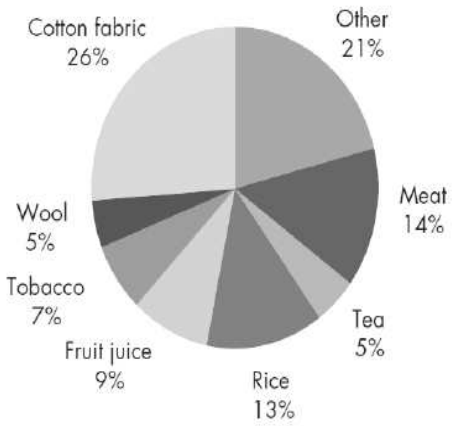
EXPORTS

The graphics below show information about exports from Zedland, a country that uses zeds as its currency.

Total annual exports from Zedland in millions of zeds, 1996 to 2000



Distribution of exports from Zedland in 2000



Question 2: EXPORTS

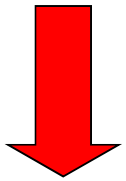
What was the value of fruit juice exported from Zedland in 2000?

A 1.8 million zeds.
B 2.3 million zeds.
C 2.4 million zeds.
D 3.4 million zeds.
E 3.8 million zeds.

Beyond the prejudices

Objectives or structured tests

- can only be useful in a summative evaluation perspective (focus on the product and not on the learning process)



Formative use of the tests is also possible!

The results can be the starting point for a reflection/discussion on the test, on the questions that have created more difficulties and on the type of skills they aim to detect, on the thought processes underlying the answers given by the students → feedback with regulatory and self-regulative function

https://kahoot.com/

Flexible and simple tool for real time learner participation and classroom feedback

Andrea Ciani & Alessandra Rosa - University of Bologna

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Learning assessment tools:

the semi-structured tests

Characteristics of SEMI-STRUCTURED TESTS

The semi-structured tests require an open answer that the student must elaborate respecting some constraints set by the stimulus, which allow the comparison with predetermined correction criteria

The stimulus proposes a specific request that:

- solicits the demonstration of specific knowledge, skills and competences
- allows to limit the subjectivity of the judgment in the interpretation of the results

Advantages with respect to the structured and unstructured tests

This type of assessment tools is more suitable than structured tests to verify higher cognitive processes and divergent thinking skills, eg. the ability to apply the acquired knowledge/skills in real-life situations and contexts or the ability to find original solution to particular problems

The proposal of well-defined and specific stimuli facilitates the identification of evaluation criteria by defining the characteristics that the student's response must possess in order to be considered adequate. This allows to overcome some limits of unstructured tests, that is the difficulty of collecting valid and reliable information on students' learning (the proposed stimuli are usually too generic and poorly representative and the evaluation criteria are usually subjective and implicit)

PLANNING of the test

Planning the test means taking a series of decisions that lead to defining and specifying different aspects

- ✓ **Test objectives**: identify and define the learning outcomes to be measured
- ✓ **Test format**: choose the type of test to be used and prepare specific and well-defined requests and instructions
- ✓ **Performance assessment criteria**: define the reference standards for scoring and evaluation (indicators and descriptors)

INDICATORS and DESCRIPTORS

The **indicators** represent relevant and observable dimensions for the analysis and evaluation of specific learning objectives (*on which fundamental aspects we need to focus our attention to verify to what extent these objectives have been achieved?*)

The **descriptors** provide an operational definition of the indicators in terms of expected performance (*in relation to each indicator, how can an adequate performance be described? what are the essential points that, depending on the stimuli presented by the test, must be present in an adequate performance?*)

This descriptions thus becomes the criteria for scoring and assessing students' performance (the "models" of performance with which to compare the performance of the students)

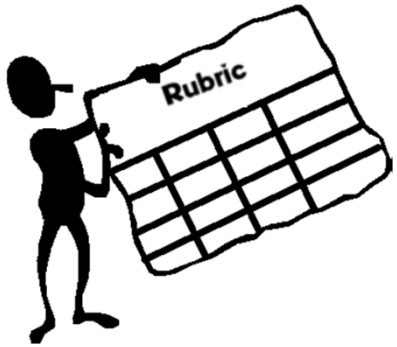
The assessment RUBRICS

What is a rubric?

A *rubric* is a tool that can be used to assess the quality of student performance in a semi-structured test

It can be defined as a description of different levels of performance useful to provide explicit, transparent and shared evaluation criteria

It is a double entry table where, for each indicator identified, a list of descriptors corresponding to different performance levels is defined



The key components of a rubric

INDICATORS	Relevant and observable dimensions for the evaluation of the performance
LEVELS	Different performance levels ordered in ascending or descending order
DESCRIPTORS	Description of the performance corresponding to each level of the scale
SCORES	Score corresponding to each level of the scale

	5	4	3	2	1
Thesis is Clearly Stated	Paper has a clear & intriguing thesis. Clearly stated in the introductory paragraph. The reader is eager to read more about the topic.	Paper states a thesis clearly within the first paragraph. Readeris able to make assumptions about the content.	Thesis is stated in the paper but it is difficult to find. Reader is not sure what this paper will be discussing.	Thesis is located somewhere in the paper. The reader is not clear what the writer is going to discuss.	What thesis?
Organization of Facts & Details Supporting Thesis	Paper is well organized. It reveals facts & details discovered during research. All of the facts & details support the thesis in an orderly manner. The reader reads on with joy.	Paper is organized. It reveals facts & details from research. Most information supports thesis and is logically stated.	Organization is rough but workable. Author reports on facts and details, but does not tie the information together. The information does not always support the thesis.	Writing is aimless and disorganized. It makes sense one moment and then loses focus. The facts & details do not support the thesis statement.	I'm suppose to be organized?
Interpreting & Analyzing	Writing sounds as if author wrote it. Research It reflects researched information reported in author's writing style. No plagiarism. Three references cited.	Writing mostly sounds as if author wrote it - reflects research information reported in author's writing style. No plagiarism. Some references citied.	Writing is bland or mechanical. It reflects researched information not written in author's own words. Some references cited.	There are a few parts where the author's voice is audible, but it sounds as if the researched information is being quoted. Less than 3 references.	I'm suppose to interpret and analyze?
Observations & Conclusions	Observations & conclusions are clearly stated and critical. Author has made personal connections to help the reader understand the significance of the topic.	Observations and conclusion are well written. There is some attempt to make personal connections.	Observations and conclusions are present, but lack depth & understanding. There is little attempt to make personal connections.	Observations are unclear. The reader is confused. There is an attempt to provide a conclusion; however, it is weak. The author makes no personal connections.	Minimal observations No conclusions
Bibliography	Bibliography is impeccably written. Author followed the handout.Nine or more sources. It looks professional.	Bibliography is well written. Author followed the rules in the handout. Nine or more sources. There are minor errors.	Bibliography is OK. There are nine sources. The author did NOT follow all the rules in the handout. There are some errors.	Bibliography is weak. Less than nine sources. The author did NOT reference all materials. Many errors.	What bibliography ?
Conventions	Author has used correct paragraph form - grammar, capitals, spelling and punctuation. There is evidence of proofreading.	Author made a few minor errors. Generally good paragraph form & mechanics. Author made a few errors in punctuation. Evidence of proofreading.	Author made some errors - reader is distracted by them. More proofreading was necessary.	The author made many errors. The number of errors made it difficult to read. The number of errors suggests no proofreading.	Spell check was a mystery button on the computer.

	Very Good	Good	Developing
Content	I have expanded and explained my topic by collecting lots of research from books and online. I didn't copy any any content from information that I found online, in books or from friends.	I sometimes expanded and explained my topic by collecting some research from books and online. I copied some content from information that I found online, in books or from friends.	I didn't expand or explain my idea so much, because I did very little research. I copied lots of content from information that I found online, in books or my friends.
Structure Main idea	My essay has a clear introduction, main body sections and conclusion. Each paragraph has a clear main idea. Each main idea has supporting sentences	My essay doesn't have a clear introduction, main body or conclusion. Some paragraphs don't have a clear main idea. Some main ideas don't have supporting sentences.	My essay has only one big paragraph. The introduction, main body and conclusion are difficult to see. There are too many main ideas and not many supporting sentences. Let's work on planning your essay more.
Grammar & Sentence structure	The sentences are clear and easy to understand. Each sentence contains one idea. There are few grammatical mistakes.	Some of my sentences are unclear and a little difficult to understand. Some sentences contain more than one idea. There are some grammatical mistakes.	The sentences are often unclear and difficult to understand. The main idea is unclear. There are many grammatical problems, this means it is difficult to read and understand the essay.
Vocabulary & Punctuation	I didn't repeat the same words so much. I often put full stops, with capitals and commas in the right place.	I sometimes repeated the same words. I sometimes put full stops with capital letters and commas in the wrong places.	I used the same words often. I didn't put full stops or commas in the right place, or aren't used at all.

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Department of Education

CHINLONE

Co-funded by the
Erasmus+ Programme
of the European Union

WP2 Teaching Quality Assurance
Assessment Methods

Formative Assessment:
strategies of assessment for learning and for quality teaching

Formative assessment

Incorporated into teaching practice. Allows for monitoring of student progress throughout the learning process in order to identify learning needs and adjust teaching appropriately

Focus on the **process** (*what and how to improve in order to achieve the established learning goals?*)

Ongoing regulatory function with a view to promoting continuous improvement of the teaching and learning processes (*internal relevance*)

- To diagnose strengths and weaknesses of students
- To regulate teaching and learning processes
- To provide feedback to students to improve their learning
- To help students to develop their self assessment skills

Summative assessment

Takes place at the end of the teaching and learning process to verify its effectiveness by assessing and grading students' achievement of intended outcomes

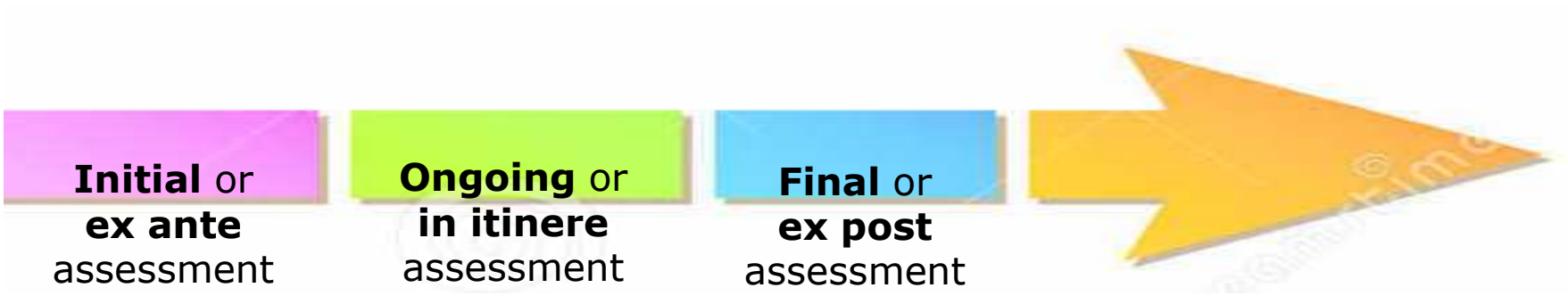
Focus on the **product** (*to what extent the established learning goals have been achieved?*)

Final verification and certification function with a view to formulating a judgment on the results of the teaching and learning processes (*external relevance*)

- To pass or fail students
- To grade or rank students
- To select for future courses
- To predict success in future courses

Extension of assessment TIMING

Assessment takes place at different times along the educational path (not just at the end)



Diagnostic function (to look for student predictable difficulties, to design the educational path)

Regulatory function (to check of student progress, to adjust the educational path)

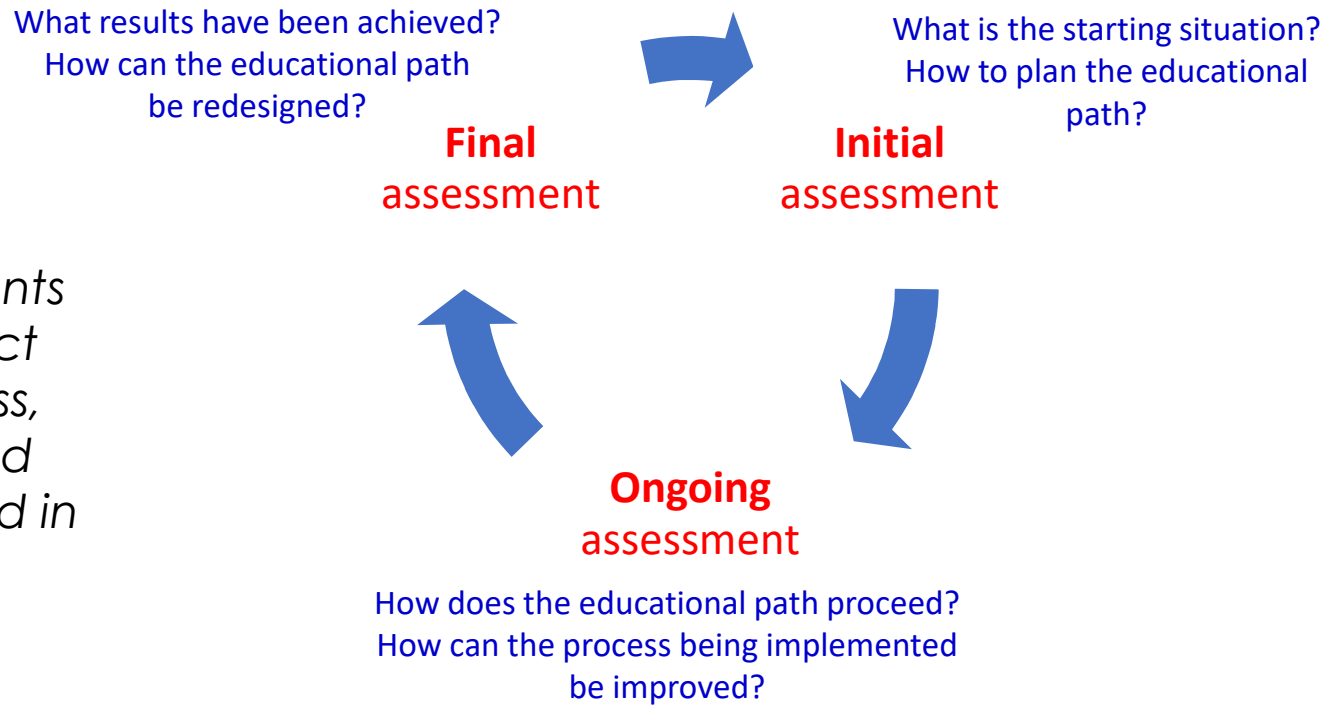
Summative function (to verify and certify student achievements, to evaluate the effectiveness of the educational path)

Assessment as a continuous and cyclical process (*assessment loop*)

Evaluation as

"*circular flow of judgments that invest every aspect of the teaching process, both in its planning and organization phases, and in those of realization*"

Vertecchi, 2003



Definitions of Formative Assessment (FA)

"Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes"

McManus (2008)

"Assessment practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited"

Black & Wiliam (2009)

Current interpretation of FA: Assessment for Learning (AfL)

Current interpretations and exemplifications of formative assessment are often linked to the concept of ***Assessment for Learning*** proposed by the UK **Assessment Reform Group** (2002) and associated work by Black & Wiliam (e.g. 1998, 2009) and colleagues

Assessment **for** Learning
(*formative assessment*)

Assessment **of** Learning
(*summative assessment*)

Definitions of Assessment for Learning (AfL)

"*Assessment for Learning is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there*"

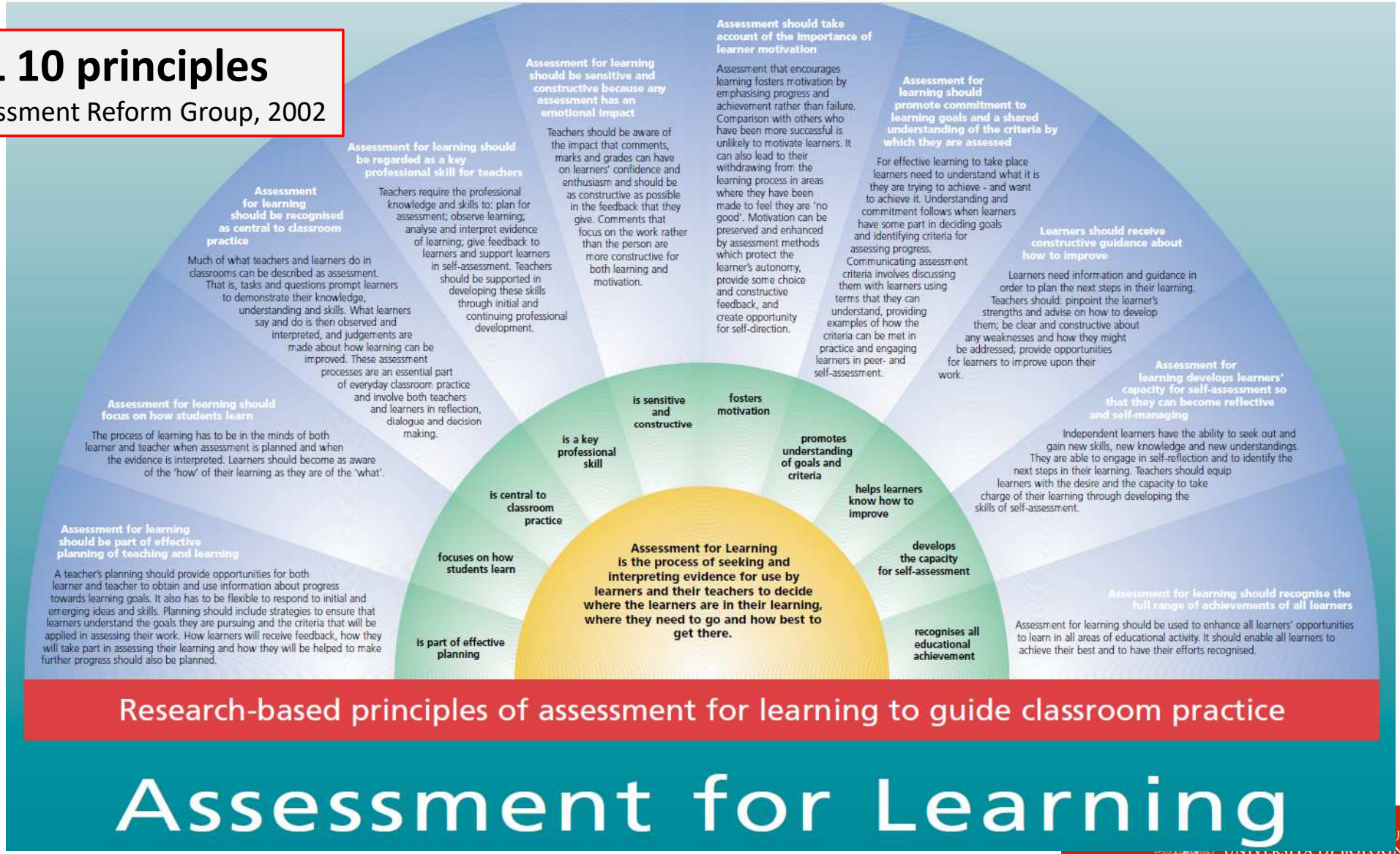
Assessment Reform Group (2002)

"*Assessment for Learning is part of everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning*"

Black & Wiliam (2009)

AfL 10 principles

Assessment Reform Group, 2002

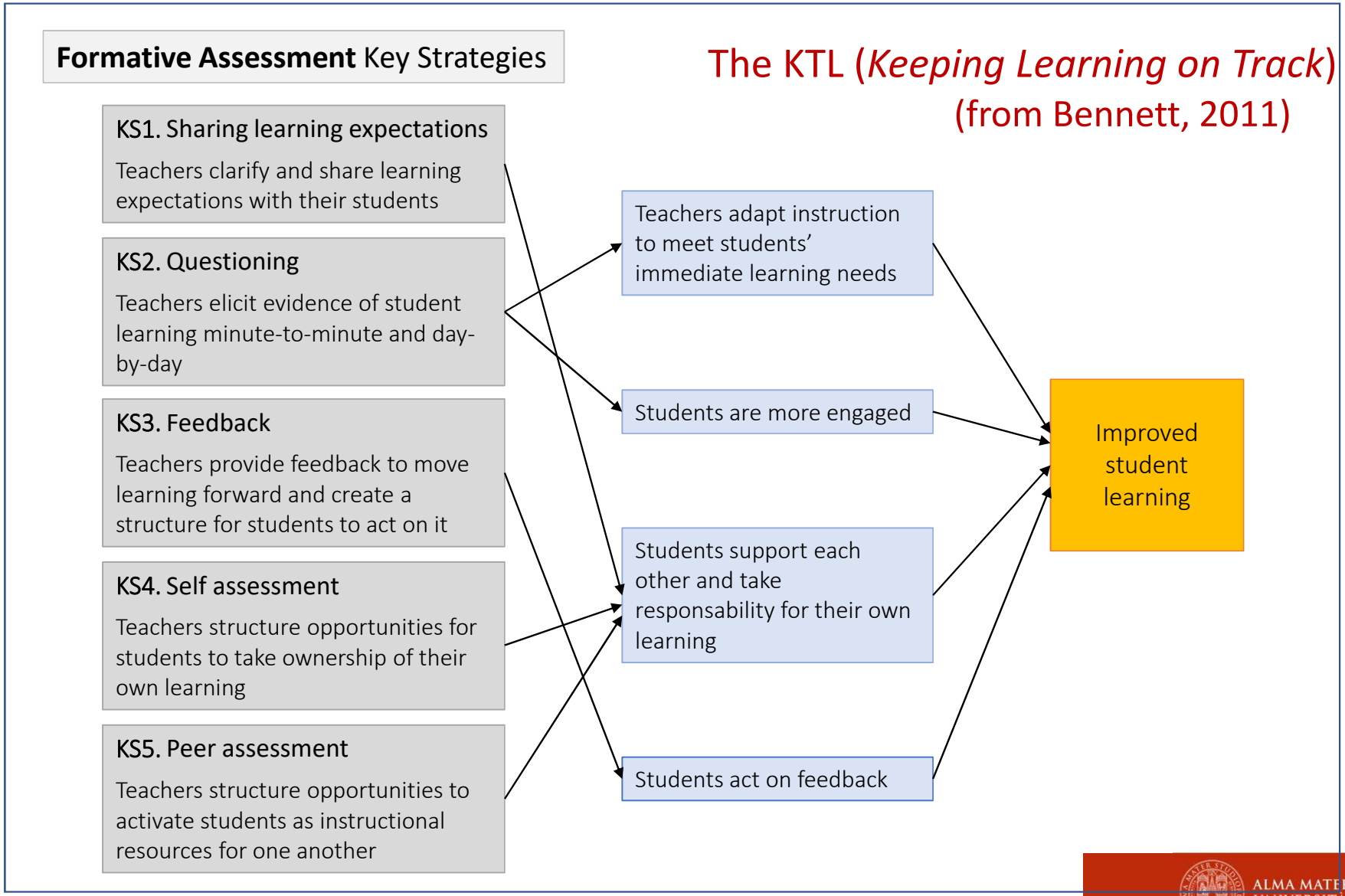


The key elements of AfL

- does not follow the teaching and learning process, but is an integral and essential part of it
- aims to make students increasingly aware, autonomous and responsible with respect to their own learning processes (metacognition and self-regulation)
- requires to clarify and to share with students intended learning goals and criteria for success
- requires providing students with frequent regulatory feedback (helping students understand how to improve)
- is attentive to emotional and motivational dimensions (focus on performance and not on the person, enhancement of potential, vision of error as a resource for learning)
- aims to actively involve students in the assessment processes (promoting self-evaluation and peer evaluation)


Assessment for Learning
Five Key Strategies
Sharing Learning Expectations
Eliciting Evidence
Feedback
Self Assessment
Peer Assessment

The KTL (*Keeping Learning on Track*) framework (from Bennett, 2011)



Effective feedback (Sadler, 1989)

Formative assessment refers to assessment that is specifically intended to **generate feedback on performance to improve learning**

From the student's point of view

Three conditions necessary for students to benefit from feedback. They must know:

1. what good performance is (the student must possess a concept of the goal or standard being aimed for)

2. how current performance relates to good performance (the student must be able to compare actual performance with the goal or standard)

3. how to act to close the gap between current and good performance (the student must be able to take action to close the gap)

From the teacher's point of view

Relate feedback:

- to the goal or standard (assessment criteria)
- to the student's performance
- to the actions to improve performance

Give feedback:

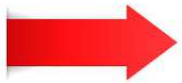
- with adequate times and frequency
- in a language understandable to the student
- with comments in **"positive-negative-positive" sandwich mode**

Feedback must be linked to opportunities for improvement and should encourage the view that mistakes are a part of learning

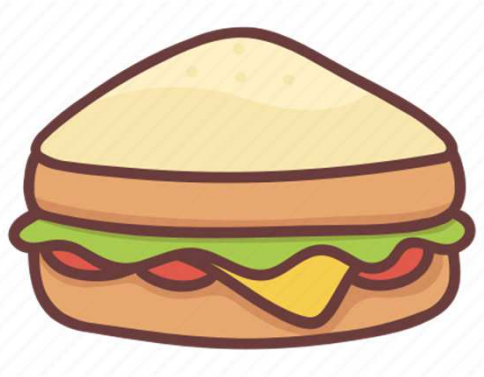
The "sandwich model" of feedback

Give students a positive comment on what did they do well

Give students clear and specific informations about errors and area of improvement



Give students clear and specific suggestions on how to improve (what strategies to do better?)



Formative Assessment acts on the student's metacognitive abilities, that is, on aspects such as:

Attribution styles (locus of control)

Learned helplessness

Motivation and orientation towards learning

Self-efficacy

Tolerance of frustration

Anxiety

….

A central argument is that, in higher education, formative assessment and feedback should be used to ***empower students as self-regulated learners****.*

This observation has led to the conclusion that, as well as improving the quality of feedback messages, teachers should focus much more effort on ***strengthening the skills of self-assessment in their students****.*

Nicol & Macfarlane-Dick (2006)

For example, a key aspect of assessment is assisting students to recognize cues from the context of study which indicate what is good quality work and helping them develop criteria which enable them to distinguish good from not so good task performance.

That is, helping students recognise external clues they can use ***to give feedback to themselves****.*

Boud (2000)

Self and Peer Assessment

❖ Self Assessment

A process in which a student monitors their own learning in relation to specific goals, and then makes necessary changes, using appropriate strategies, to meet the goals. It involves metacognition and self-regulation (an internal conversation led by a series of questions: Where am I now? Where am I going? How can I get there?)

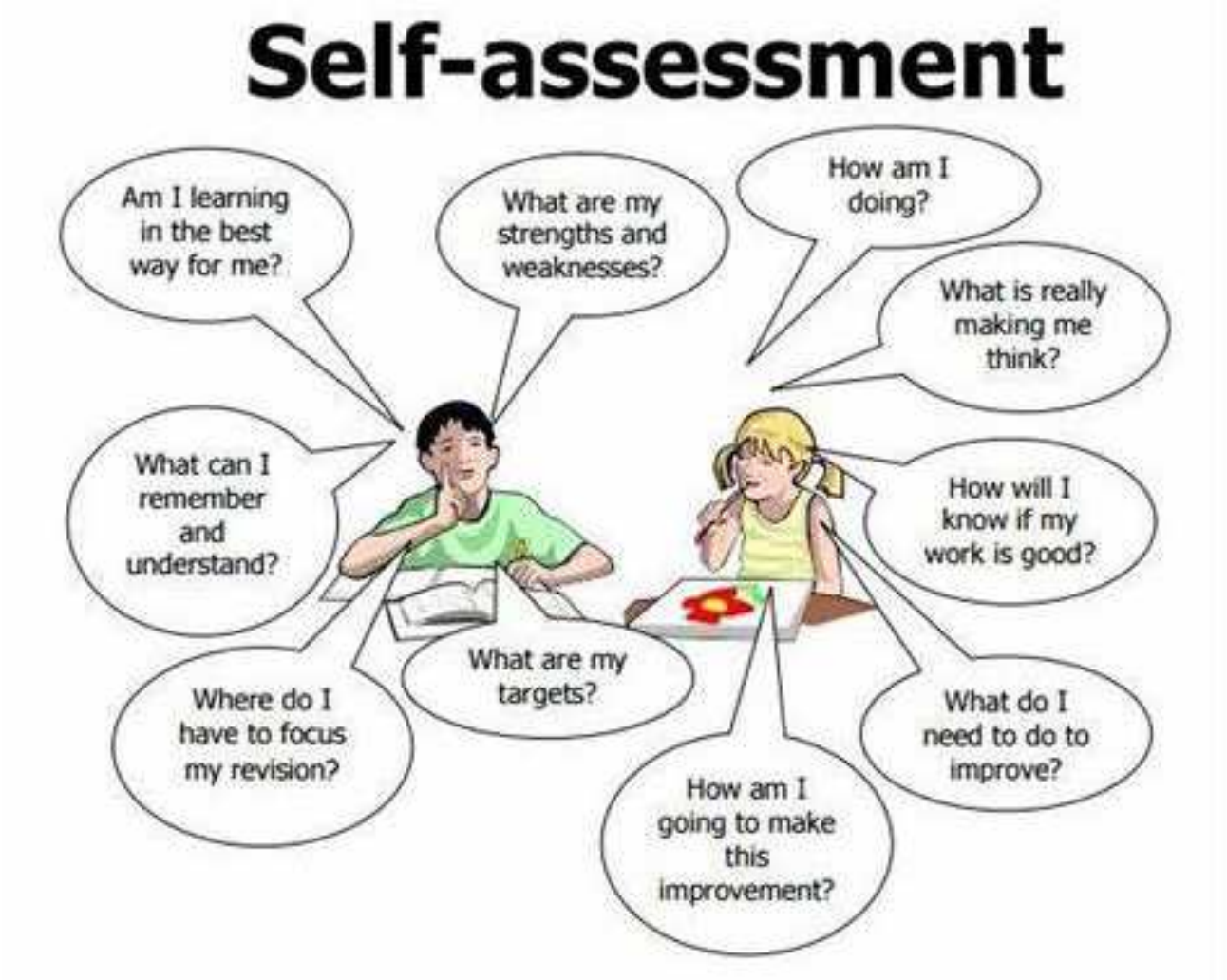
e.g. *Students are involved in and responsible for assessing their own piece of work (they must know criteria for success)*

❖ Peer Assessment

A process in which students provide feedback to each other in relation to specific learning goals

e.g. *Students are involved in assessing the work of other students (they must have a clear understanding of what to look for in their peers' work)*

Peer and self assessment are often considered together. The first can help the second: by judging the work of others, students gain insight into their own performance


Self-assessment
Am I learning in the best way for me?
What are my strengths and weaknesses?
How am I doing?
What is really making me think?
What can I remember and understand?
How will I know if my work is good?
What are my targets?
Where do I have to focus my revision?
What do I need to do to improve?
How am I going to make this improvement?

How to peer-assess

It's important when you are peer-assessing someone's work to provide them with honest and helpful comments. Follow these steps for perfect peer-assessment.

What went well?

Give your partner a positive comment. What did they do well? What did you enjoy reading in their work?

Even better if...

What could your partner do to improve their work? Give them clear and honest feedback about the areas for improvement.

Next step

Tell your partner what they need to do next time or in their next piece of work to make it even better. Give them a short target.

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Department of Education

CHINLONE

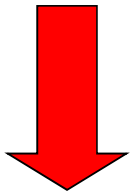
Co-funded by the
Erasmus+ Programme
of the European Union

WP2 Teaching Quality Assurance
Assessment Methods

Authentic Assessment:
definition and construction of authentic tasks

Authentic Assessment

Authentic assessment was proposed in contrast to more traditional forms of assessment centered on the disciplinary contents and decontextualized



Need to ascertain and verify not only what the student knows, but what he can do with what he knows

Wiggins (1993)

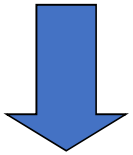
Authentic Assessment

Expressions used to indicate forms of assessment in which the performance of a subject in carrying out **authentic tasks** is examined

Authentic tasks

Complex and open problems *posed to students*
as a means to demonstrate mastery of something

Glatthorn (1999)



They allow the student to demonstrate "what he can do with what he knows" by activating his own resources (knowledge, skills, attitudes ...) to performe **contextualized tasks based on relevant and real-life situations and problems**

Authentic tasks: quality criteria

- The task is **contextualised**: the learning is activated and detected in relation to specific situations and problems
- The task is **realistic**: it simulates or refers to real-life situations and problems
- The task is **open and complex**: it is not merely executive and does not provide a single solution, it asks the student to elaborate and use in a critical and autonomous way what he has learned

Check-list for the construction of an authentic task

Characteristics of an authentic task	
1. reality and authenticity	☐
2. complexity and novelty	☐
3. willingness to act and confront to others	☐
4. focus on problem solving processes and multiple solutions	☐
5. definition of context, place, time (reality principle)	☐
6. definition of recipients and purpose	☐
7. reference to performance or skills	☐
8. use of acquired knowledge and skills (at the end of an educational path) or to be acquired (if you want to create a task for diagnostic/ educational purposes)	☐
9. disciplinary or interdisciplinary (in the second case it is necessary to coordinate the specific objectives with colleagues from other disciplines)	☐
10. individual or group	☐

Authentic tasks for the assessment of competences

Competence as a complex object to be assessed

Some features of the concept of competence that pose challenges for its evaluation

- it is multidimensional
- it is situated and contextualized
- it is not merely applicative/executive
- it has a component not directly visible

What is COMPETENCE?

Guy Le Boterf

"A set of representations, knowledge, skills and behaviors mobilized and combined in a relevant manner in a given context" (1990)

"*Representations, knowledge, skills and behaviors can be summarized with the term "resources", leading us to affirm that competence is a specific quality of the individual: that of knowing how to combine different resources to manage or deal effectively with specific situations in a given context*" (1994)

"*The competence does not lie in the resources (knowledge, skills, ...) to be mobilized, nor is it given by their "sum", but in the very act of mobilization/combination of resources that allows to realize a performance and to achieve a result. In this sense it means knowing how to mobilize resources, combining them in an original and effective way in a given context*" (1997)

What is COMPETENCE?

Michele Pellerey

Subjective or internal resources (knowledge, skills, attitudes and personal dispositions)

*"Ability to activate and coordinate the **internal resources** possessed and the **external ones** available to positively deal with a typology of **challenging tasks or situations**"* (2009)

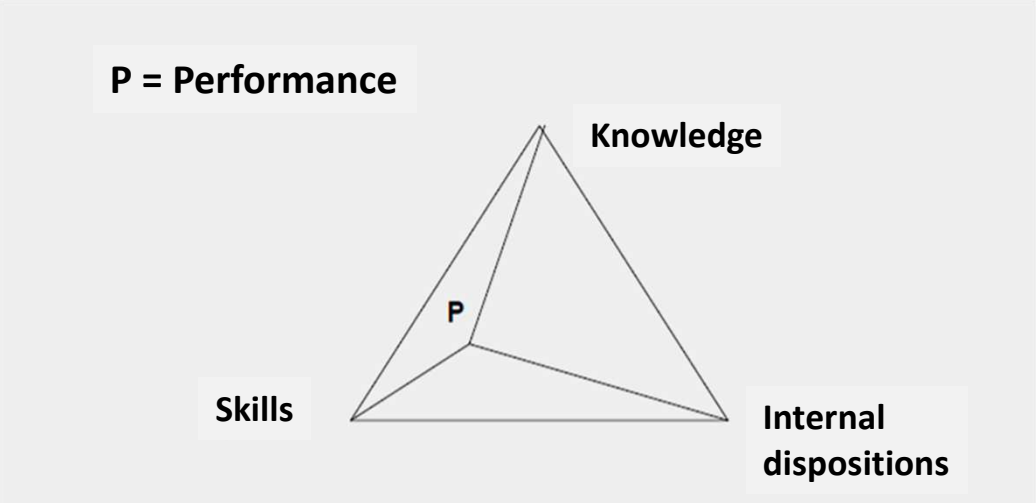
External or contextual resources (material, cultural and social)

Non-replicative and complex tasks that require a certain interpretative effort

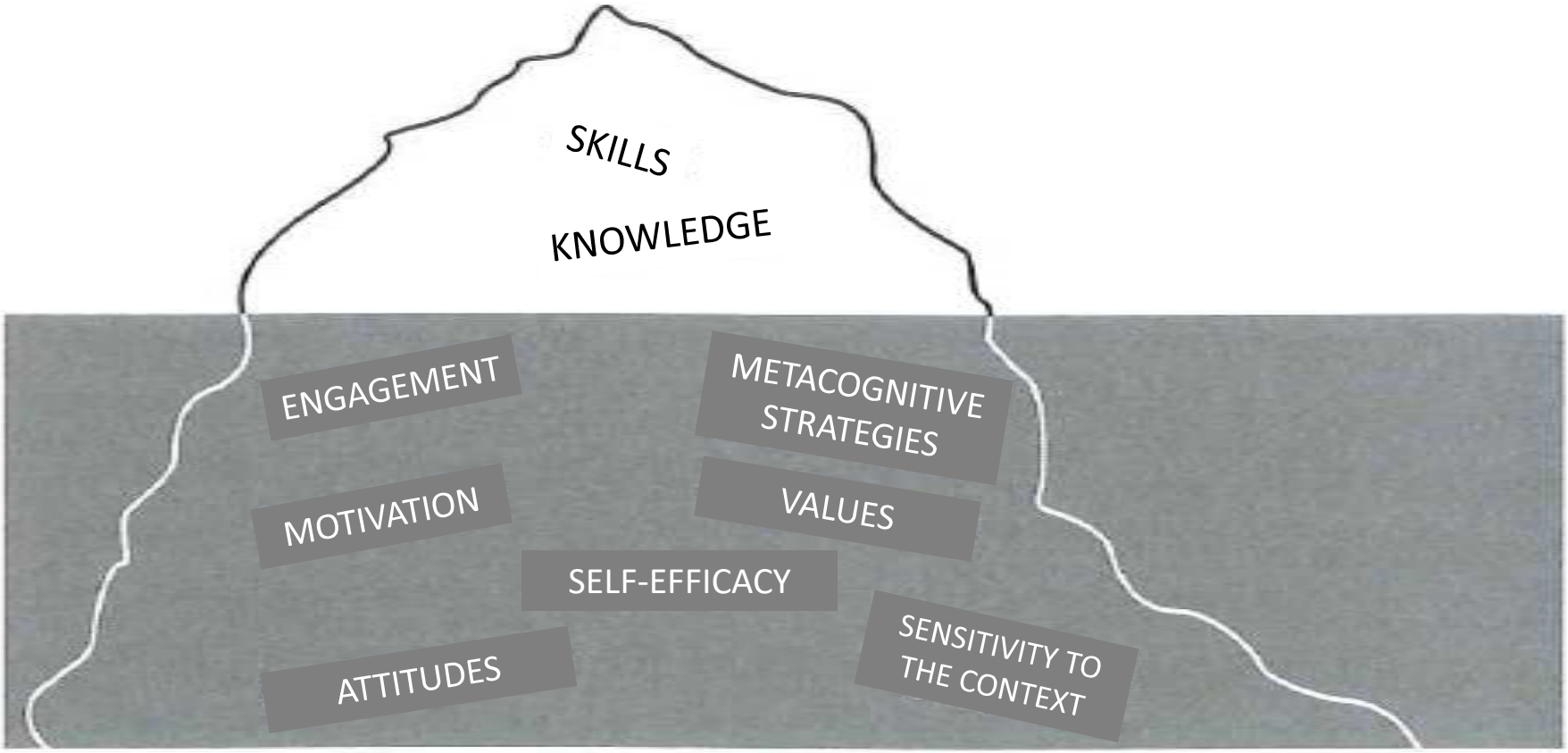
What is COMPETENCE?

The **fundamental components** of each competence:

- **knowledge** (*knowing what*: concepts, terms, facts, principles, theories, methods...)
- **skills** (*knowing how*: cognitive and practical skills)
- **internal dispositions** (attitudes, values, aspirations...)



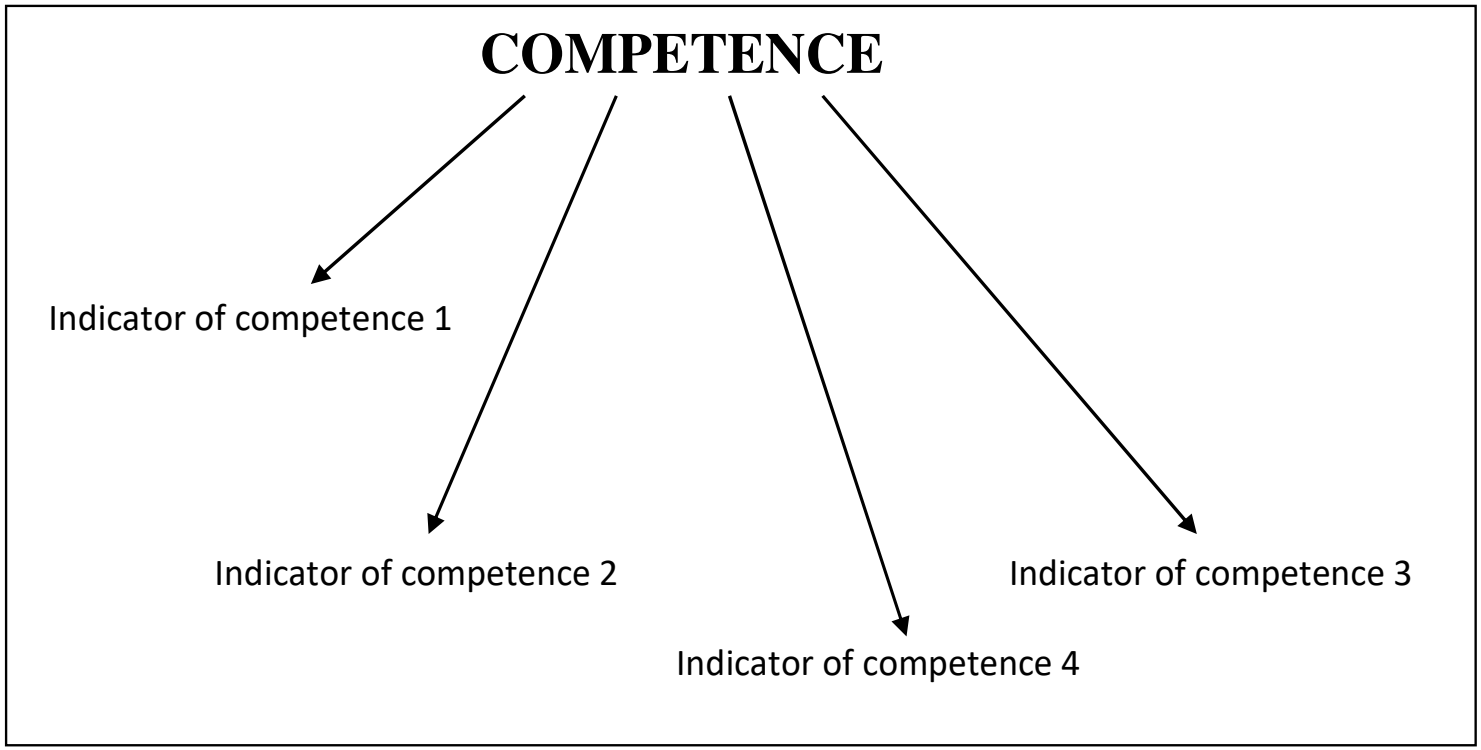
The iceberg of competence



Assessment challenges and perspectives

- The competence itself is not measurable, but it is possible to detect “clues" that constitute indicators of competence
- It is necessary to adopt a multi-perspective view and to use and integrate several sources of information (triangulation)
- Assessment methods must be adopted that are consistent with the construct of competence (e.g. authentic assessment)

Indicators of competence = observable performance of the student



What can be verified and measured directly is not the competence,
but some performances considered valid indicators of that competence

Example of authentic tasks

http://myp8science.blogspot.com/2018/04/authentic-task-climate-change.html

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Department of Education

Thank you!

Andrea Ciani & Alessandra Rosa

alessandra.rosa3@unibo.it
andrea.ciani5@unibo.it