

ITC + ATP • TECHNOLOGY-BASED ASSESSMENT

Technology earns its place when it improves the inference.

*A quality-centered read of a
community-built field
guide.*

VERSION 1.1 • JULY 2025



Technology is already woven through the assessment lifecycle.

The guideline treats technology as a lifecycle condition—not a feature layer.



A community standard holds the whole chain to one public promise.

Capability is not the same as value.

CAPABILITY

≠

VALUE

- 01** Scale, automation, and engagement are capabilities.
- 02** Value depends on intended use, validity, fairness, and consequence.
- 03** The guideline asks what the score means—and for whom.

A COMMUNITY-BUILT FIELD GUIDE

This is a field guide, not a mandate.

100+ CONTRIBUTORS.
ONE PUBLIC PROMISE.

- CO-CHAIRS / EDITORS
JOHN WEINER + STEPHEN SIRECI
- 178 PDF PAGES • 11 GUIDELINE CHAPTERS
- VERSION 1.1 • JULY 2025 UPDATE

PREFACE + ACKNOWLEDGMENTS • ATP + ITC



New technology can add a new thing to the score.

It can expand the construct—or add what the test never meant.

EXPAND OR CONTAMINATE

WHAT TECHNOLOGY CAN REVEAL

Listening • speaking
Simulation • authentic tasks
Higher-order performance

WHAT TECHNOLOGY CAN ADD

Device • interface
Bandwidth • unfamiliar tools
Irrelevant variance

SAMUEL MESSICK (1989): “...include something that should be left out, or both.”

WHAT THE SCORE MEANS

The question is not whether the tool works; it is what the score means.

**“EVIDENCE + THEORY
SUPPORT INTERPRETATION”**

- AERA + APA + NCME (2014)
- INTERPRETATION OF SCORES
- PROPOSED USE



THE CONSTRUCT CAN GROW

Technology can expand what an assessment can see.

LISTENING • SPEAKING
SIMULATION • AUTHENTIC TASKS

- HIGHER-ORDER SKILLS
- RICHER EVIDENCE
- THE INTENDED CONSTRUCT MADE VISIBLE



Novelty is not a quality argument.

NOVELTY $\neq$ **EVIDENCE**

- 01 “Technology for the sake of technology” should be avoided.
- 02 Validation is the bridge.
- 03 Claims of engagement still require evidence.

DESIGN STARTS WITH DIVERSE USERS

Quality starts before the first item is built.

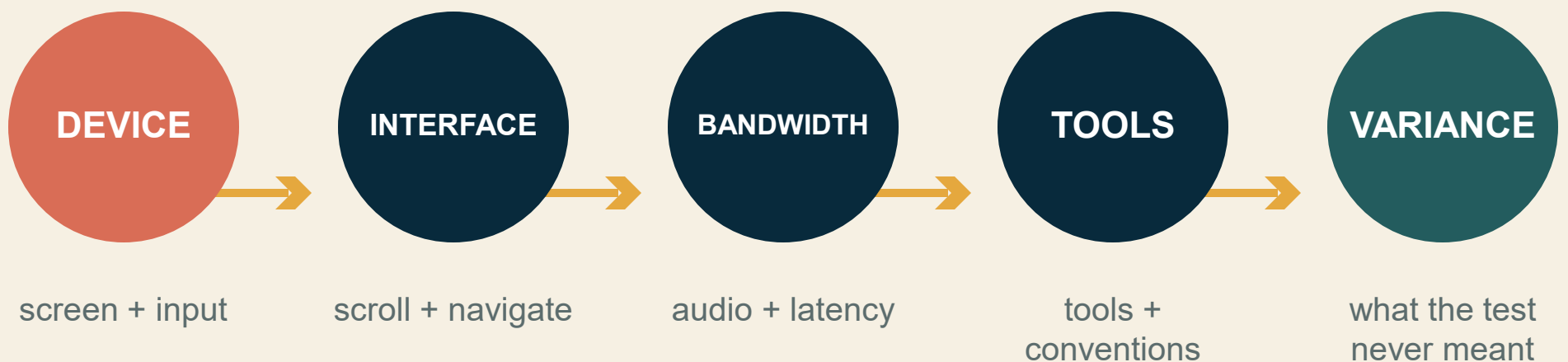
“CONSIDER DIVERSE USERS
FROM THE ONSET”

- MARTHA THURLOW • UNIVERSAL TEST DESIGN
- CONSTRUCT + ACCESS NEEDS
- DESIGN + PROTOTYPE + VALIDATE



The interface can become part of the construct by accident.

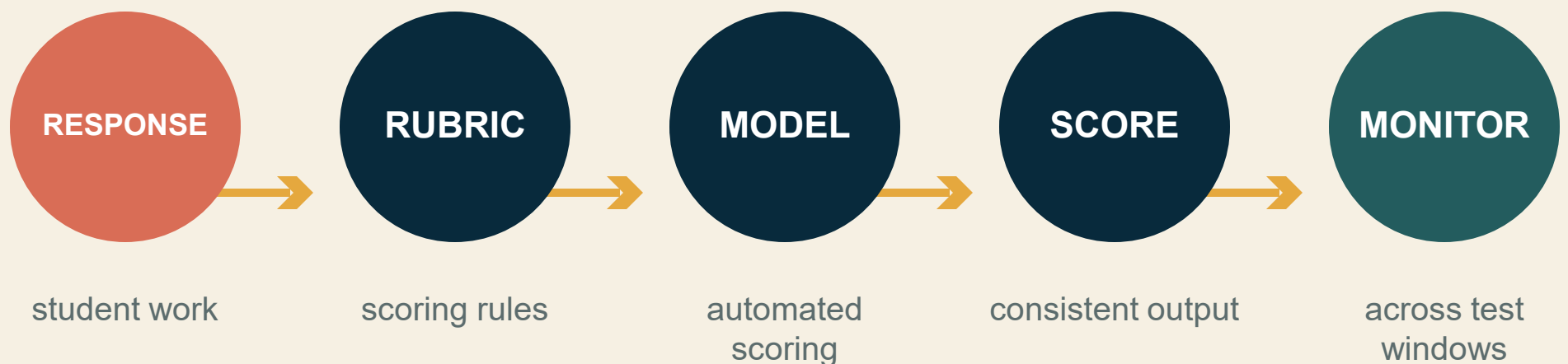
The assessment can measure the interface instead of the construct.



Every nuisance factor is a claim against the inference.

Automation can improve consistency.

A real benefit deserves a fair statement—and evidence.



SUSAN LOTTRIDGE: “...ensure scoring consistency within and across test administration windows.”

CONSISTENCY STILL NEEDS A HUMAN CHECK

Consistency still needs a human quality check.

“HUMAN SCORING
ALONGSIDE AUTOMATED SCORING”

- SUSAN LOTTRIDGE
- AUTOMATED + HUMAN SCORING
- QUALITY CONTROL



Human involvement belongs inside the decision loop.

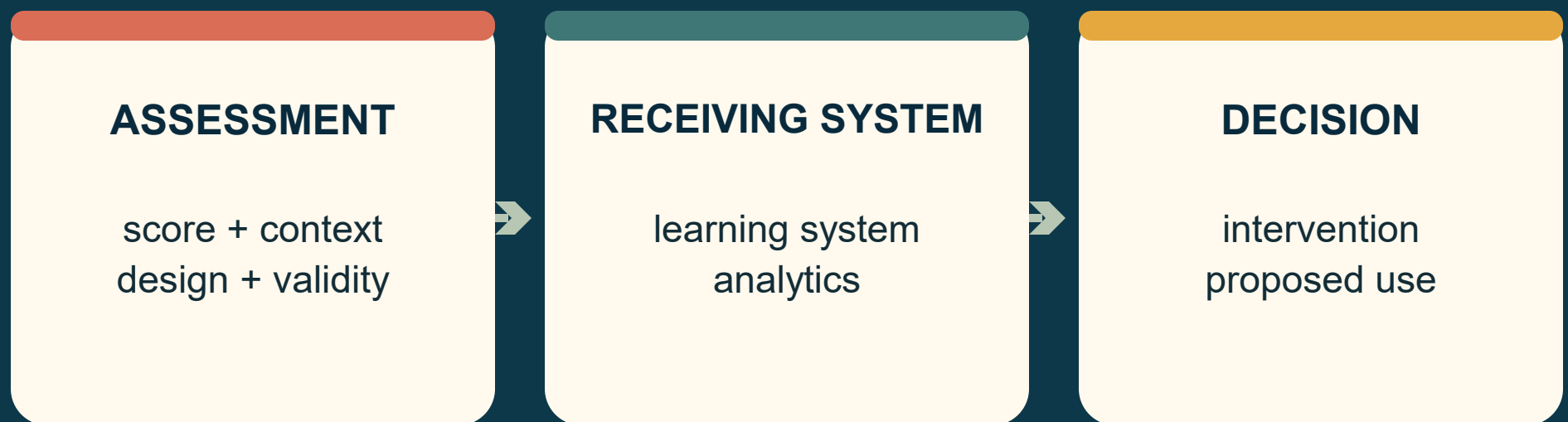
Human review is a system principle—not an exception path.



SUSAN LOTTRIDGE: “...humans in the decision-making loop [are] an essential component of AI systems.”

An assessment result travels—and can lose its meaning.

Validity evidence must move with the data.

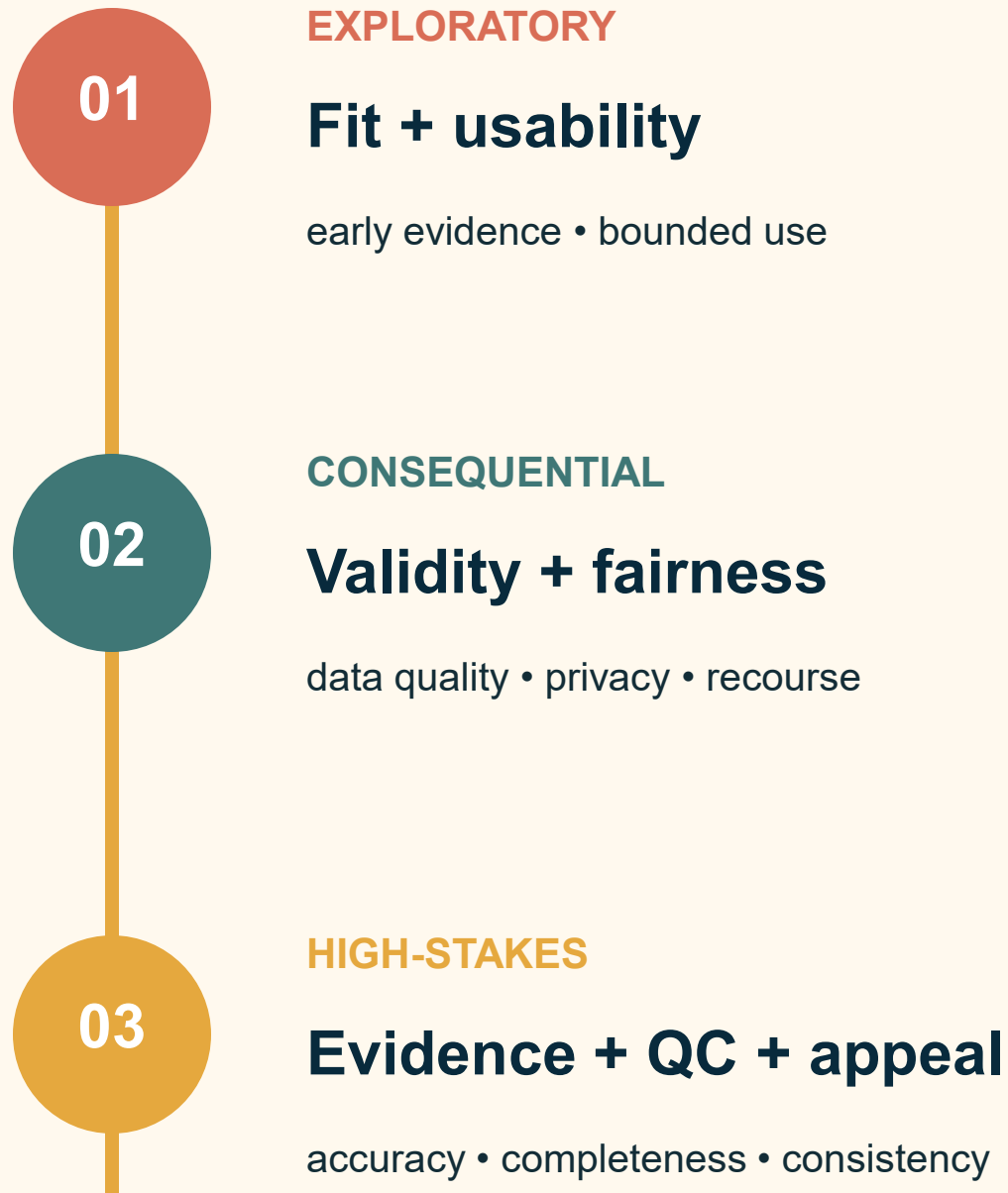


VALIDITY EVIDENCE



MICHELLE BARRETT + DAVID
KING
CHAPTER 6 CONTRIBUTORS

The higher the stakes, the more evidence and data quality matter.



A qualitative ladder—not a quantitative scale.

AI opens new assessment possibilities.

Responsibility and accountability have to scale with the opportunity.

ADAPTIVE + SCORING	AUTHORING + TRANSLATION	SIMULATION + NEW TESTS
OPPORTUNITY 01	OPPORTUNITY 02	OPPORTUNITY 03
adaptive scoring • fraud detection	item authoring • translation • reporting	simulation • richer tasks • new test types

MIKE SPARLING • BEN TAYLOR • JOHN WEINER • WAYNE CAMARA

Generative AI widens the use cases—and the risk surface.

Appropriate processes and capabilities must mitigate potential risks.

USE CASES & RISKS

WHAT IT CAN ASSIST

Job analysis • blueprints
Item writing • graphics
Translation • scoring • reporting

WHAT IT CAN AMPLIFY

Limits • hallucination • bias
Privacy • security • integrity

HUMAN INVOLVEMENT + OVERSIGHT

Deployment is not the end of the evidence cycle.

01

PLAN + BUILD

Intended use + safeguards

construct • users • consequence

02

VALIDATE + DEPLOY

Evidence + fairness + quality

launch within human decision rights

03

MONITOR + RETIRE

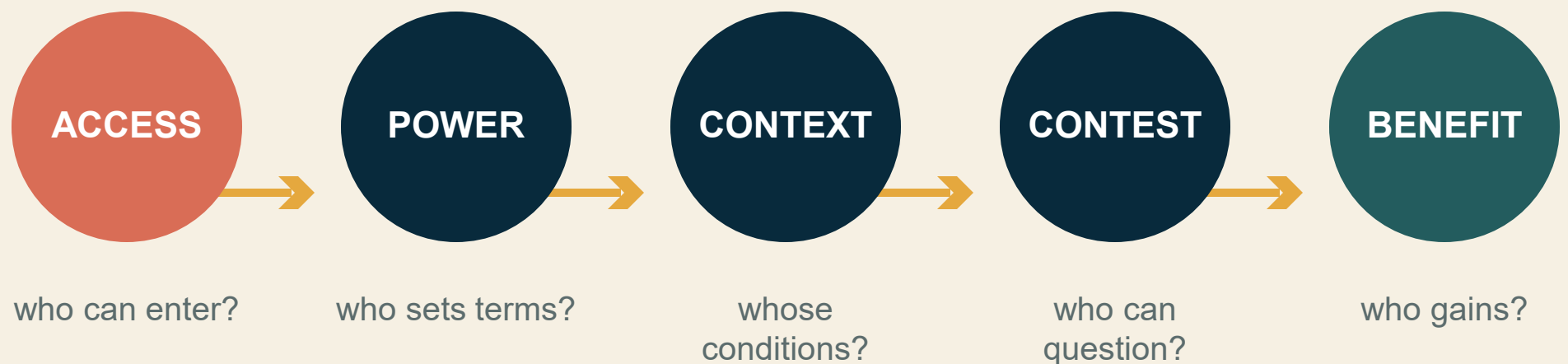
Drift + appeal + correction

feed learning into the next cycle

JOHN KLEEMAN • ONGOING EVALUATION

Quality includes who can access, contest, and benefit.

Consider power, authority, access, and privilege at every stage.



JENNIFER RANDALL + MICHAEL WALKER • EQUITY ISSUES IN TBA

A technically elegant system can still fail the person.

“STABLE, WELL-LIT, PRISTINE HOMES” CANNOT BE ASSUMED.

- JENNIFER RANDALL + MICHAEL WALKER
- HOME + PRIVACY • DEVICE + BANDWIDTH
- ACCESS + ANXIETY • DESIGN FOR REAL CONDITIONS



Assessment is an ecosystem, not a feature.

01

DESIGN + DELIVERY

Items + platform + access

construct • users • context

02

SCORING + REPORTING

Human + automated + meaning

validity • fairness • use

03

DATA + CONSEQUENCE

Integrity + lineage + intervention

privacy • security • recourse

One quality rail runs across the whole system.

THE TURN

Quality is not the brake on innovation.

QUALITY IS THE
OPERATING CONDITION.

- VALID ENOUGH TO INTERPRET
- FAIR ENOUGH TO USE
- ACCOUNTABLE ENOUGH TO LEARN

SYNTHESIS • PARTS I–IV



Carry five questions into the next AI experiment.

If the answer is unclear, the innovation is not ready for consequential use.



INTENDED USE • VALIDITY • ACCESS • PRIVACY • HUMAN DECISION RIGHTS

The guideline leaves us with a productive question.

WHAT EVIDENCE



WORTH USING?

- 01 For whom?
- 02 At what stakes?
- 03 With what appeal when it fails?

THE DURABLE TEST

The durable test remains the same.

MIKE SPARLING • BEN TAYLOR
JOHN WEINER • WAYNE CAMARA

“Valid, reliable, fair and unbiased, accessible, and secure—without introducing irrelevant variance or unintended consequences.”

[Use the guideline for the next experiment →](#)

PART IV CONCLUSION • ATP + ITC • VERSION 1.1

