



香港大學

THE UNIVERSITY OF HONG KONG

Evidence-based assessment in the era of big data

Dr. Maggie Zhao

Teaching and Learning Evaluation and Measurement Unit
The University of Hong Kong

Presentation at the Higher Education for Tomorrow Summer Institute, June 20, 2019.

The Fundamental Question

WHAT IS ASSESSMENT?



SERIES: Assessment of 21st Century Skills



EDUCATION PLUS DEVELOPMENT

Using assessment data to help students thrive: Classroom perspectives from Mongolia

Khandjav Terbish, Munkhbayar Baljinnyam, and Esther Care
Thursday, December 20, 2018

EDUCATION PLUS DEVELOPMENT



Editor's Note: This is the third post in a series about *using data from 21st century skills assessments*.

We are teaching students who will eventually work, use technologies, and solve problems yet unknown. How do we identify the skills that will help prepare them for the challenges of tomorrow's complex communities and workplaces?

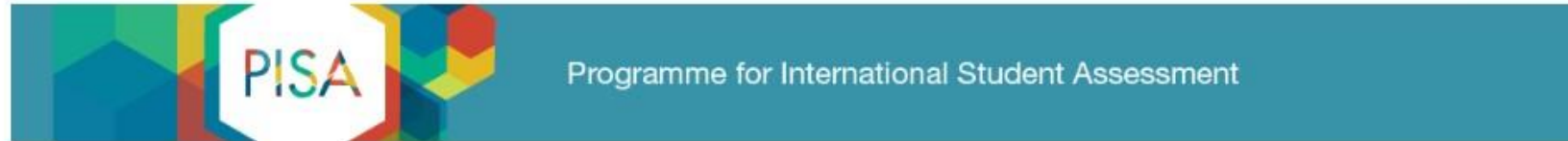
Traditional methods of assessing students' core skills have dominated education for past decades in Mongolia. That is, assessment has been used mainly for **monitoring and accountability**. Mongolian schools have primarily assessed the abilities that students acquire at the end of an academic semester as a summative type of assessment. However, as discussed in a recent Brookings **report**, we have realized that this approach is insufficient.

South Korea grounds all planes so students can focus on college exam

Airport take-offs and landings suspended so students can hear language test, while roads are cleared for late-running teenagers



▲ Students sit the annual college scholastic ability test at a high school in Seoul, South Korea, on Thursday.
Photograph: Jung Yeon-Je/AFP/Getty Images



[Home](#)

[About](#)

[PISA Test](#)

[Data](#)

[Publications](#)

[Webinars](#)

[Join PISA](#)

[FAQ](#)

About

What is PISA?

The Programme for International Student Assessment (PISA) is a triennial international survey which aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students who are nearing the end of their compulsory education. PISA assesses how well they can apply what they learn in school to real-life situations. Over 90 countries have participated in the assessment so far which began in 2000. Every three years students are tested in the key subjects: reading, mathematics and science. In one assessment there is a focus on one of the subjects. For example in the year 2000, the focus was on reading which allowed us to get more in-depth information on the students' reading skills. In addition to the three core subjects, students are tested in an innovative domain such as collaborative problem solving in 2015 and global competence in 2018. Some countries also choose to administer an assessment in financial literacy.

Not only do the students take a test, but also, they fill out a background questionnaire to provide some contextual information e.g. they are asked about the level of education of their parents and how many possessions there are in their household. In addition school principals fill out a questionnaire about how their schools are managed. This gives us insights into how the students' background and school environment could influence their performance. Some countries also choose to administer extra background questionnaires to the students' to find out about

PISA Country Participation

See the list of countries and economies that have participated in PISA to date.

Map of PISA Countries and Economies





FREE School Year Success Kit

DOWNLOAD NOW

AdChoices 

AdChoices 

EDUCATION 12/14/2011 05:49 pm ET | Updated Dec 15, 2011

Test Scores Often Misused In Policy Decisions

By Joy Resmovits

Education policies that affect millions of students have long been tied to test scores, but a new paper suggests those scores are regularly misinterpreted.

According to the new [research](#) out of Mathematica, a statistical research group, the comparisons sometimes used to judge school performance are more indicative of demographic change than actual learning.

For example: Last week's [release](#) of National Assessment of Educational Progress scores led to much finger-pointing about what's [working](#) and what [isn't](#) in education reform. But according to Mathematica, policy assessments based on raw test data is extremely misleading — especially because year-to-year comparisons measure different groups of students.

“Every time the NAEP results come out, you see a whole slew of headlines that make you slap your forehead,” said Steven Glazerman, an author of the paper and a senior fellow at Mathematica. “You draw all the wrong conclusions over whether some school or district was effective or ineffective based on comparisons that can’t be indicators of those changes.”

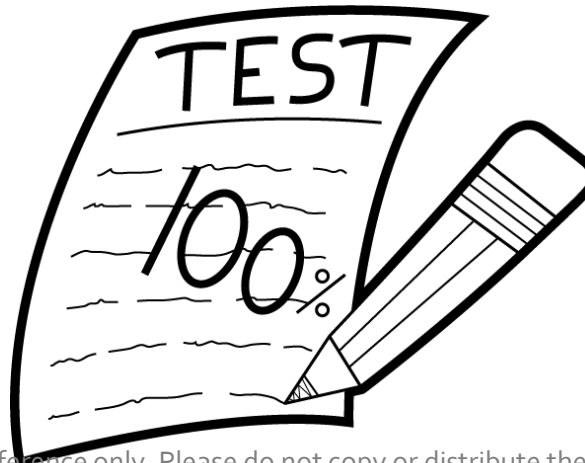
The Fundamental Question

WHAT IS ASSESSMENT?



Summative
Assessment

Assessment of learning



Formative
Assessment

Assessment for learning

Outline

- Brief introduction
- Fundamentals and myths of assessment
- Using assessment for quality assurance and enhancement: Emerging trends globally and HKU's practice
- How can next generation assessment best support teaching and learning: Reflections and recommendations
- Reimagine next generation assessment for teaching and learning: Discussions and insights from participants

<http://tlem.hku.hk>



t&lemu

Teaching and Learning Evaluation and Measurement Unit



About Us Our Team SLEQ Survey Portfolio

Network on Education Quality Monitoring in the Asia-Pacific (NEQMAP), UNESCO Bangkok

Reporting directly to the Vice-President and Pro-Vice-Chancellor (Teaching and Learning), the Teaching and Learning Evaluation and Measurement Unit (T&LEMU) is a central unit that supports the University's teaching and learning strategies, quality assurance and enhancement through collecting stakeholders' feedback and conducting evaluation and measurement studies.

The Teaching and Learning Evaluation and Measurement Unit is an institutional member of the Network on Education Quality Monitoring in the Asia-Pacific (NEQMAP), facilitated by UNESCO Bangkok.



Asia

- Afghanistan
- Bangladesh
- Bhutan
- **Cambodia**
- Hong Kong
- India
- **Indonesia**
- Japan
- Lao PDR
- **Malaysia**
- Maldives
- **Mongolia**
- **Myanmar**
- Nepal
- Pakistan
- Philippines
- Russian Federation
- Singapore
- South Korea
- Sri Lanka
- Thailand
- Vietnam

Australia

- Australia
- Fiji



INSTITUTIONAL SURVEYS

To collect, analyse and report institutional data through institutional surveys for informing and supporting the University's quality assurance and enhancement in relation to teaching and learning.



EVALUATION & MEASUREMENT

To conduct evaluation and measurement studies with the goal of furthering the University's teaching and learning strategies.



RESEARCH

To advance understanding of evaluation and measurement practices and processes through research in support of teaching and learning.

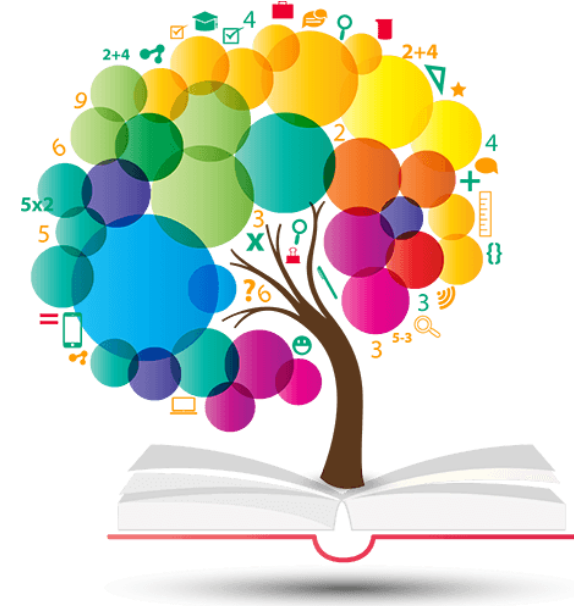
For internal reference only. Please do not copy or distribute the PPT slides without permission.

Fundamentals of Assessment

- Assessment serves multiple purposes:
 - ✓ Informing teaching and learning
 - ✓ Assessing student outcomes
 - ✓ Informing decisions about students

Fundamentals of Assessment

- Assessment drives the way we learn (Boud, D.)



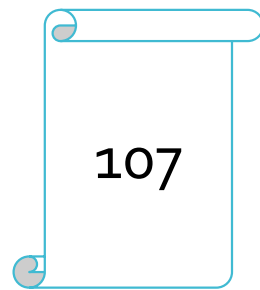
Fundamentals of Assessment

- When an assessment is designed or used to serve multiple purposes, evidence of **validity**, **reliability**, and **fairness** should be provided for each intended use" (*Standards for Educational and Psychological Testing*, AERA, APA & NCME, 2014)

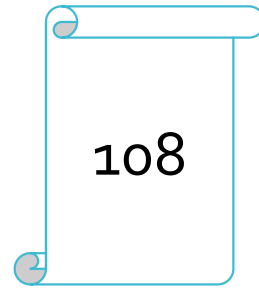
Fundamentals of Assessment - Reliability

- **Reliability** refers to the **consistency** of scores across replications of an assessment procedure.
- Reliability is a characteristic of assessment scores and not the assessment itself.

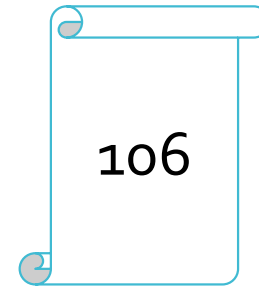
Mary's IQ test scores



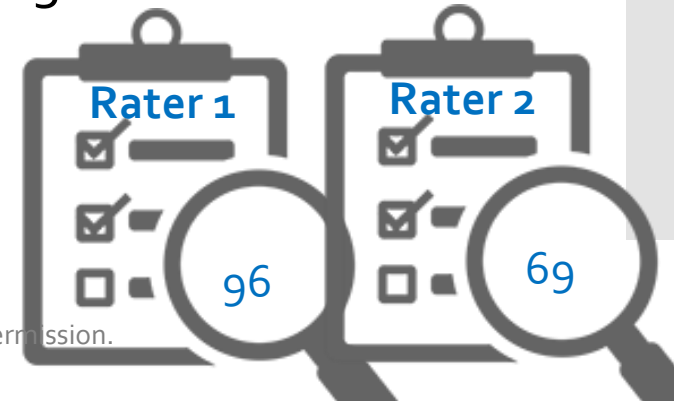
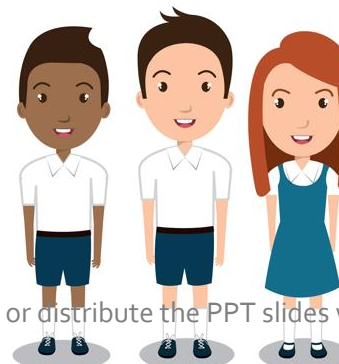
Score 1



Score 2



Score 3



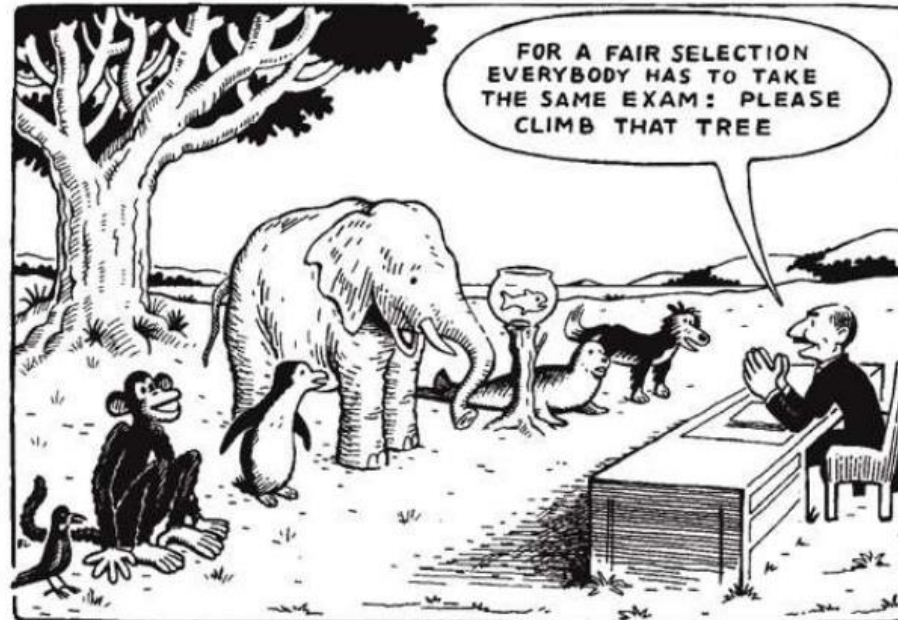
Fundamentals of Assessment - Validity

- We do not validate an assessment. We validate a particular interpretation/use of scores.
- Validity of scores is situation specific.
- Validity is a matter of degree. It is not all or none (“valid” or “invalid”).

“Statements about validity should refer to particular interpretations for specified uses. It is **incorrect to use the unqualified phrase “the validity of the test.”** (*Standards for Educational and Psychological Testing*, AERA, APA & NCME, 2014)

Fundamentals of Assessment - Fairness

- Providing students with
 - ✓ multiple opportunities to demonstrate their capabilities
 - ✓ curriculum and instruction that afford them the opportunity to learn
 - ✓ appropriate testing accommodations to address particular access needs



For internal reference only. Please do not copy or distribute the PPT slides without permission.

Fundamentals of Assessment - Fairness

- The equivalence of the construct being assessed is a central issue in **fairness**, whether the context is, for example, individuals with diverse special disabilities, individuals with limited English proficiency, or individuals across countries and cultures.



SUSTAINABLE DEVELOPMENT GOAL 4

Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all



Myths of Assessment

Myth 1

Assessments permanently categorize students.

Myths of Assessment

Myth 2

One assessment can serve all purposes.

Myths of Assessment

Myth 3

School and community leaders know how to use assessment for improvement. Teachers are trained to assess productively.

Myths of Assessment

Mthy4

Faulty interpretations of assessment results



Temptation is to discard tests, but what are the alternatives? Are they used properly? Are decisions made based on the alternatives reliable, valid, fair and free of error?

Fundamentals and myths of assessment

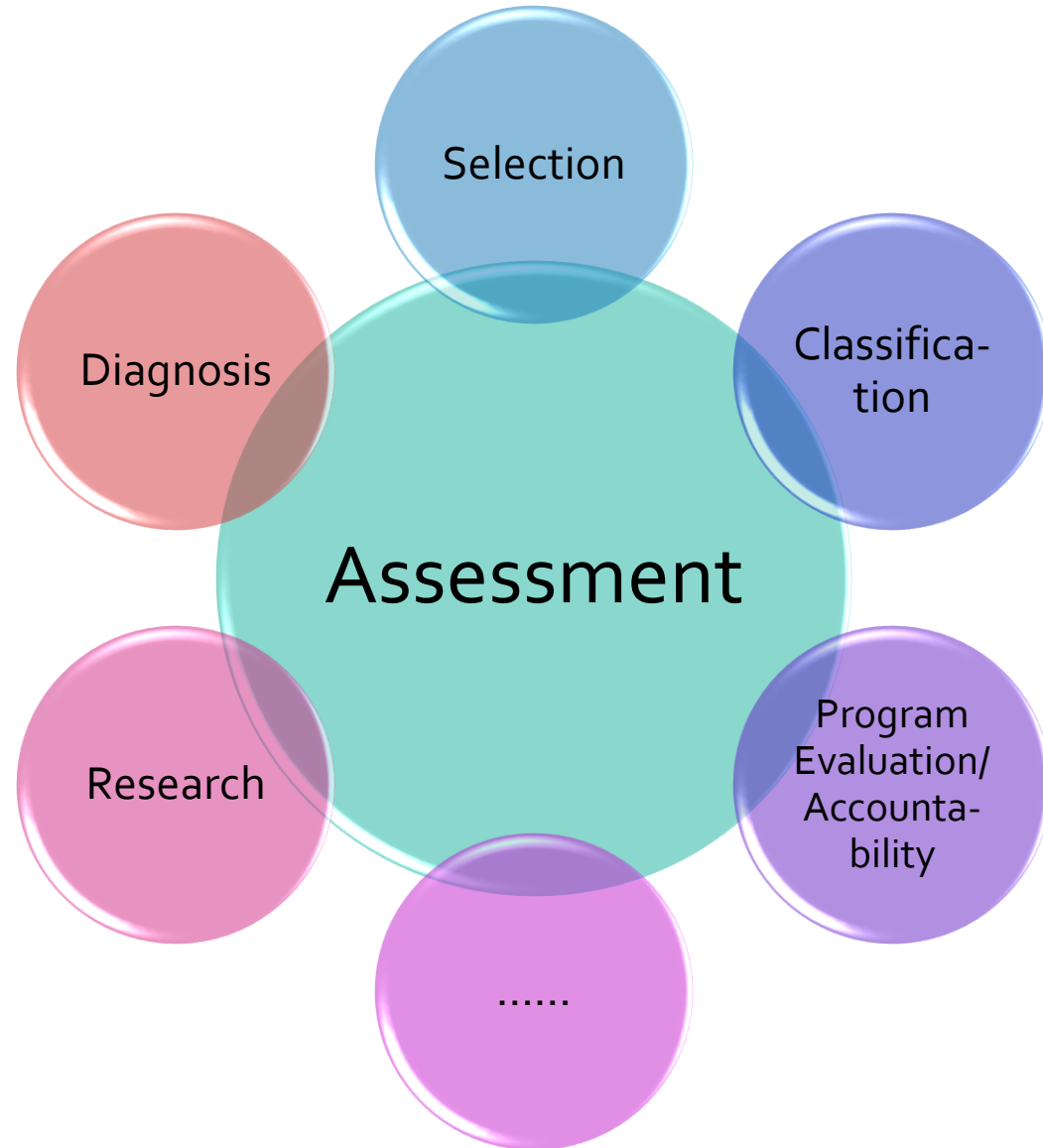
- No assessment is perfect, but the use of assessments can lead to better decisions.





Using assessment for quality assurance and enhancement: Emerging trends and HKU's practice

Major Uses of Assessment



Higher Education



Images: speakermagazine.com/how-do-you-define-value/

For internal reference only. Please do not copy or distribute the PPT slides without permission. novustechsolutions.com/value_addtion.php

Higher Education



Images: thechronicleofeducation.com/

For internal reference only. Please do not copy or distribute the PPT slides without permission. setasc.pnwbores.org

Quality Assurance in Higher Education

- Higher education institutions are becoming increasingly accountable to stakeholders. Evidence of student learning is of urgent demand (Liu, 2011).



• Outcomes Assessment:

The *systematic and ongoing* method of *gathering, analysing and using information* from various sources about an education programme and measuring programme/institutional outcomes in order to *improve student learning*.

Continuous process . . .



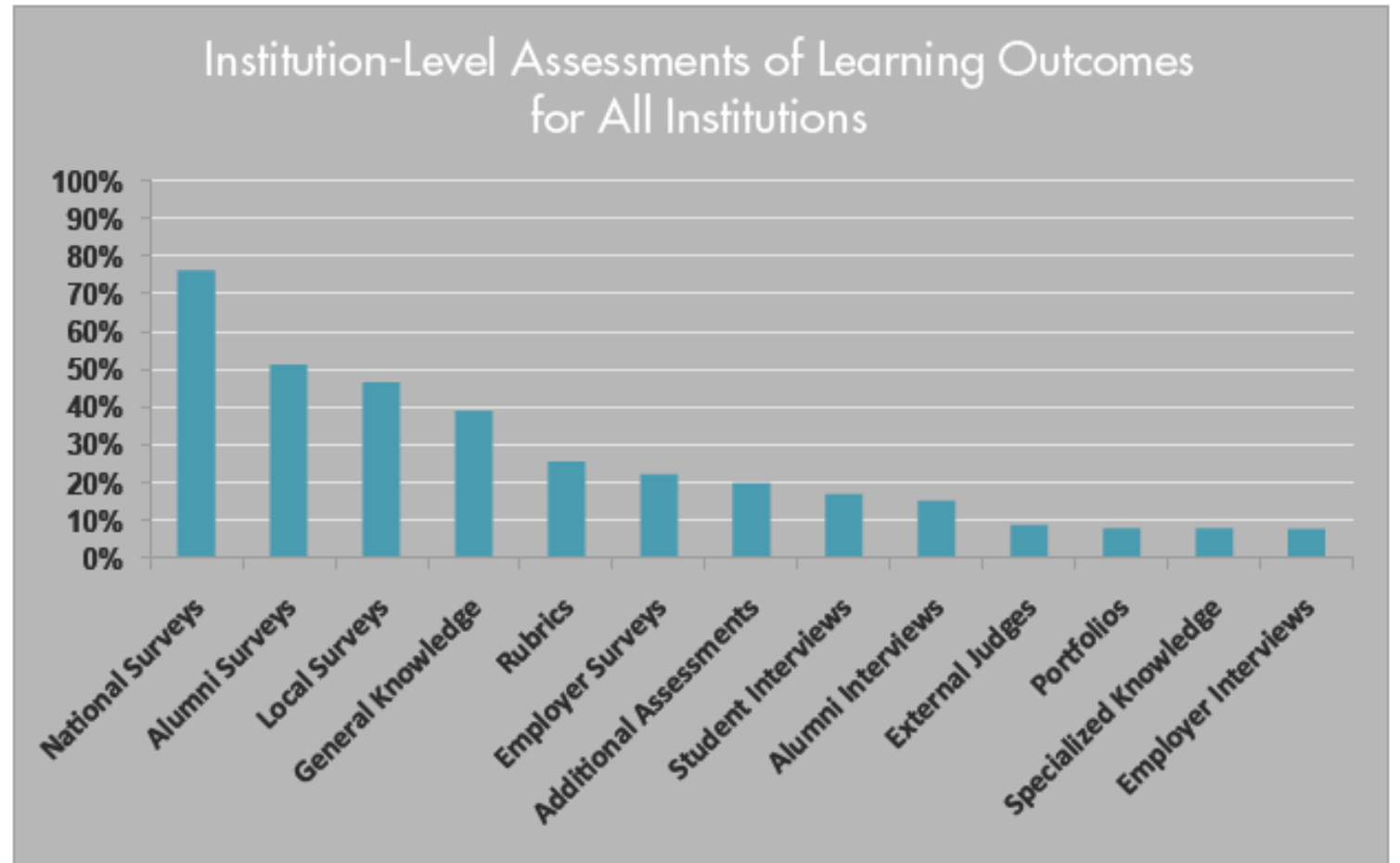
. . . instead of an event



<u>Direct Evidence</u>	<u>Indirect Evidence</u>
<ul style="list-style-type: none">• Student work• Capstone experience• Portfolios• Standardized tests• Embedded course assessment• Employer/internship ratings of students' performance• External examiner reports	<ul style="list-style-type: none">• Student surveys• Focus groups• Structured feedback session• Registration/course enrollment information• Department or program review data• Job placement indicators• Graduate school placement rates

National Survey of Provosts in US (National Institute for Learning Outcomes Assessment, 2009)

- The vast majority (92%) of all colleges and universities use at least one assessment approach or tool with institutionally valid samples; two-thirds of all schools use three or more.



Assessing
Student
Learning
Outcomes
in the form of
**standardized
tests & student
surveys**
- Global Practice

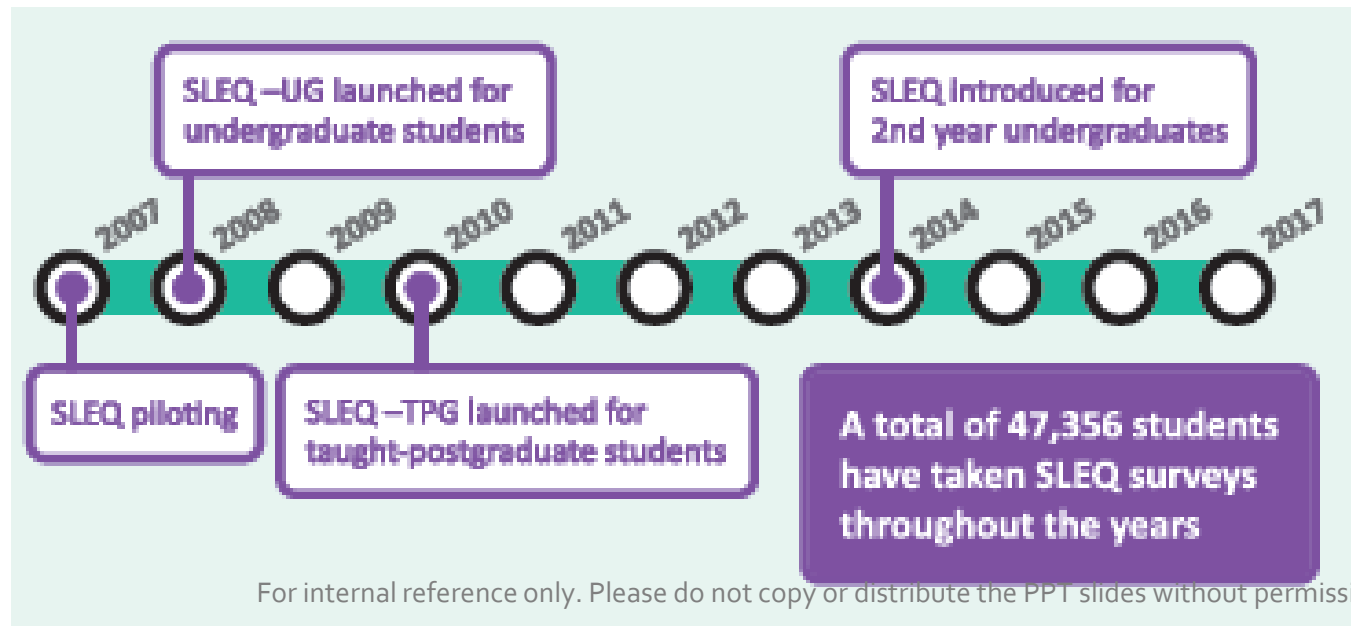


HKU Student Learning
Experience Questionnaire



Student Learning Experience Questionnaire (SLEQ)

- The aim of the survey is to collect data about students' perceptions on the achievement of their educational aims and their learning experience at HKU, with a focus on their experience of the degree curriculum as a whole.
- All 1st, 2nd, and final year undergraduate students and taught-postgraduate students are invited to participate in the survey exercise annually.





Teaching and Learning Environment:

- Active learning
- Feedback from teacher
- Good teaching (Motivation)
- Clear goals & standards
- Assessment for understanding
- E-Learning
- Learning communities
- Coherence of curriculum
- Using English as a medium of learning
- English support

Achievement of Education Outcomes:

- Pursuit of academic / professional excellence
- Critical intellectual inquiry
- Lifelong learning
- Tackling novel situations and ill-defined problems
- Critical self reflection
- Greater understanding of others
- Intercultural understanding
- Communication
- Collaboration
- Upholding personal and professional ethics
- Global citizenship
- Leadership
- Advocacy for the improvement of the human condition

Overall Satisfaction

Overall Comments:

- The best aspect of my University experience
- My university experience could have been improved if the University had done

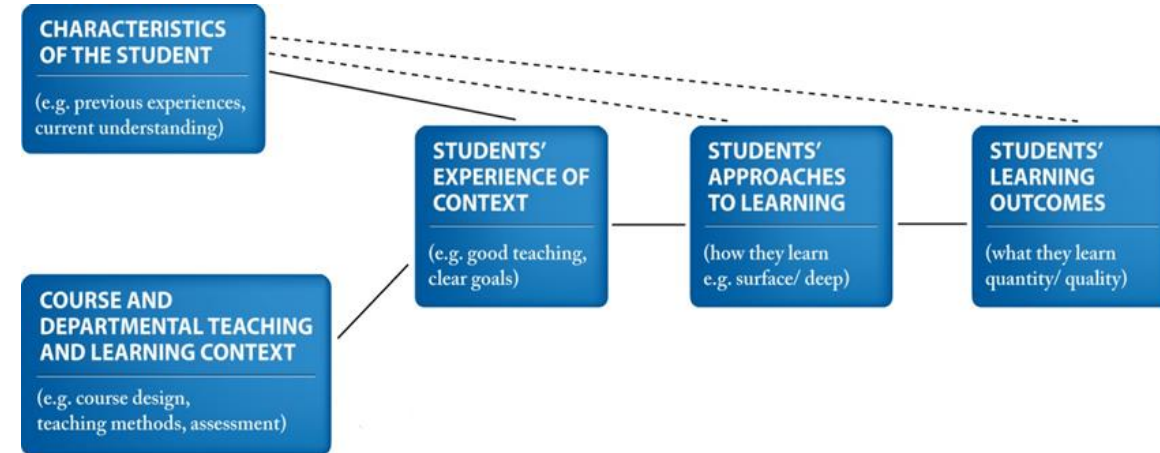


Common learning experience of the undergraduate curriculum



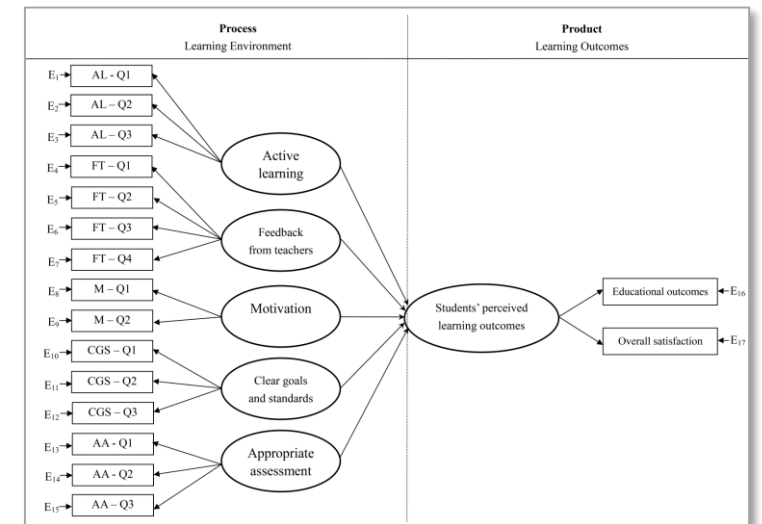
Scientific Quality of the SLEQ

- Gone through piloting and focus group discussions for **survey development** and revision. Psychometric properties were evaluated and revisited regularly to provide evidence of **reliability** and **validity**.



• SLEQ in higher education research:

- Zhao, Y., Huen, J.M.Y., & Chan, Y.W. (2017). Measuring longitudinal gains in student learning: A comparison of Rasch scoring and summative scoring approaches. *Research in Higher Education*, 58(6), 605-616. doi: 10.1007/s11162-016-9441-z.
- Zhao, Y., Huen, J.M.Y., & Prosser, M. (2017). Comparing perceived learning experiences of two concurrent cohorts under curriculum reforms in Hong Kong: A multiple-group confirmatory factor analysis approach. *Quality Assurance in Education*, 25(3), 270-286. doi: 10.1108/QAE-11-2016-0070.



Open-ended Comments from Students -

The best aspect of my university experience(s) is / are:

All the classes I have attended as of yet at HKU have been great. Not only is the content dealing with current technologies and times, the professors and tutors have constantly provided an **interactive** class. HKU has provided a platform to **meet people from around the world** and given us **the ability to experience different cultures** as well. This first year at HKU has not only been a learning experience but has been **great for personal growth**.

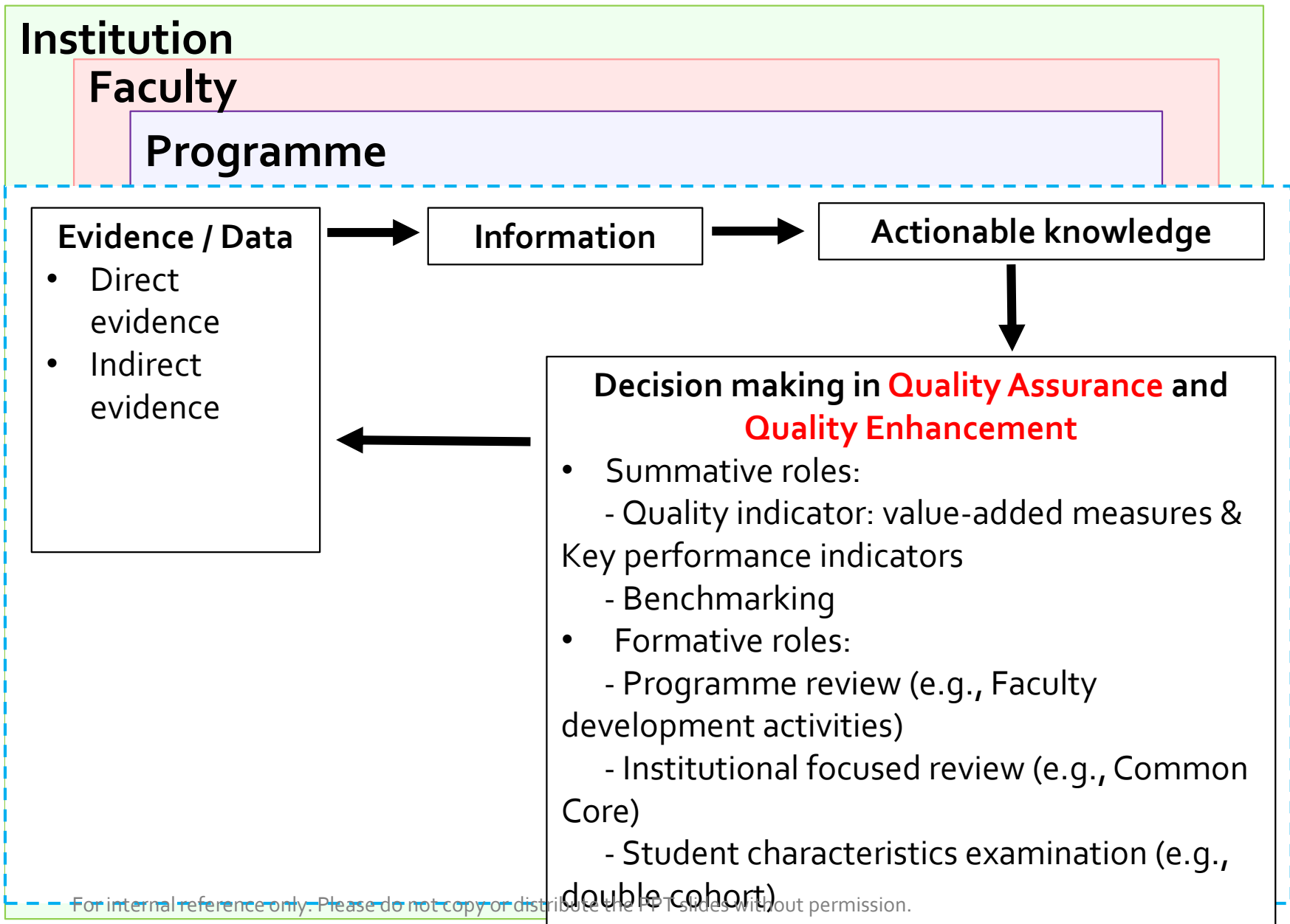
First year student,
Bachelor of
Medicine and
Bachelor of
Surgery,
2017/18

My university experience could have been improved if the University had done the following:

It is hoped that **more interactive** activities **across different discipline** of studies can be organised since as a Arts major student, I barely communicate with other students of Medicine Discipline and Law Discipline. I think meeting different people is also a key factor to develop different points of views.

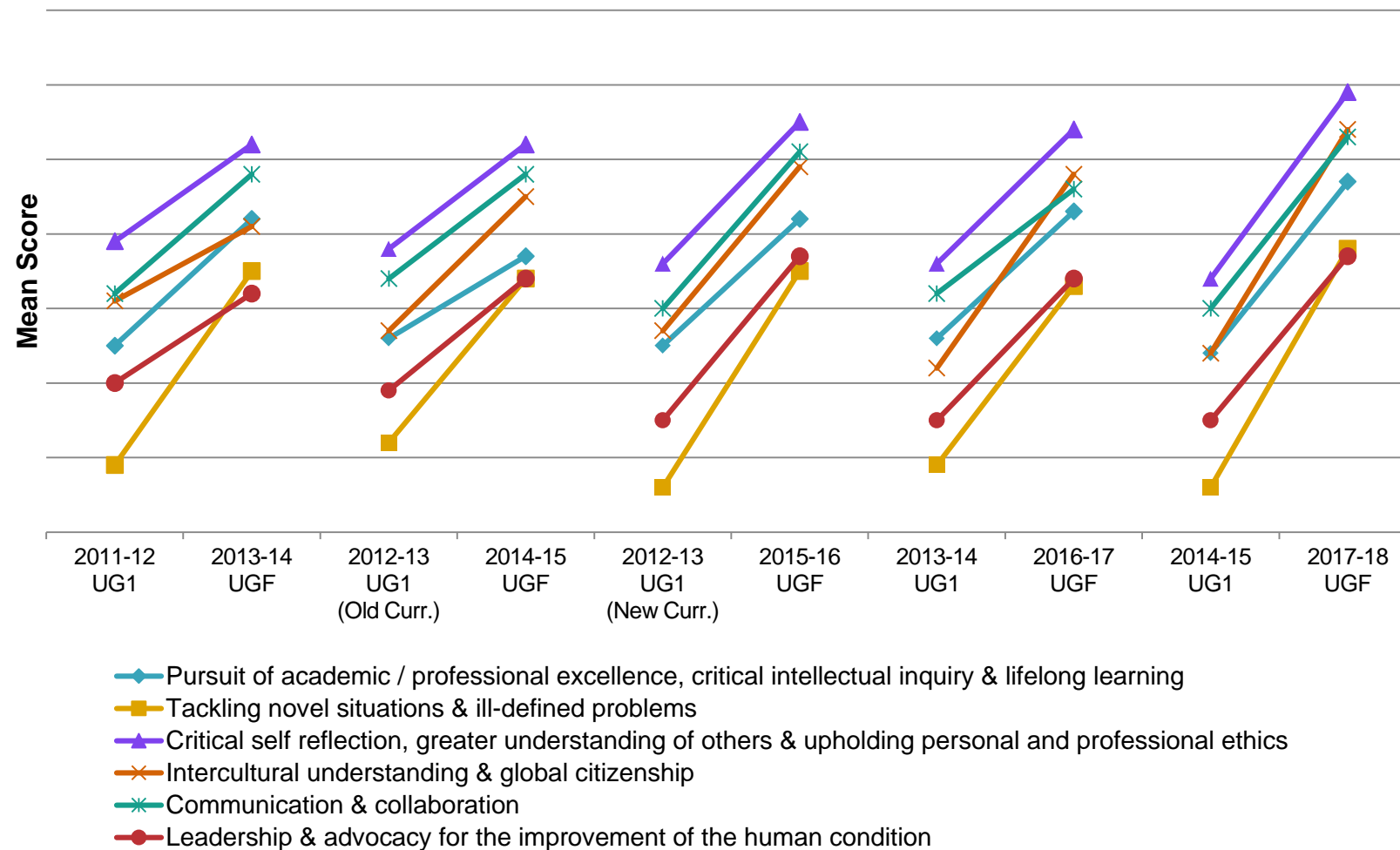
Final year student,
Bachelor of Arts,
2017/18

Using assessment for quality assurance and enhancement



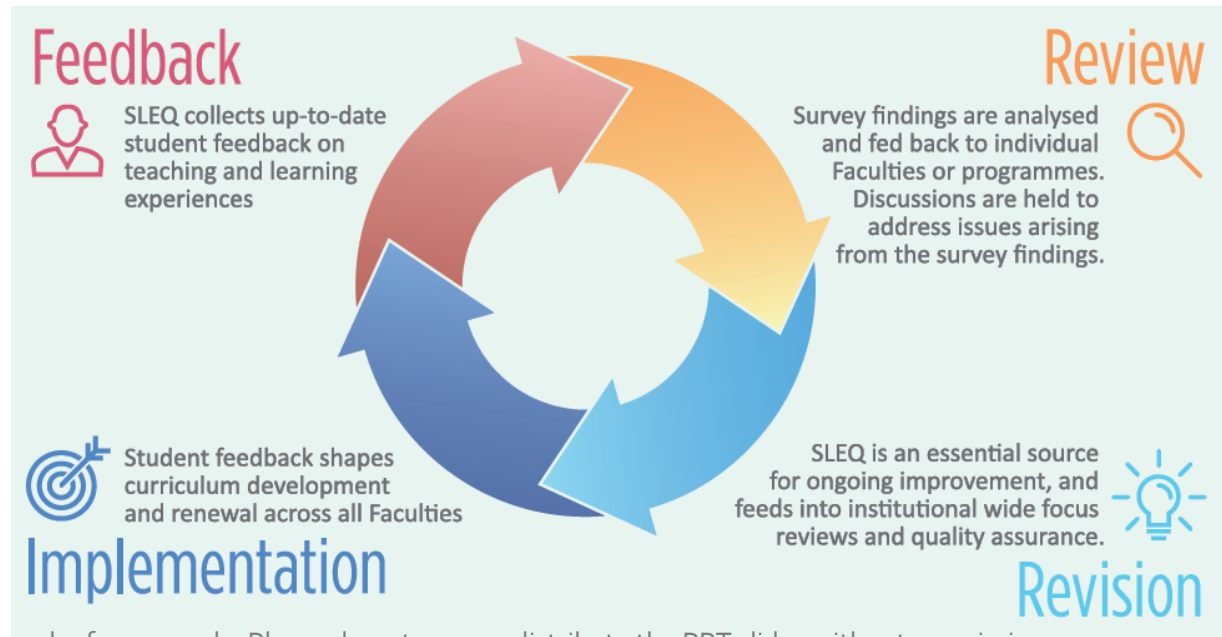
Value-added measures

Longitudinal Tracking for the Same Cohorts of Students in Their Perceptions of Achieving Learning Outcomes



Programme review

The survey findings are fed back to individual Faculties or programmes. Discussions are held with Deans, Associate Deans, Programme Directors and relevant staff members to address issues arising from the findings for the purpose of curriculum and pedagogy monitoring and renewal.



Close the feedback loop



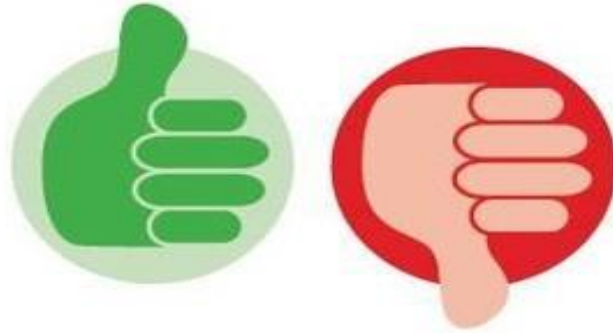
- **Observation based on the SLEQ**

In the qualitative comments, students indicated that they wanted more feedback, especially on examination performance.

- **Actions planned/taken**

- 1) Monitoring – It has been a policy of the Department of xxx to require a generic feedback report on examination from every course. Teachers may also opt for individual feedback in addition or as an alternative to the generic report. This new policy was implemented after the last SLEQ exercise.
- 2) Sharing good practice
- 3) Collecting students' views

Is collecting evidence of student learning as outcome assessment just an exercise of paper/data collection?



Advantages

- * Evidence-based approach & data driven decision making
- * Provide value-added measures
- * Facilitate the identification and implementation of effective practices

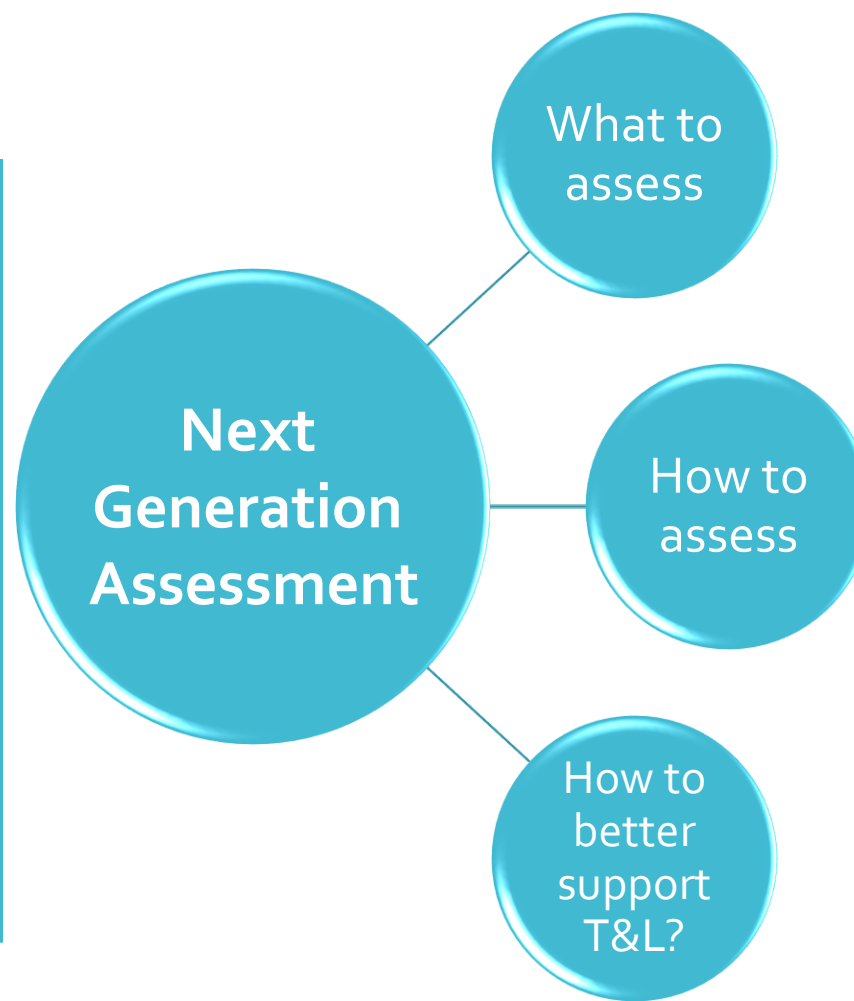
Disadvantages

- * Too much data
- * Time consuming



How can next generation assessment best support teaching and learning: Reflections and recommendations

Next Generation Assessment



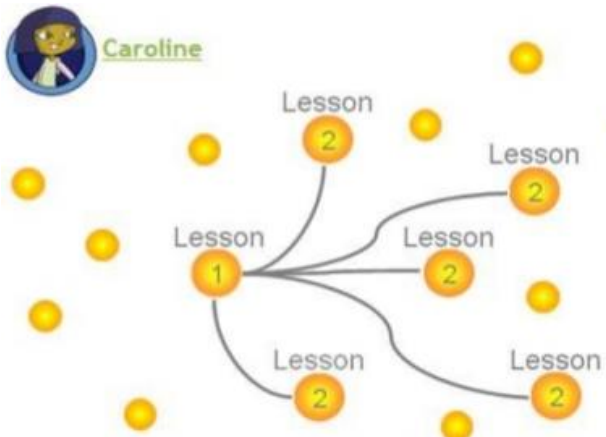
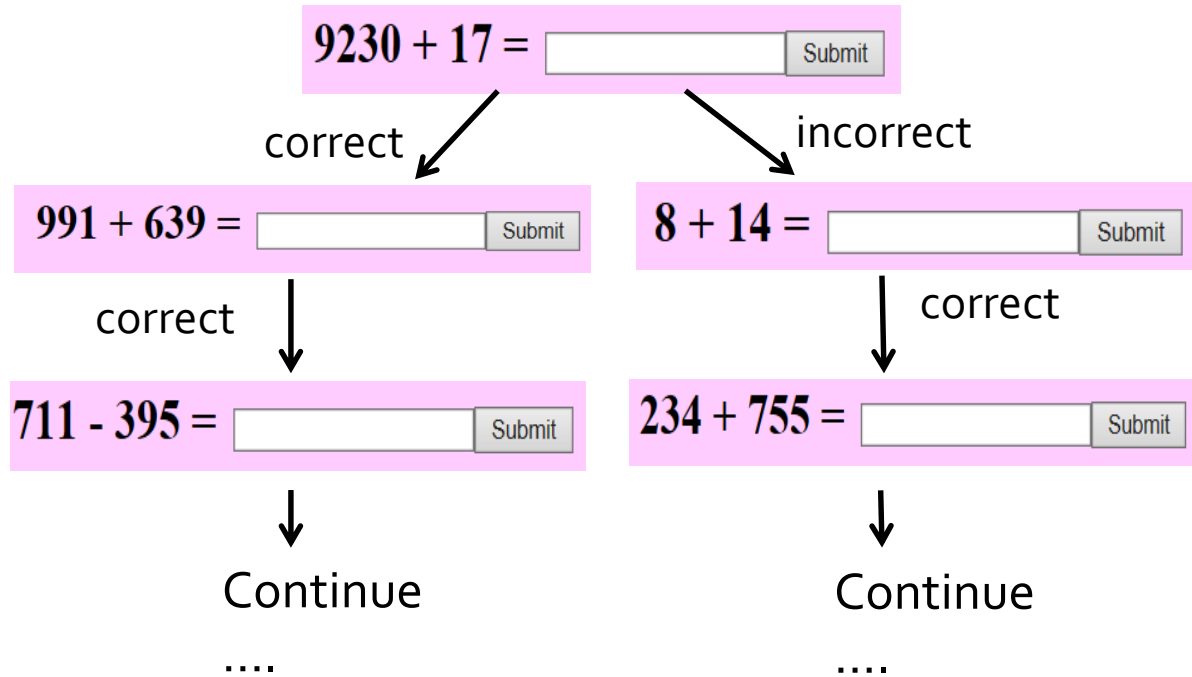
Artificial Intelligence
Big data
Computer technology
Data mining
Learning sciences
Psychometrics
.....



"As education enters a new world of assessment with advanced computer/human interaction, intelligent tutorials, adaptive feedback and data mining, the fact remains that there is much to be learned about how to appropriately leverage advances in technology, statistics, and the learning sciences toward the creation of a new generation of valid assessments" (Mazzeo, J.)

For internal reference only. Please do not copy or distribute the PPT slides without permission.

Personalized Learning



- Same ability ?
- Same level of mastery?
- Same set of items?

For internal reference only. Please do not copy or distribute the PPT slides without permission.

A Is For Adaptive

Monday, Jun. 17, 2013

TIME

By Kayla Webley



Photo-Illustration by Alexander Crispin for TIME

Source:

<https://www.slideshare.net/DreamBoxLearning/defining-adaptive-learning-technology>

Webley, K. (2013). A is for adaptive. Time, 181(23), 40-45.



Recommendation 2

In evaluation or accountability settings, assessment results should be used in conjunction with information from other sources.

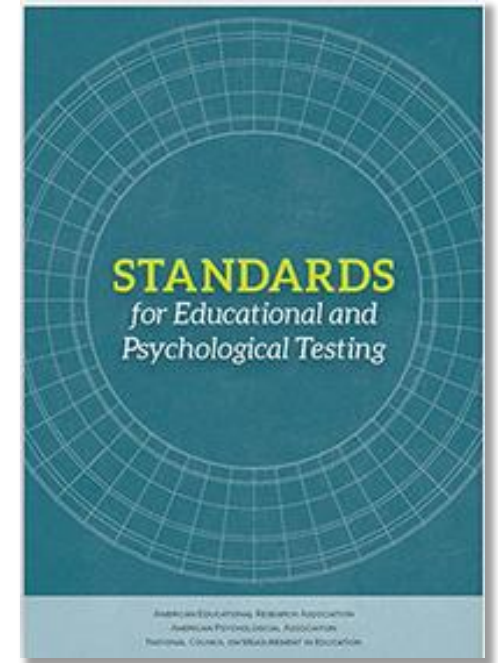


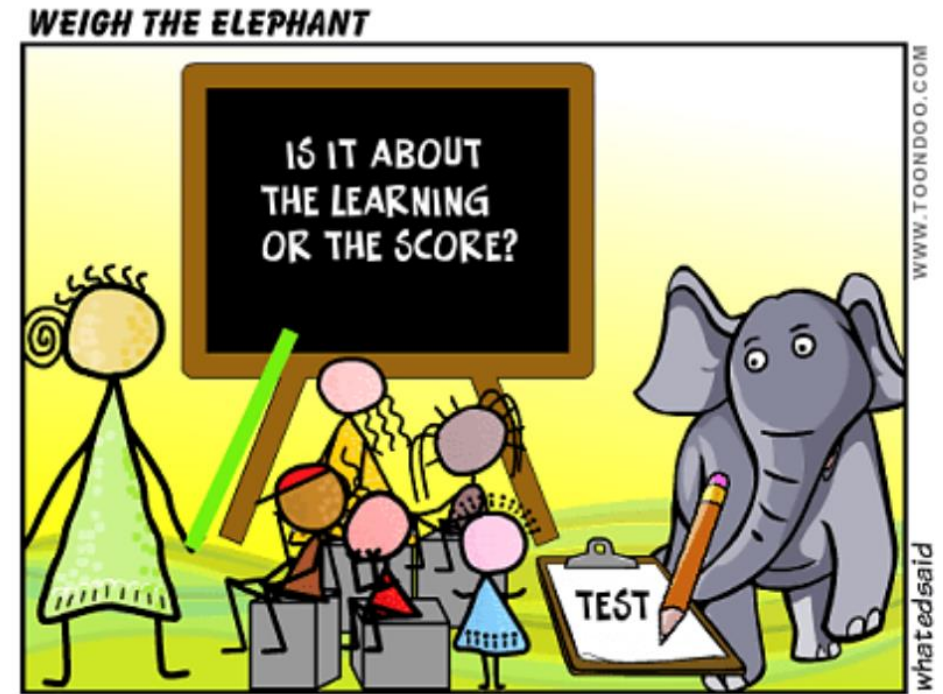
Image: ccctechedge.org



Recommendation 3

Assessment can fulfill both formative and summative functions for effectiveness/ improvement measures.

Balancing assessment *of* learning and assessment *for* learning.





Recommendation 4

Assessment results should be interpreted properly and used appropriately.

Assessment literacy makes for better learning.



Image: blog.edmentum.com



Recommendation 5

When interpreting group differences in assessment performance, relevant contextual information should be considered to enable meaningful interpretation of group differences.

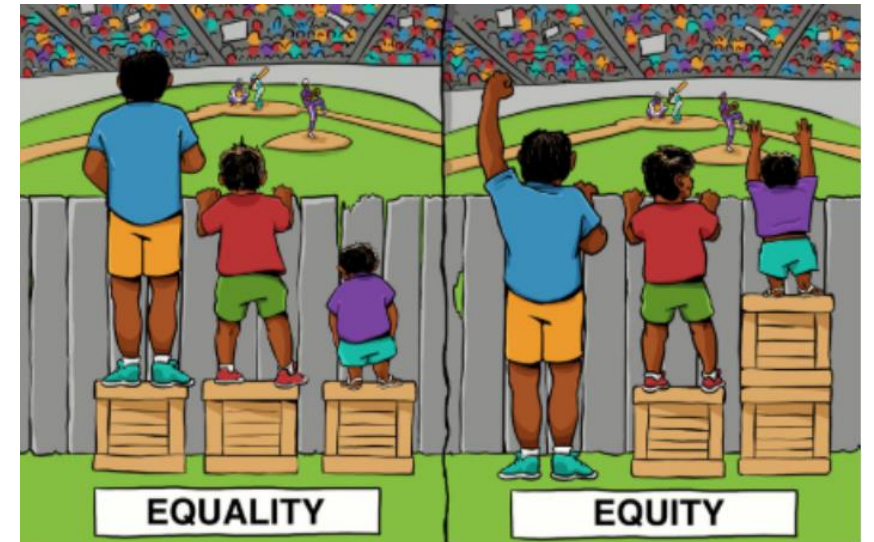


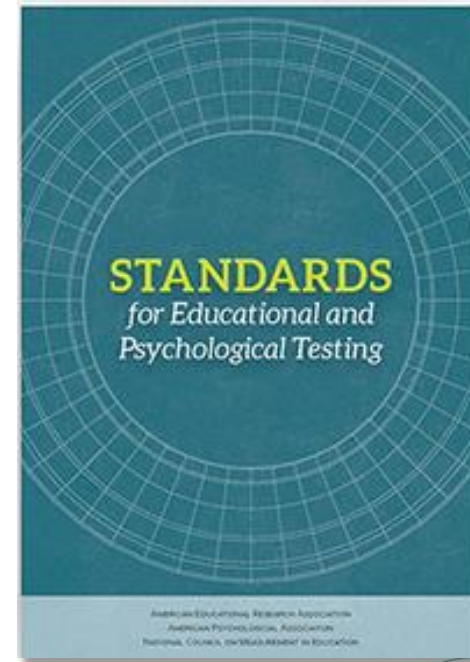
Image: culturalorganizing.org

For internal reference only. Please do not copy or distribute the PPT slides without permission.



Recommendation 6

Evidence of validity, reliability, and fairness for each purpose for which an assessment is used should be collected and made available.



The essential
(classical view)
of assessment

Assessment ~ the Latin root, *assidere*,
means to sit beside.

to assess means to sit beside the learner





t&lemu

Teaching and Learning Evaluation and Measurement Unit

Thank you!

myzhao@hku.hk

tlem.hku.hk



Using Evidence of Student Learning
for Improvement