

Making sense of 21st century assessment in education

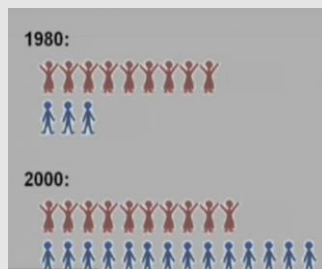
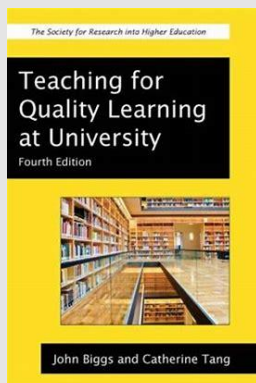
Constructive alignment & validity-related challenges

Jönköping university 2023-01-10
Åsa Hirsh



1

Times change



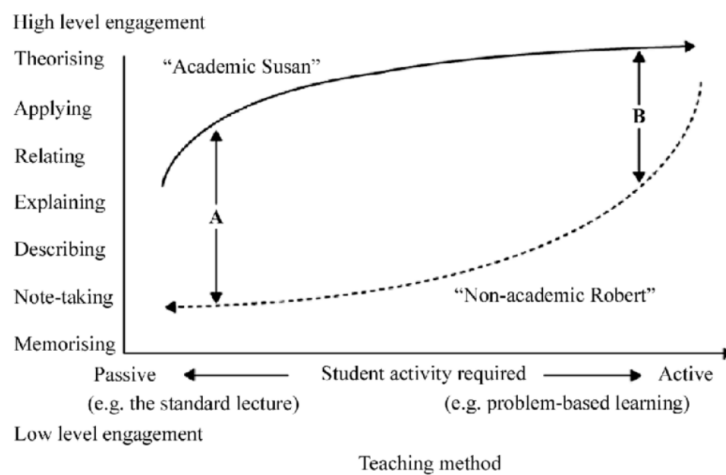
We don't have the same
students in higher
education as we have
traditionally had

- but we have the students
we have!

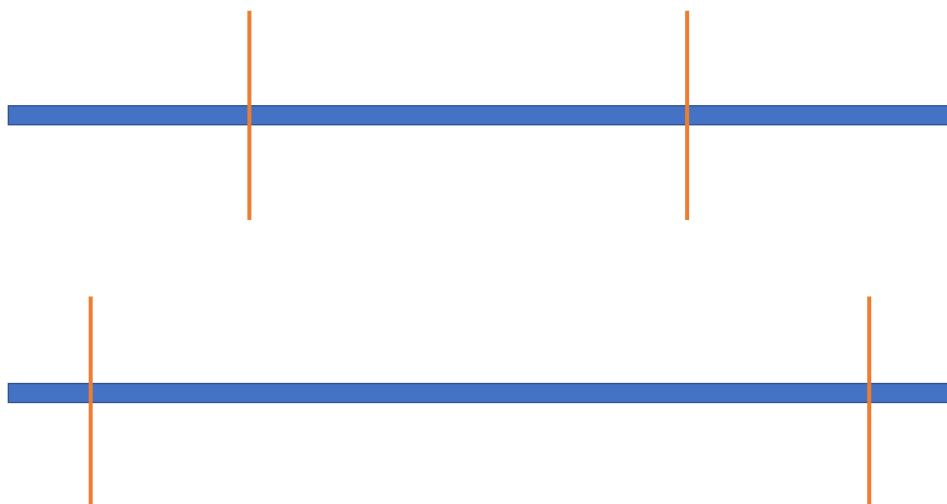
“Academic Susan” & “Non-
academic Robert”

2

Reducing the distance between Susan och Robert

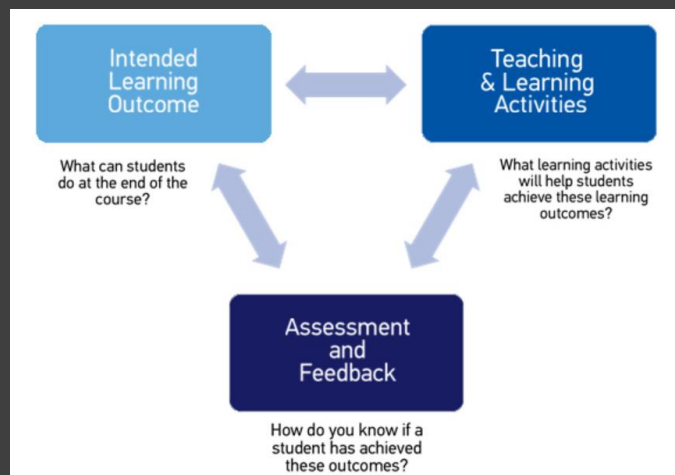


3



4

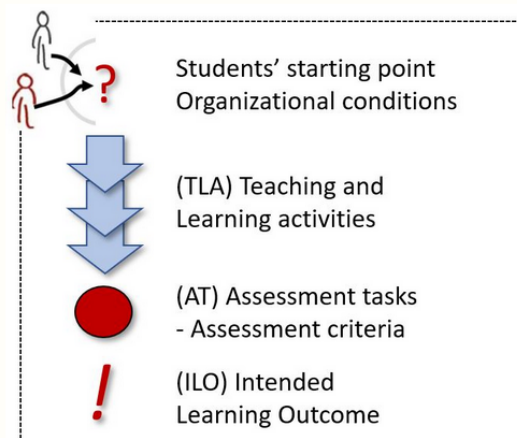
Constructive alignment as a teaching principle in higher education



5

All should be aligned

The logical flow through a course can be described as follows, based on the components of constructive alignment: The students begin the course with a certain level of knowledge, during the course they participate in learning activities (TLA), to finally do assessment tasks (AT) which demonstrate their acquired new knowledge and skills, which prove that they have reached the course objectives (ILO).



Karlstad universitet:

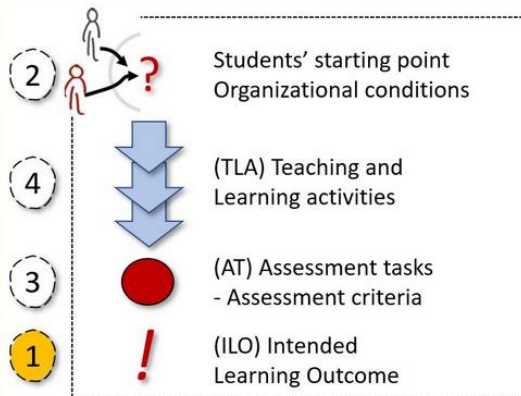
<https://sola.kau.se/keeponteaching/constructive-alignment-as-a-working-method/?lang=en#more-958>

6

1. Start with the course objectives (ILO)

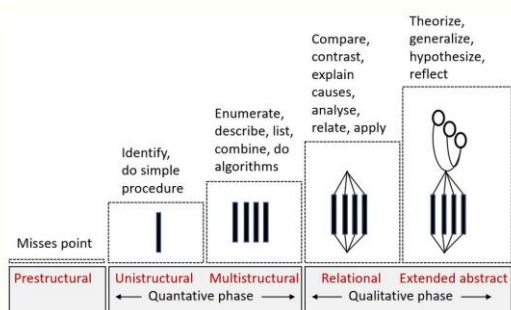
What should students learn and what level of knowledge and skills should they achieve?

We use descriptive verbs such as know, account for, compare, combine. The most common model for describing levels of knowledge and skills is [Bloom's taxonomy](#) in different versions, but there are more such models, e. g. the [SOLO model](#) (Biggs & Tang, 2011).



Karlstad universitet:
<https://sola.kau.se/keeponteaching/construtive-alignment-as-a-working-method/?lang=en#more-958>

7

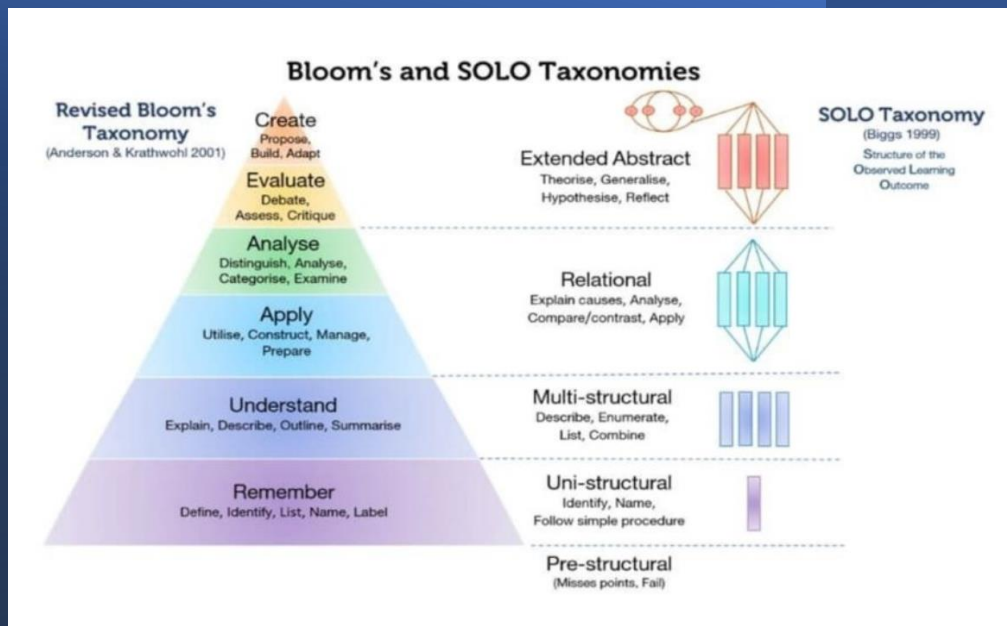


The hierarchy of the SOLO taxonomy in five levels. Adapted from Biggs and Tang (2011, Figure 5.1, p. 91).

Bloom's taxonomy - revised

Creating	Can the learner create new or original work?	assemble construct create	compose design develop	formulate generate produce	revise rewrite write
Evaluating	Can the learner justify a stand or position? Make informed judgments?	appraise argue assess	defend evaluate judge	rank rate recommend	select support value
Analysing	Can the learner draw connections among ideas?	breakdown categorise compare	contrast criticise differentiate	discriminate distinguish examine	predict question simplify
Applying	Can the learner use information in new ways?	choose compute demonstrate	dramatise employ illustrate	interpret operate sketch	solve use write
Understanding	Can the learner explain ideas or concepts and construct meaning?	classify compare describe	discuss distinguish explain	identify locate recognise	summarise translate paraphrase
Remembering	Can the learner find or recall information?	define draw duplicate	list label match	name outline recall	repeat select state

8



9

2. The external preconditions?

Who are the students?

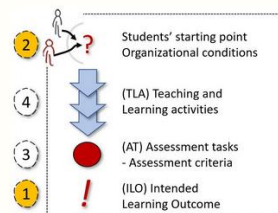
Previous knowledge, study experience, their short- and long-term goals with the course?

Organizational conditions?

How large is the student group? Campus/distance? Pace of study? The course's purpose in the study program? Any ongoing parallel courses? Available rooms, equipment, etc.?

The course design must always be adapted to the current student group and accommodated within the organizational framework.

The first two steps thus frame the course: Students' starting point and external conditions in one end, and end goals in the other end.

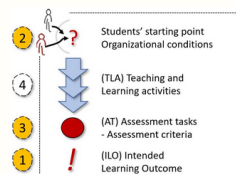


Karlstad universitet:
<https://sola.kau.se/keepon-teaching/constructive-alignment-as-a-working-method/?lang=en#more-958>

10

3. How should we measure if the student has reached the goal? (AT)

The third step is about concretizing the final goal: How should we and the students know that we have succeeded in reaching the Intended Learning Outcome?



Assessment task

Behind the ILO's is of course a notion of a context in which the intended knowledge should be used, that is to say what makes the ILO's meaningful. Design a task (or several) that demonstrates that the student can use the knowledge in this way. In order to be relevant as a goal for the students, the task must be described to them in a way that enables them to self-monitor their progress towards being able to succeed with it.

Assessment criteria

Also describe the criteria that will be used to determine whether the student has reached the objectives or not and the criteria for graded reports.

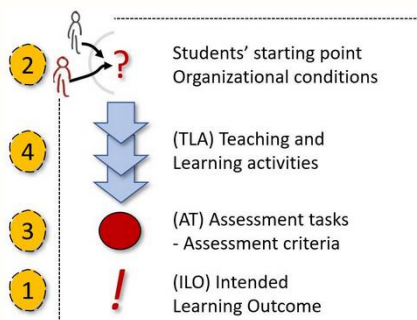
In Biggs and Tang's book (Chapters 10-12) there are many excellent examples of how to design Assessment Tasks and assessment criteria for different types of course objectives.

Karlstad universitet:
<https://sola.kau.se/keeponteaching/constructive-alignment-as-a-working-method/?lang=en#more-958>

11

4. Design the course's learning activities (TLA)

Now it's finally time to draft the course's learning activities. As the basis, we have the description of course objectives, how these are to be measured, which student group it is and what organizational framework we need to work within. Now it's time to design the teaching and learning: the work that students need to do to achieve the course objectives.



Note that the important thing is *what the students need to do*, not what the teachers should do. The latter comes as a result of the former. If we ask ourselves the question in the opposite way ("*How should we, the teachers, conduct the teaching?*"), we easily lose sight of the fact that it is the students' activity that creates their learning.

Karlstad universitet:
<https://sola.kau.se/keeponteaching/constructive-alignment-as-a-working-method/?lang=en#more-958>

12

Dialogue

- What are your experiences of planning courses based on constructive alignment?
- What opportunities and difficulties have you experienced?
- Give examples of how you have managed to increase student activity, i.e., go from "standard lecture" (transmission) to more student involvement and student activity!
- Is the latter also possible in really large student groups? How?



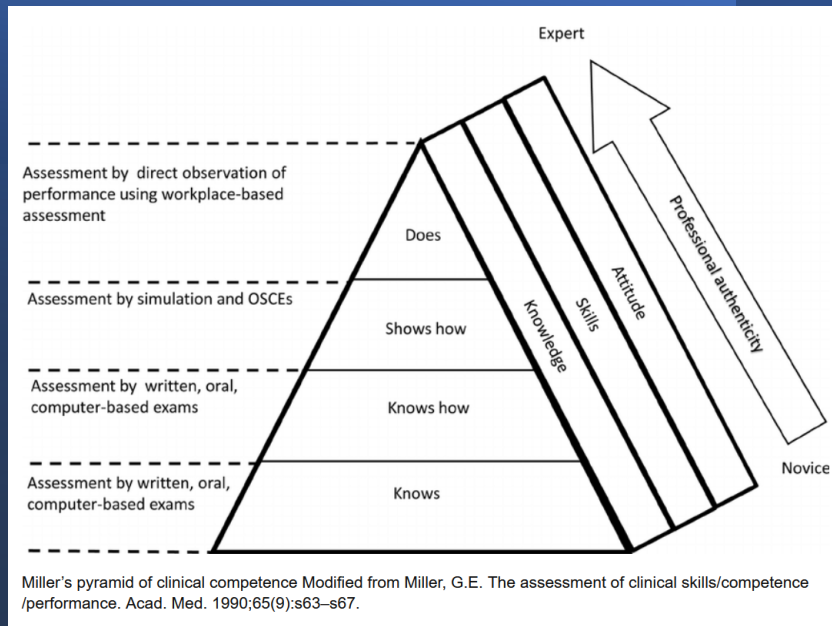
13

Assessing professional knowledge

Give examples of how you assess, i.e., organize assessment situations and tasks, to grasp the different levels in the pyramid!



14



15

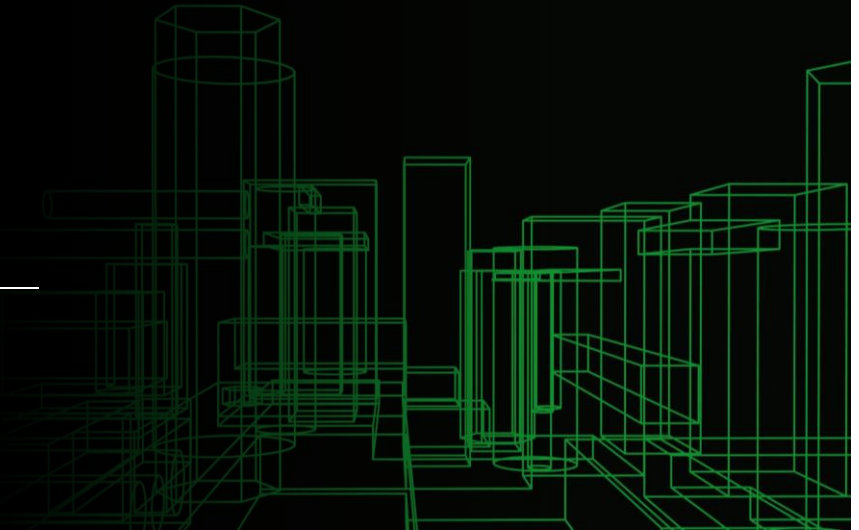
OSCE

- OSCE (objective structured clinical examination) is a practical examination that typically is applied to various educations in health care (e.g., doctor/nurse/dental program) around the world.
- In short, it involves assessing clinical and practical skills as well as communication skills.
- Example from medical education in Sweden: The OSCE is made up of different stations that the student must go through in a limited time. For example, there could be 6 stations with a 6-minute time limit at each station. At each station, the student must solve a problem on her/his own and from a doctor's perspective. There is an examiner (doctor) inside each station. One station may mean that you perform a heart status (= examine the heart and describe what you see/hear/feel...) and another station may mean interpreting an X-ray image and coming up with diagnostic suggestions.
- Objective..?

16

Making sense of 21st Century Research on Assessment in Education –

An overview of research
topographies and contemporary
research interests



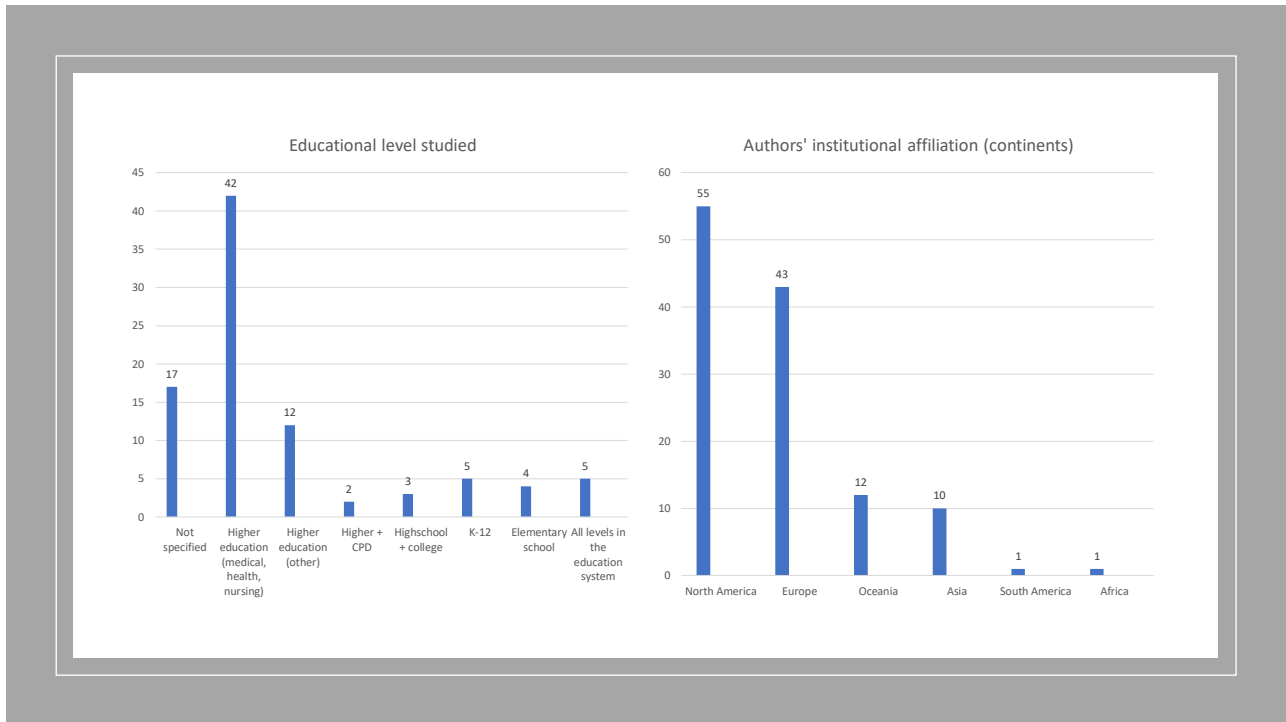
17

Aim and research questions

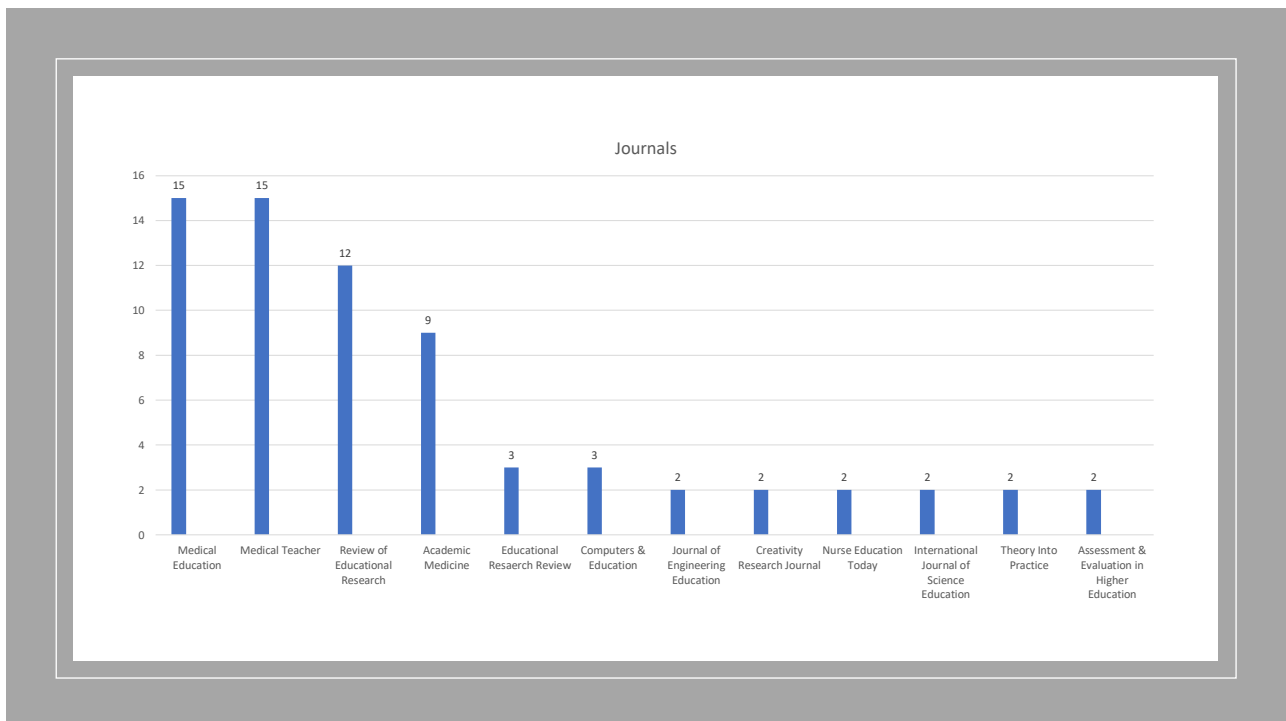
- Aim: To provide a state-of-the-art description of a research area whose scope has been significantly broadened during the 2000s, and to identify current and future challenges.
- How? By mapping and analysing research topographies and interests in high-impact research. 90 top-rated (WoS) 21st century research reviews, where assessment of/for students in any part of the formal education system is a central aspect, were included.
- RQ1: What characterizes the included reviews regarding:
 - a) geographical distribution of authors' institutional affiliations,
 - b) journals in which the reviews are published, and
 - c) educational level studied?
- RQ2: Which issues of inquiry are addressed how can these be understood in relation to contemporary development in society and the formal education system?



18



19



20

Research interests in the 21st century



Validity issues and validity discussions run like a common thread through virtually all the underlying material. In 66 reviews, i.e., 73% of the underlying material, various validity issues form the very core of the questions asked and discussions undertaken.



Validity is not only problematized when formal standardized tests are studied and discussed. Rather, the validity concept is used in a broader sense to problematize the degree to which assessment information obtained from various types of formal or informal assessment situations, for summative as well as formative purposes, provides information about the construct and content that one needs information about (in accordance with, for instance, learning objectives).



While various validity issues basically seem to be eternal (e.g., the consideration of construct, criterion, content, and predictive validity), it is possible to discern time-typical trends in the material where validity questions are raised, and perhaps become even more critical, in new contexts in line with developments in society and the formal education system.

21

Six main areas were distinguished which can be said to constitute 21st century trends that expose validity-related tensions and dilemmas.

Multiple purposes and uses of assessment information

The assessor role

Growing student groups

Technologies for educational assessment

Complex skills

Increasing diversity

22

Multiple purposes and uses of assessment information

- The increased interest in multiple uses of assessment information is reflected in the underlying reviews
- 16 reviews focus solely on formative purposes and uses of assessment information, 32 focus on both summative and formative purposes and uses, and 43 delve purely into summative purposes and uses.
- Validity issues are brought to the fore regardless of whether summative or formative assessment purposes and uses are discussed. They become more critical when the purpose of the assessment is summative and of a high-stake nature than when the purpose is formative.
- However, the educational decisions made based on assessments carried out for formative purposes (whether such decisions concern how instruction can be (re)organized or how feedback to the student should be designed) are directed at students' learning relative learning objectives that will ultimately be assessed summatively. Therefore, it is not possible to disregard the validity issues when it comes to formative purposes

23

The assessor role

- Issues relating to the assessor role attracts interest among researchers in the 21st century.
- Partly a consequence of a development where key strategies for formative assessment, involving peer and self-assessment, have increasingly come to impact classroom practices throughout the formal education system.
- In addition, technological development has led to the use of non-human actors, i.e., computers, as assessors.
- Another aspect of the assessor role that is frequently discussed in the underlying reviews is the use of laymen in assessment contexts. Although these often lack in-depth assessment knowledge, they are assumed to provide important input and perspectives in authentic assessment.
- Thus, the assessor is no longer just the educator, but can instead be a peer, oneself, a workplace supervisor, a layman, or a computer.
- This naturally raises questions about how well sent to make valid assessments and provide valid feedback that different assessors are.

24

Growing student groups

- The rapid social and economic societal transformations of recent decades have led to an increasing number of students remaining in the formal education system for longer.
- The size of student groups has increased, as have the demands placed on students' knowledge and skills. Resources have, however, not increased proportionally.
- Valid assessment of students' knowledge and goal attainment is resource intensive. Several underlying reviews address the question of how assessment of large groups of students can be made more efficient, preferably more or less automated.
- Technical solutions are developed with the goal of reducing the need for human involvement, such as self-correcting multiple-choice tests or automated essay assessment.
- While such solutions solve certain problems, others arise. There is an imminent risk that the easily measurable ends up at the center at the expense of valid assessment of the more complex knowledge and skills that learning objectives state.

25

Technologies for educational assessment

- 20% of the underlying reviews address assessment issues that are technology related.
- The reasons for the increased presence of digital tools can be divided into three main motivators:
 - ✓ the use of technology is a response to policy requirements,
 - ✓ the use is a response to resource issues, and
 - ✓ the use is motivated in terms of educational benefits
- Using digital solutions for resource-saving purposes is the most common motivator, which relates to the dissonance between growing student groups and the efficiency requirements. Assessing and grading student work in large groups is highly labor intensive, wherefore institutions and instructors turn to automated grading software.

26



Dialogue

- What's it like at your respective institutions?
- When are digital tools for assessment purposes used as a response to resource issues and when is the use motivated in terms of educational benefits?
- Discuss concrete examples you have encountered of both types of situations!

27

Validity issues (technology for educational assessment)

- Frequently discussed: To what extent can computer aided assessment be used when it comes to determining more complex abilities in students?
- Health care has grown more complex and demands for patient safety have increased, while resources for instruction and assessment have become more limited. Technology has come to be regarded as a solution, but most studies offer limited validity evidence.
- For instance: Although intelligent parsing of free text is gradually improving, it is still a long way from matching human scrutiny and interpretations. Automated essay-grading cannot identify truth and set a limited standard for what good writing is. Automated essay scoring cannot score, for instance, the use of complex metaphors or humor.
- A central validity question concerns which information about students' knowledge that becomes available versus which does not when automated testing and assessment is used. There is imminent risk that knowledge and competencies that are to be developed in accordance with learning objectives are not validly assessed, as the assessment instruments are not optimal for providing that information.

28

Complex skills

- In one third of the included reviews, the question of how complex skills can be validly assessed lie at the very center.
- Apart from one review that concerns assessment of 21st century skills in engineering education and two reviews that address the question of how creativity and creative potential can be measured, the rest of all reviews that concern valid assessment of complex skills are from medical education and health care.
- Examples of complex skills frequently studied and discussed are professionals' (mainly doctors and nurses) interaction and communication with patients, professionalism in general, ethical and affective competence, independent decision-making ability, and clinical competence.
- The valid assessment of complex abilities is clearly linked to what appears in the material as an ever-present dilemma, namely that between valid assessment and cost-effectiveness.
- A certain triangulation (i.e., the use of various complementary assessment forms and the involvement of different assessors' perspectives), or multiple biopsies of the same content domain, is required to achieve validity when assessing complex knowledge and skills.
- Triangulation and multiple biopsies are, however, resource intensive. It is argued, for example, that it takes 30-50 patients who assess one medical student's physician's competence for the assessment to be considered reliable and valid. Difficult to defend from a cost perspective.
- In addition, when real patients are used as assessors alongside instructors, their ability to understand and relate objectively to established assessment criteria needs to be considered.

29

Increasing diversity

- Another consequence of the fact that a growing number of students remain in the formal education system for longer is increasing diversity in student groups. The fact that there is no standard student and that the inclusion perspective needs to be considered in the assessment context is a prominent theme in a number of reviews.
- Some reviews address test accommodations for students with learning disabilities (LD). The validity issues that come to the fore concern the possibility of ensuring that the adjustments made to the assessment situation will benefit the students with disabilities, while the same adjustments would not also benefit students without disabilities.
- Some also question whether it is reasonable to assume that 'one size fits' all when it comes to assessment practices that are a consequence of the formative paradigm.
- The effects of self- and peer assessment for students with LD are studied. Since they often struggle with organizing abstract information, they tend to be overwhelmed by the abstract process of setting reasonable goals for themselves and evaluating whether the goals have been achieved or not.
- Some claim that peer-assisted learning and peer feedback do not help students with LD, since students with LD are often so far below the average performance of classmates that peer feedback is insufficient to help them understand difficult concepts.

30

-
- Societal development in terms of increased global mobility and ethnic diversity among students, is also reflected in the underlying reviews.
 - Some address the impact of language barriers in assessment situations, and whether poorer test results can be translated into a lack of knowledge about the construct under measure, or whether it can be attributed to a purely linguistic inability to understand the content of tasks or to express knowledge.
 - In one review, examining the assessment of English language learners (ELL:s) in science, a concern is raised that their science achievement is underestimated when they are not allowed to demonstrate their knowledge in their home language, which in turn raises questions about educational equity.
 - Arranging assessment situations that are adapted to students with special needs, or situations where it is possible for students to demonstrate subject-specific knowledge in such a way that the language itself does not constitute an obstacle, is often resource-intensive. Accommodations are unlikely to be implemented if they are not logistically manageable or affordable.
 - The question of greater diversity in student groups is also reflected in the discussions and conclusions of many reviews, concerning which further and future research is crucial for advancing assessment research. Research is requested that to a greater extent considers how factors such as gender, ethnicity, culture, etc. affect and interact with forms of assessment and assessment situations.

31

Summing up

- Several reviews address the question of how that which has so far been considered immeasurable can be made measurable and assessable.
- This is largely a result of a growing insight into the fact that professional knowledge, not least in interpersonal professions, is complex. While knowledge of hard facts is still essential, students need to be provided with knowledge and skills beyond that. Phenomena such as 21st century skills, general skills, clinical skills, ethical skills, etc. are increasingly incorporated into education and learning objectives.
- While this can be considered highly commendable, it nevertheless poses challenges to the achievement of valid assessment practices. A major dilemma is that while knowledge of valid assessment of complex competencies has increased, the (often wise) solutions proposed in research are usually so resource-intensive that they are difficult to apply in a resource-limited practice.
- Since valid assessment of complex abilities is costly and the total monetary resource always limited, the gains at one end risk being made at the expense of losses at the other. Thus, valid assessment may come to take place at the expense of e.g., high-quality instruction.

32

Summing up

- As major advancements in technology have taken place, hopes have increased that automated computerized assessments will solve this dilemma. Despite remarkable developments, however, we have not yet reached the point where a computer can replace human judgment and scrutiny.
- Until we do (if we ever will) we should know that the measuring of the (to some extent) immeasurable is, in a sense, a pseudo-quantification.
- Liedman describes a pseudo-quantity as a quality that would more accurately be characterized verbally. It is a spurious quantity that claims to indicate the dimensions of something, but can in fact, if taken at face value, be misleading. Pseudo-quantities, Liedman argues, “proliferate in an age where all types of quality assessments are expected to be expressed in quantitative terms and in which impartially and illusory exactitude appear to be more reliable than qualified expert judgment”

33

Summing up

- The cost-effectiveness versus validity dilemma also arises in relation to the increased diversity in student groups. Over the years, research has contributed greater awareness and knowledge of how assessment situations for students with LD, other special needs, or where language barriers exist, can or should be handled validly.
- From an equity point of view, these students should indisputably be offered the compensation required to have equal opportunities to demonstrate their knowledge. This too, however, is often highly resource-intensive, and may inevitably mean reduced resources for high-quality instruction for all students.
- The dilemma of cost-effectiveness versus validity is a dilemma in the true sense. There are no given answers. When two seemingly incompatible goals must be achieved and neither can be opted out, a reasonable balance must be sought. This balancing act is anything but simple, not least because it largely involves ethical considerations. The question of how a reasonable balance between cost-effectiveness and validity can be achieved is perhaps impossible to answer, but nonetheless one of the most important to problematize both in education and in research.

34



The dilemma of cost-effectiveness versus validity

- Is a dilemma in the true sense. There are no given answers.
- When two seemingly incompatible goals must be achieved and neither can be opted out, a reasonable balance must be sought. This balancing act is anything but simple, not least because it largely involves ethical considerations.
- The question of how a reasonable balance between cost-effectiveness and validity can be achieved is perhaps impossible to answer, but nonetheless one of the most important to problematize both in education and in research.

35

Oral examinations



36

A matter of validity – fitness for purpose

- The form of assessment we choose needs to give us the information we need in relation to the ILO - about each individual
- We have three overall forms of assessment to choose from - written, oral and observational
- If we are to assess someone's writing ability, we must use the written form.
- If we are to assess someone's oral abilities, we must use the oral form.
- If we are to assess someone's ability to perform something in practice, we need to use the observational form.
- But often we have the opportunity to choose - not least between oral and written form

37

To think of when it comes to group presentations as assessment form

- Oral presentations in groups have many advantages, but they need to be constructed so that the individual student's knowledge is made visible within the frames of what the group jointly expresses at the presentation.
- It cannot be assumed that a student who does not receive or take so much space during an oral presentation in a group lacks sufficient knowledge.
- It may indeed be so, but it may just as well be because one or a few other students in the group have been dominant and decided who should say what at the time of presentation.
- It may also be the case that a student in a group has carried a very large workload during the group work and that other group members step in and report based on a script that student has written.
- Personally, I would probably not use group presentation as the only form of examination if I needed to assess the students based on a multi-grade grading scale. *Maybe* if it's only about Pass or Fail.

38

Before and during an oral examination

- Prepare carefully what aspects you will look for when you listen
- Share (and explain) these aspects with students in advance
- Keep the aspects in front of you as you listen, preferably in a matrix-like document where you can tick preliminary levels and write some short notes as you listen
- Add any comments and make a preliminary overall assessment as close to implementation as possible

	Utmärkt	God	Att utveckla vidare
Skolutvecklings-modell och dess utgångspunkter	Beskrivning och analys av skolutvecklingsmodellen är väl kopplad till dess utgångspunkter och knyter an till litteraturen med relevanta begrepp	Beskrivning och analys av skolutvecklingsmodellen är kopplad till dess utgångspunkter och knyter an till litteraturen	Kopplingar mellan skolutvecklingsmodell och dess utgångspunkter saknas eller är ej relevanta, bristfällig hänvisning till litteratur
Skolutvecklings-modell och egen erfarenhet	Utifrån väl valda egna erfarna eller tänkta exempel och resonemang granskas och bedöms skolutvecklingsmodellen	Utifrån egna erfarna eller tänkta exempel och resonemang granskas och bedöms skolutvecklingsmodellen	Granskning och bedömning av skolutvecklingsmodellen utifrån egen erfarenhet saknas eller kopplingen mellan egen erfarenhet och modellen är inte relevant
Referenser/referenslista	Korrekt skrivna och korrekt återgivna Citat	Behov av mindre korrigeringar	Ikke korrekt skrivna
Språklig utformning	Språkligt god	Mestadels god	Svårbegriplig

Kommentarer:

Bra och kärnfullt sammanfattat i textens inledning. Du visar sedan genomgående en fördjupad teoretisk förståelse på ett brett plan. Det blir ibland lite oklart, i vissa delar av texten, vad som handlar om pedagogisk dokumentation specifikt, samtidigt som det blir kraftfullt när man gör de jämförelser utifrån så många olika källor som du gör. Det hade varit illustrativt med fler egna exempel från den egna praktiken. Referenshantering och språk mycket bra.

39

IRL versus pre-recorded oral presentation

A big advantage of oral presentations IRL is the opportunity to ask clarifying and deepening questions

Digital solutions offer great (and often fun) opportunities and can be "kinder" to students who get nervous about oral exams

40

Dialogue

- Which form(s) of assessment dominate where you are?
- What do you think are the reasons for choosing one form or the other?
- Do you feel more confident using one form or the other? Why?
- Share some experiences of using oral forms of assessment!
- What opportunities and difficulties have you experienced when it comes to using oral examinations?

