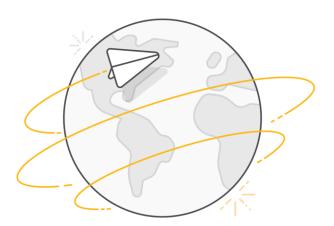
"Are we there yet?" The promise of the Holodeck for the Future of Education

Alina A. von Davier | Duolingo April, 2023



The world is changing.



Education is being redefined by

Global events

Digitally-mediated life

Technology and AI

The Promise of the Holodeck

Learning & Assessment are personalized, adaptive, and immersive



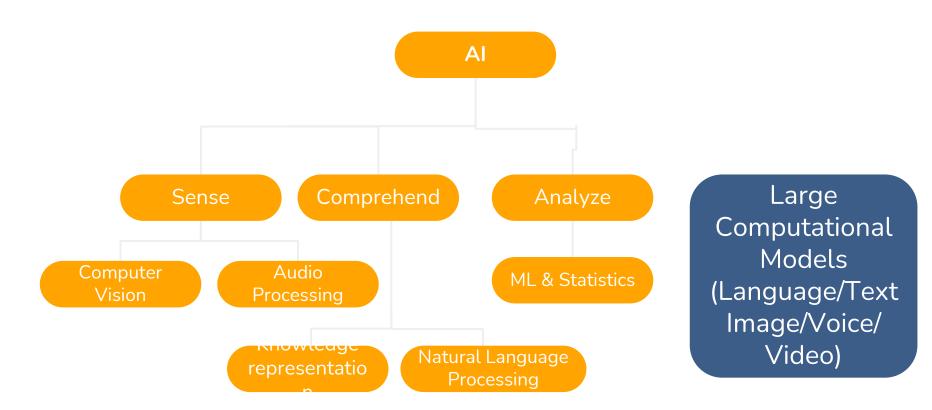
William Riker (Jonathan Frakes) entering a Holodeck simulation. Star Trek/Screenshot/Memory Alpha

Star Trek's Holodeck: from science fiction to a new reality

Learning & assessment (L&A) has shifted to digital platforms

- A lot as happened in the pagnerative A!
- Are we there yet?
 - Digital-first learning & assessments (DfL&A)
 - Personalized & adaptive L&A
 - Game-based and VR/AR L&A => immersion
 - Collaborative L&A
- What are novel fairness considerations for AI and AI-adjacent methods for digital L&A?
 - Do we need novel fairness considerations for digital L&A with head-spinning momentum of innovation associated with large language models (e.g., GPT-4), and large multimodal models (e.g., image generation)?

Generative Al



ChatGPT swept us off our feet

"Any sufficiently advanced technology is indistinguishable from magic"

— Arthur C. Clarke

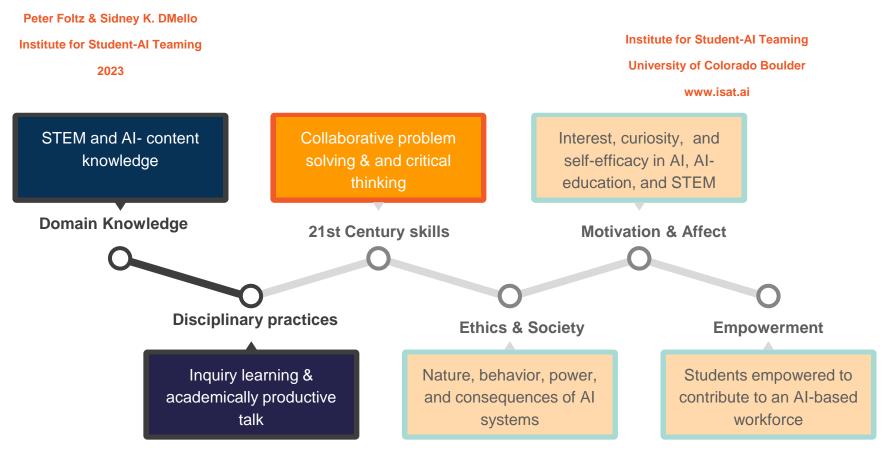
Emerging Trends for Learning & Assessments

- 1. Holistic
- 2. Engaging, gamified, immersive
- 3. Personalized diagnostic & recommendations
- 4. Analytics for evaluation
- 5. Digital and multimodal delivery
- 6. Delightful experience (any time, anywhere)



The power is in the integration of frameworks & tools

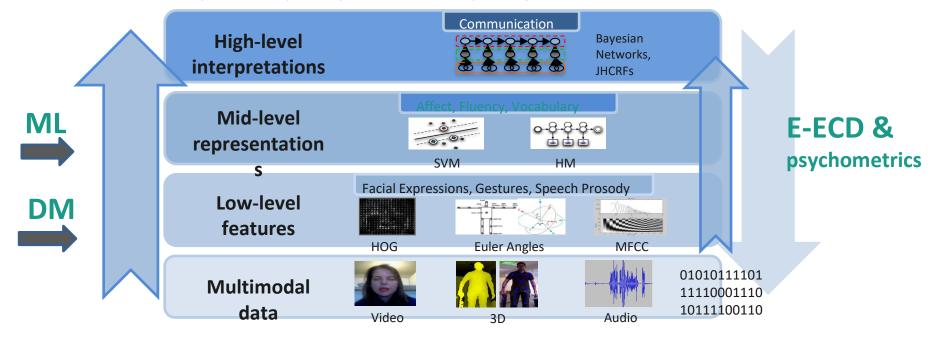
iSAT integrates Al-education in science & tech courses to provide measurable learning outcomes*



^{*}Adapted slide

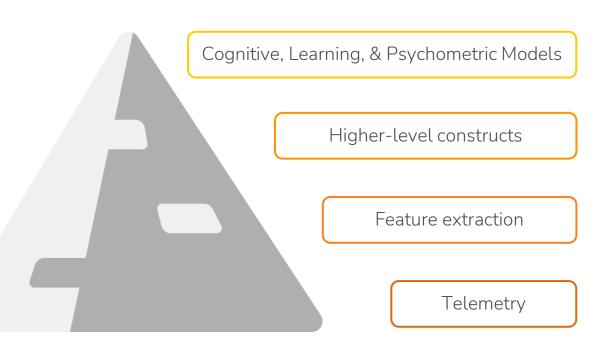
Computational Psychometrics: Hierarchical Inference Model for Multimodal Data for L&A

- Design the system (learning &/or assessment) based on theory
- Identify **constructs** associated with competency of interest (*Expanded Evidence*)
- Find evidence for these constructs from low-level multiple sensory data (*Evidence about Evidence*) --Khan (2015); von Davier (2015)



Computational Psychometrics: An Integrative Framework for AI-based Capabilities

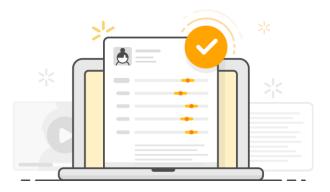
Computational psychometrics is a blend of theory-driven psychometrics & machine learning methods used to measure latent abilities in real time.



We need a paradigm shift in assessment!



Digital-first assessments as a solution for the new normal



"The future is already here. It's just not evenly distributed, yet."

— attributed to W. Gibson

Characteristics of digital-first assessments

Born digital

- Digitally-mediated constructs
- Designed with computational models & AI-algorithms in mind
- Theory-based & evidence-centered
- Valid & reliable

Integrated

- Fluid infrastructure
- Integrative theoretical frameworks
- Integrated (automatic) tools & databases

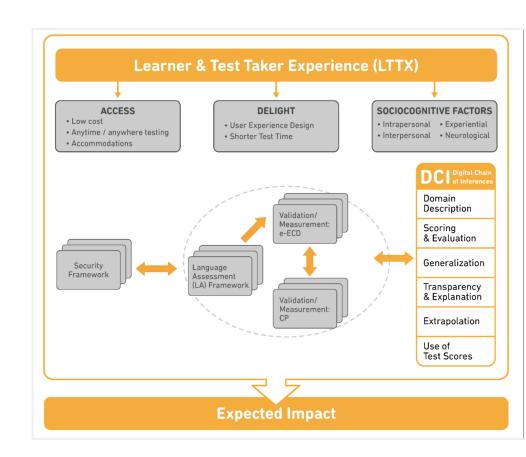
Superior test-taker experience (TTX)

- Available on demand, anytime, anywhere
- Online testing with remote proctoring



DLFA ecosystem

- Integrated learning & assessment frameworks
 - Digitally-mediated constructs
 - Security, Design, Measurement
 - o LTTX
- Guiding questions for fairness auditing in the ecosystem
 - Where is AI used in the frameworks?
 - What practices support audits within & across frameworks?



AI & ML within the Duolingo English Test

Item development

Content generation
Question/prompt generation
Option generation
Review models (FAB & IQR)
Review tools (Items factory & AQuAP)
Pilot tools (Pre-pilot platform)

Test administration

Items microservice Adaptive algorithm Monitoring computer use Interactive tasks

Scoring

Automated speech recogition
Feature development
Scoring speaking
Scoring writing
Score monitoring (AQuAA)
Pilot tools (Full-legth practice test)

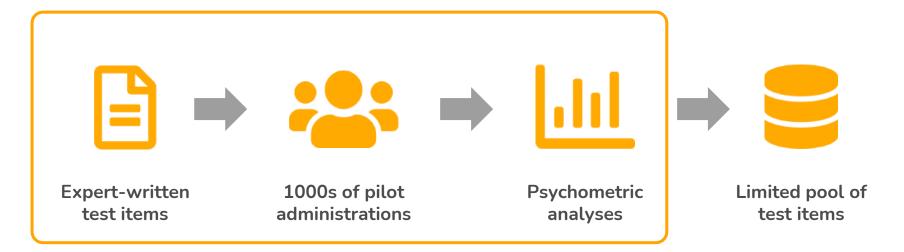
Security

Proctor portal Al behavior flags Computer use reports

Score reporting

Automated test taker report Score report API for test score users

Traditional test development



Time-consuming & expensive
The pilot samples may not represent the target
population

Creation at scale: Human-in-the loop Al

Goal: Humans leverage AI to perform a task

- Difficult for both humans and Al
- Have "expert" eyes on all content on the test

"The Loop":

- Humans design the items
- Al systems make predictions or generate content
- Humans review and/or edit the outputs
- Human labels and edits are used to improve Al

At scale and with a high degree of efficiency

Content creation at scale: Al-Driven Item Generation

Generative AI (GPT4, BERT, ChatGPT, Stable Difusion, etc)

can be used to automatically generate

- Fill-in-the-blank vocabulary questions
- Multiple choice keys and distractors
- Free response comprehension questions
- Writing and reading prompts
- Dialogues
- Images

Passage generation

GPT-3/4 can mirror style, format, and content given extremely few examples

Goal

Generate self-contained passages in a range of domains - especially at an academic level

See Attali et al. (2022) for a detailed description of the creation of an interactive reading task

See Attali et al. (2022) for a detailed description of the creation of the creation of an interactive seed the compact of the creation o

EXAMPLES

Topic: Sociology

Title: The process of socialization

Passage: Sociology is the study of society and social behavior. One of the main areas of study in sociology is socialization, the study of how people become social beings. Sociologists are interested in...

Topic: Business

Title: Business ethics

Passage: Business ethics is a term used to describe the standards of conduct used by businesses, corporations, and individuals in the business world. It refers to the behavior and practices of individuals....

Topic: Biology

Title: The history of evolution

Passage: More than 4 billion years ago, life began on Earth. According to evolutionary theory, living organisms have changed over time. Evolution occurs over millions and tens of millions of years. The process of evolution can be summed...



Title: The History of X-rays

Passage: In 1895, German physicist Wilhelm Röntgen produced the first photographs that showed the skeletal structure of living things. These pictures were made when x-rays (electromagnetic radiation) passed through the air...

Technology

Passage: The medical

Passage: The medical profession can be traced back about 5,000 years.
Primitive medical care in the form of herbal remedies was common and early civilizations tended to look after the sick and injured.
Although medical knowledge and technology have advanced over time...

Title: The Use of Computers in Medical Research

Passage: Science and technology have improved health care greatly over the past few decades. Computer technology allows researchers to store, process, and analyze massive amounts of data much more efficiently.

This has resulted in many lifesaving discoveries, such as...

Recent Developments

Responsible AI for digital learning &

as seasingent Large Language Models

- Duolingo partnership with OpenAl
 - GPT-3 & GPT-4 for content generation
- Responsible AI Standards for the Duolingo English Test (2023)
 - First <u>Responsible AI Standards</u> for an assessment
 - Open for public comment
 - Aligned with DET's long term use of Al as a digitalfirst assessment

The Future is Now!

Science fiction is any idea that occurs in the head and doesn't exist yet, but soon will, and will change everything for everybody, and nothing will ever be the same again.

— Ray Bradbury

DET 2023

Thanks! Questions?

avondavier@duolingo.com

Acknowledgement:
The Duolingo English Test R&D Area

