Leveraging ideas from adaptive testing to adaptive learning

*The HERA showcase*

Meirav Arieli-Attali  Alina von Davier
ACTNext, by ACT
MARC, 2019
Outline of the talk

• Introduction
  • Adaptive Testing
  • Adaptive Learning

• Ideas from adaptive testing and formative assessment
  • Adaptation by difficulty
  • Self-adaptation
  • Multi-stage Adaptivity
  • Assessing partial knowledge; hints & feedback
  • Learning progressions and diagnostic tests
  • Development framework: Evidence Centered Design

• Application: design of the HERA system (+demo if time allows)

• Findings from early pilot
Adaptive Testing: Intentions & Outcomes

• **Goal:** improve measurement
  • Increase reliability (reduce measurement error)
  • shorten tests
  • Maximize test information
  • Statistical models: primarily IRT

• **Outcome:** assign question at the ability level of test taker
  • Usually test takers will get items that they have a probability of 50% to answer them correctly
  • Usually, item selection is defined item-by-item
  • Item selection is by **difficulty**
  • ➔ Similar experience for all individuals (in terms of relative test difficulty)
  • ➔ high performing test takers are not bored, low performing are less frustrated
Adaptive Learning

• Tutoring systems
  • Adaptivity by content/skill
  • Rule-based or algorithm-based
    • Within task (step loop) vs. between tasks (task loop)
  • Mastery-model for knowledge
  • Provide feedback on correctness
  • May provide hints
  • Statistical models: mainly Bayesian Knowledge Tracing (BKT)
Ideas from Adaptive Testing and Formative Assessment

• Why and how ideas from assessment can leverage learning?
• Rigor methods for ensuring validity
  • Adaptation by difficulty; assessing ability on-the-fly
  • Self-adaptation – research findings
  • Multi-stage Adaptivity
  • Assessing partial knowledge; hints & feedback
  • Learning progressions and diagnostic tests
  • Development framework: Evidence Centered Design
Adaptation by difficulty; Assessing ability on-the-fly

• In contrast to adaptive learning (usually by content/skills)
• Ability assessed on-the-fly → ability measure reliable and valid
• Valid & stable measures of item difficulty (not just expert evaluation)
• Can be flexible – change the window of input to estimate ability (to allow measure of change/learning)
• Based on psychometric models (IRT; CDM); can also adopt Elo and Urning models / mathematically also linked to BKT (Deonovic et al., 2019)
• ➔ can combine adaption by difficulty & skill (CDM)
Self-adaptation – Research Findings

• Giving test takers choice to choose the difficulty (Arieli-Attali, 2016; Rocklin & O’Donnell, 1987; Wise et al., 1992)
  • Test takers overall choose level of difficulty that corresponds to their ability level
  • Test takers overall choose difficulty of 65%-75% probability correct (CAT algorithm often selects items at 50% difficulty)
• If test takers are rewarded for difficulty of items – they tend to challenge themselves more
Multi-stage Adaptivity

• Instead of selecting item-by-item, can select a group of items (testlets) adaptively
• Content balanced
• Information Targeted at Cut versus at Ability
• Influence of Multiple Cut Scores
• Tree-based multistage adaptive
Assessing partial knowledge; hints & feedback

• Assessing partial knowledge (Ben Simon & Budescu, 1997)
• Assessing knowledge when feedback and multiple attempts are provided (Attali & Powers, 2010; Attali, 2011)
• Assessing knowledge/ability when hint is used (Bolsinova et al., 2019)
Learning progressions and diagnostic tests

• Designing task models based on a map of skills that reflects progression ➔ student model
• Diagnostic models ➔ statistical models to diagnose where students are
Development framework: Evidence Centered Design

A Schematic Representation of the Models in the ECD Framework (Mislevy et al., 2006)

Expanded framework

- Assembly Model
- Evidence Models
- Task Models
- Presentation Models
- Delivery Model

- e-Assembly model
- e-Proficiency model
- KSA-change model
- KSA model

- e-Evidence model
- Transitional-Evidence model
- Observational-Evidence model

- e-Task model
- Task-support model
- Task model

ACTNEXT
THE HERA SHOWCASE
Meet HERA!

An Adaptive

- Holistic
- Educational
- Resources and
- Assessment System

for Science

--Research-based prototype
--Bridging assessment & learning
--Using science simulations as context
--Adaptive scaffolding (self-adaptive help options)
--Adaptive sequencing
Task model from HERA
An item with scaffolds after incorrect response

Target skill

Proximal precursor (+ distal)

Initial + distal+ proximal
The HERA pilot
May-Aug 2018

• Collaboration between ACT, ACTNext, Smart Sparrow & PhET
• Pre-pilot in May 2018 – to examine functionality
• Large pilot in Aug 2018 – to examine learning-supports usage
  • Participants:
    • 2,775 Amazon Mechanical Turk; in 10 conditions; each participant completed 3 lessons
  • Materials:
    • Six lessons (Physics; Chemistry; Biology) - Four lessons include simulations as preview (two sims from PhET); Each lesson includes 10 items with learning supports
  • Conditions:
    • Examine different ways to offer learning supports: (1) before response or after; (2) with or without cost; and (3) with different cost systems

*Lessons topics: Restitution, Specific Heat, Hooke’s Law, Beer’s Law, Hinges, Self-pollination
Item with Learning Supports

Learners can choose between three learning supports:
1 – **Rephrase** the question
2 – **Break-down** the question to steps
3 – **Teach me** the content by solved example or full explanation

Differential cost was:
- **Rephrase** = 1 beaker
- **Break-it-down** = 2 beakers
- **Teach me** = 3 beakers

Equal cost : 2 beakers per support
Learners Preference of Learning Supports

Trends:
• Overall, learners prefer “Teach me” (on average 45%), over “Break-it-down” (on average 25%) and “Rephrase” (on average 15%).
• When scaffolds are offered before answer - learners are using more help, particularly more “Teach me” but less “Rephrase”.
• When scaffolds are offered at a cost – students use less help (by about 3%- 4% compared to no cost).
• When the cost is differential as 1-2-3, there is almost no change in the distribution compared to no cost; equal cost of 2-2-2 increases the use of “Teach me” in the expense of “Rephrase”
Summary

- We implemented ideas from CAT, MST and Self-Adapted tests; adaptivity by difficulty, content & student choice
- Adaptivity by units – based on multistage adaptive and balancing content
- Based on progression of skills and student maps (student model)
- Based on statistical models when using hints, feedback, & multiple attempts
Next steps

• Based on pilot with adults ➔ developing a prototype with more content
• Will pilot with middle school students in 2020
Thank you

meirav.attali@act.org
alina.vonDavier@act.org