

ATP ASIA

Annual Conference 2026 | Kuala Lumpur, Malaysia

From Measurement to Governance: What Assessment Leaders Need to Know About AI, Digital Evidence, and Proxy Governance Failure

Dongyue Louis Liu - Gakken Holdings Co., Ltd.



A leadership perspective on the transition from testing ecosystems to evidence ecosystems

- The market environment surrounding assessment products and services is changing rapidly.
- Recent developments such as AI, learning analytics, digital credentials, and learning records are expanding the evidence landscape.
- The question is no longer simply how you can provide good assessment.

The assessment market is changing from testing ecosystems to evidence ecosystems.

- We are no longer operating in a world where tests serve as the sole or even primary source of evidence about human capability.
- AI, learning analytics, digital credentials, portfolios, skills records, and continuous learning records have dramatically expanded the range of evidence available to make relevant decisions.

This is NOT an anti-testing argument. This shift creates new opportunities for many of us.

- Many forms of evidence now being promoted, such as credentials, portfolios, badges, learning records, and endorsements primarily describe what an individual has accomplished in the past.
- They provide potentially valuable information about experiences, achievements, participation, and prior performance.

Why tests still matter? (1) Because tests can provide current information.

- In evidence ecosystems, tests continue to perform a different function from other emerging evidences.
- Tests provides evidence of what an individual can demonstrate under specified conditions at a particular point in time.

Why tests still matter? (2)
It can serve as a form for safety net of social infrastructure.

- Not all individuals possess extensive portfolios, trusted learning records, or recognized credentials in standardized format.
- Tests provide a broadly accessible mechanism through which individuals can demonstrate capability, regardless of prior educational history or access to opportunity.

From Measurement to Governance: The issue is no longer merely producing evidence.

- Traditionally, the primary focus of assessment organizations has been “measurement.”
- As AI-generated outputs, predictions, dashboards, and other engagement metrics become increasingly common, the challenge expands beyond measurement.

Why Is Governance Becoming More Important?

- Recent technological changes have made evidence:
  - Easier/cheaper to generate, to share, to store, to analyze, to compare, to automate
- As evidence becomes more abundant and influential, governance becomes increasingly important.

Governance Challenges in Testing and Assessment.

- The governance challenges associated with evidence ecosystems can be understood through three domains:
  - operational governance
  - purpose governance
  - proxy governance

Operational Governance: Are we generating evidence under trustworthy conditions?

- For decades, testing organizations have invested substantial effort in controlling:
  - administration procedures (timing, security, accommodation)
  - psychometrics, scoring processes, quality assurance
- Assessment leaders must consider not only what evidence shows, but also how that evidence was generated.

Purpose Governance: Are we measuring the correct target construct?

- Purpose governance concerns the relationship between purposes and the constructs that assessment systems seek to measure.
- With recent technological advanced, conditions under which an individual perform a certain task has been changed dramatically.
- Still measure the same construct?

Proxy Governance: Are our measurements and indicators being interpreted and used appropriately?

- A proxy is something we can observe or measure that we use as a stand-in for something we actually care about.
- Proxy governance concerns the governance of measurements and indicators, particularly how they are interpreted, weighted, and used in decision-making.

Proxy Governance Failure (PGF): When an indicator evolves from supporting judgment to governing behavior.

- PGF occurs when an indicator gradually evolves from a tool that informs judgment into a mechanism that governs behavior, incentives, and institutional priorities.
- The challenge is no longer only measuring evidence. The challenge is governing increasingly influential evidence.

Why the risk of PGF is increasing with recent technological advancement?

- With recent digital technologies, we can
  - Generate far more indicators than before
  - Make indicators more visible
  - Make decisions more connected, embedded, and automated
- Under such conditions, indicators often becomes targets.
  - Goodhart's Law: "When a measure becomes a target, it ceases to be a good measure."

Are measurements and indicators being granted appropriate authority?

- Responsibilities expected from assessment leaders are expanding with technology.
- Assessment leaders must consider not only what evidence your service provides, but also how that evidence is consumed in society.
  - Consumer education on how your assessment service should be used is increasingly more important.

Leadership in the Evidence Ecosystem: Response to Governance Challenges

- Assessment leaders cannot assume:
  - Operational conditions will remain stable.
  - Assessment purposes/target constructs will remain unchanged.
  - Proxies/indicators will remain being viewed appropriately.
- Effective assessment leadership requires revisability.

Leadership in the Evidence Ecosystem: Response to Governance Challenges

- Assessment leaders cannot assume:
  - Operational conditions will remain stable.
  - Assessment purposes/target constructs will remain unchanged.
  - Proxies/indicators will remain being viewed appropriately.
- Effective assessment leadership requires revisability.

Assessment leaders will need to develop systems that remain open to revision.

- Strong assessment systems and business models are not those with the largest number of indicators or the most sophisticated dashboards.
- They are those that can revise their purposes, proxies, and operations to sustain future value.

Takeaway 1: The assessment market is shifting from a testing ecosystem to an evidence ecosystem.

- Assessment organizations are no longer operating in a world where tests are the only trusted source of evidence. AI, digital credentials, learning analytics, portfolios, skills records, and learning histories are expanding the evidence landscape.
- What role will testing play within a broader evidence ecosystem?

Takeaway 2: The future challenge is not only measuring evidence, but governing evidence through operations, purposes, and proxies.

- Assessment leaders must increasingly consider:
  - how evidence is generated (Operational)
  - what evidence is intended to represent (Purpose)
  - how evidence is interpreted, consumed, and used (Proxy).
- How should increasingly influential evidence be governed?

Takeaway 3: Effective assessment leadership requires revisability.

- Assessment leaders therefore need the ability to review, challenge, revise, and replace purposes, proxies, and operational practices when necessary.
- How can we sustain the future value of assessment services in a changing evidence ecosystem?

Proxy Governance Failure: Preprint

- Proxy Governance Failure in Data-Driven Education DX: When Learning Indicators Drift from Signals to Targets
- DOI: 10.35542/osf.io/fpmkv_v2


ATP
ASIA
Stepping into the Future!
Empowering Asia's Workforce
Thank You!