Listening. Learning. Leading.®

A Research Programme for Collaborative Assessments

Alina A. von Davier

Educational Testing Service

OUTLINE

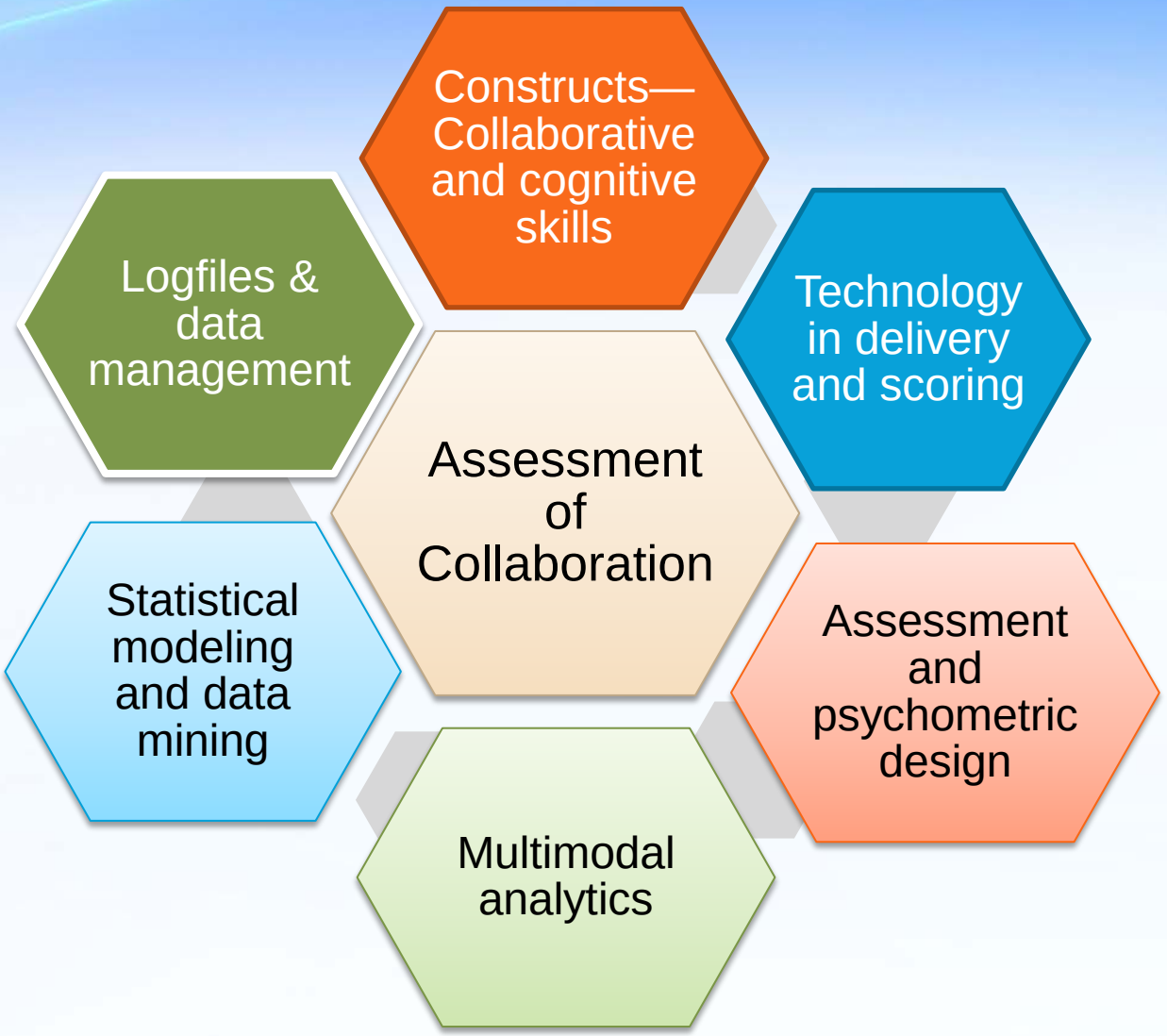
- Motivation for a comprehensive research programme
- Assessment design & data challenges
- Statistical models that account for dependencies

- The focus is on the measurement of

  - Collaboration skills and
  - Cognitive skills of
    - individuals
    - teams

- Many contributors:
  Lei Liu, Jiangang Hao, Saad Khan, Mengxiao Zhu, Yoav Bergner, Patrick Kyllonen, Deirdre Kerr, Diego Luna-Bazaldua (TCC), Zouwei Wang (UMi), Lei Chen, Peter Halpin (NYU), Art Graesser (UMe)

A RESEARCH PROGRAMME AT ETS



EXAMPLES OF INSTRUMENTS & DATA

Tetralogue

Construct: Science

Auto-scoring of dialogue

```
<?xml version="1.0" encoding="UTF-8"?>
<gameLog>
 <session>
  <sessionID>CPSSingle</sessionID>
  <attemptID>1</attemptID>
  <playerID>A17F1JAIWT9GYH</playerID>
  <sessionExtData/>
 <events>
  <event>
   <eventName>Initialize</eventName>
   <eventStartTime>2013-12-13 02:09:22</eventStartTime>
   <eventBy>
  </eventBy>
   <eventTo>-1--1</eventTo>
   <eventResult>Started new session</eventResult>
   <eventSceneID>NA</eventSceneID>
   <eventExtdata/>
  </event>
  <event>
   <eventName>Changed slide</eventName>
   <eventStartTime>2013-12-13 02:09:22</eventStartTime>
   <eventBy>
  </eventBy>
   <eventTo>Slide Manager</eventTo>
   <eventResult>New current slide is: 000 - INIT EVERYTHING</eventResult>
   <eventSceneID>NA</eventSceneID>
   <eventExtdata/>
  </event>
  <event>
```

Assessment design

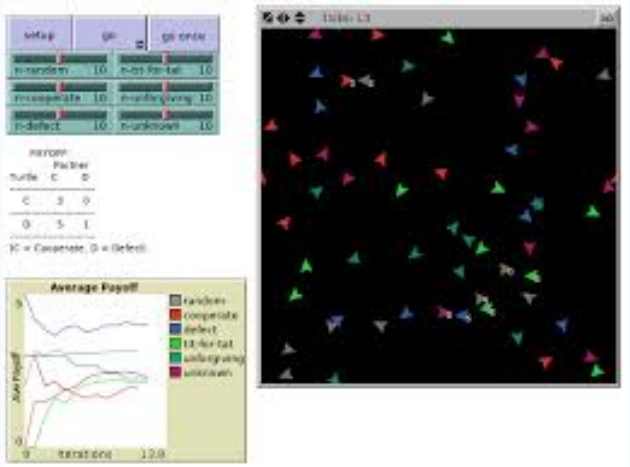
Log file data management

Data-mining & dynamic models

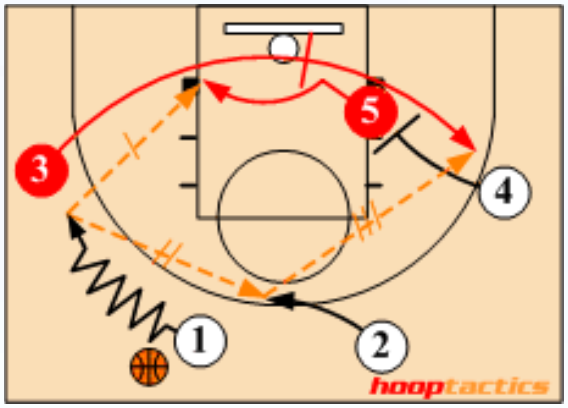
EXAMPLES OF INSTRUMENTS & DATA

NAEP TEL

Agent Based Modeling



Basketball



DATA, SCORES, AND INTERACTIONS

Analyzing data from an assessment that includes collaborative problem solving (CPS) tasks involves several modeling aspects that are not encountered in the traditional tests

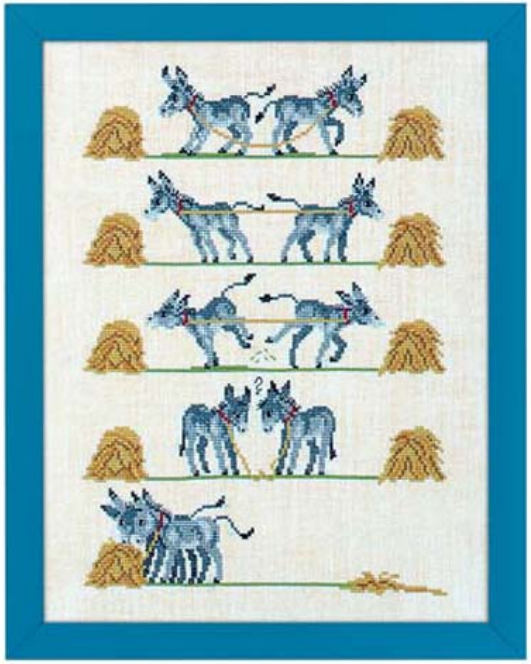
- Items and people exhibit several types of dependence, that is, the local independence assumptions do not hold (more to come)
- People behave differently when they interact in teams from when they work alone (e.g., O'Neil, 1999),
- The individual domain skills might not correlate highly with the team's outcome (e.g., Woolley et al., 2010).

Consider multiple scores

- an individual domain score measured in isolation,
- an individual domain score measured in collaboration, and
- an overall team score.

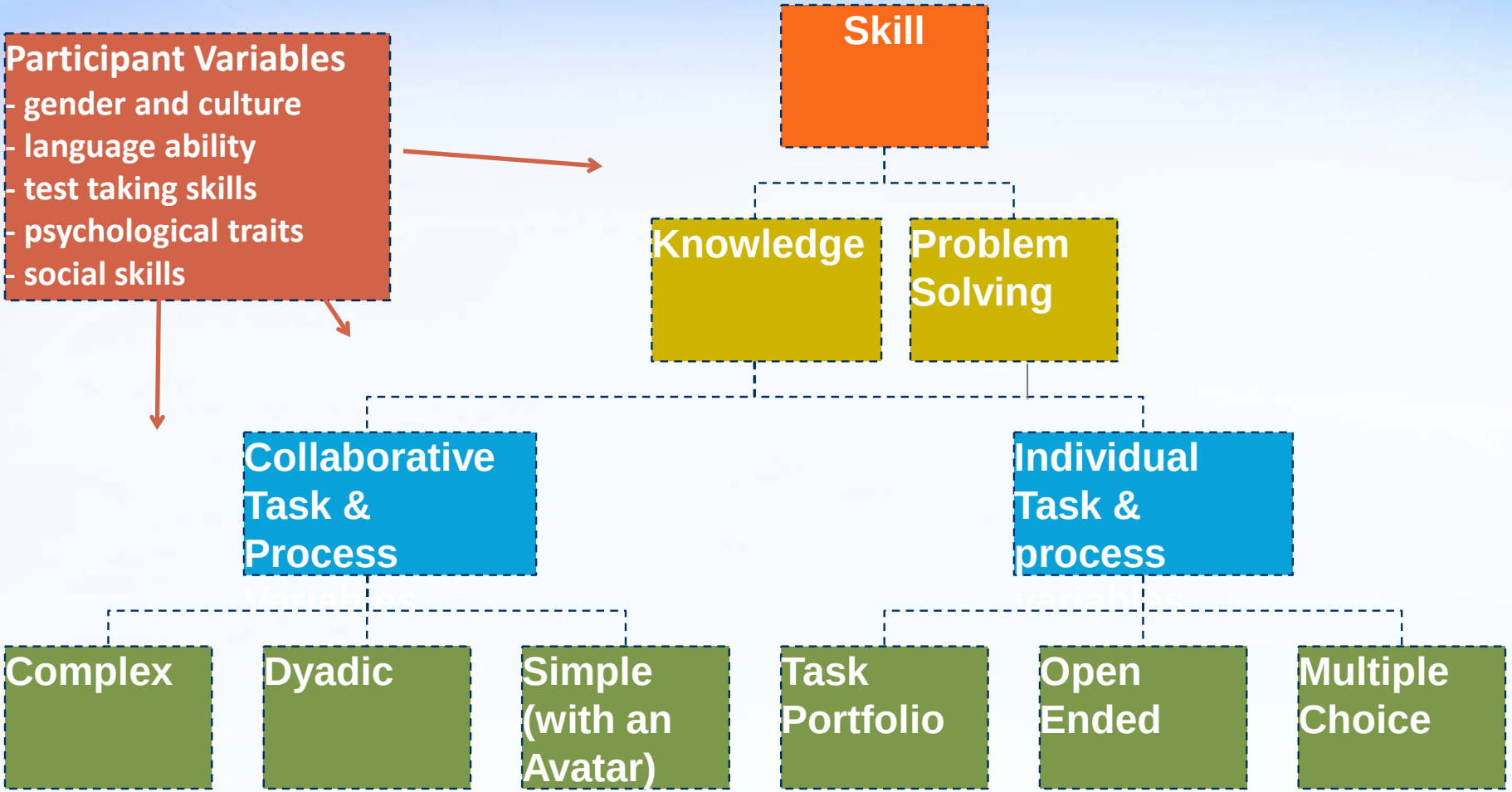
QUALITY MEASUREMENT OF THE COLLABORATIVE PROBLEM SOLVING CONSTRUCT. HOW TO DO THAT?

- Interdisciplinary teams
- Evidence Centered Design to design the simulations/games/collaborative tasks
- Large samples
- Representative samples
- Robust measurement models
- (Predictive) validity studies
- Fairness studies
- Process data/modeling the interactions



TEST DESIGN

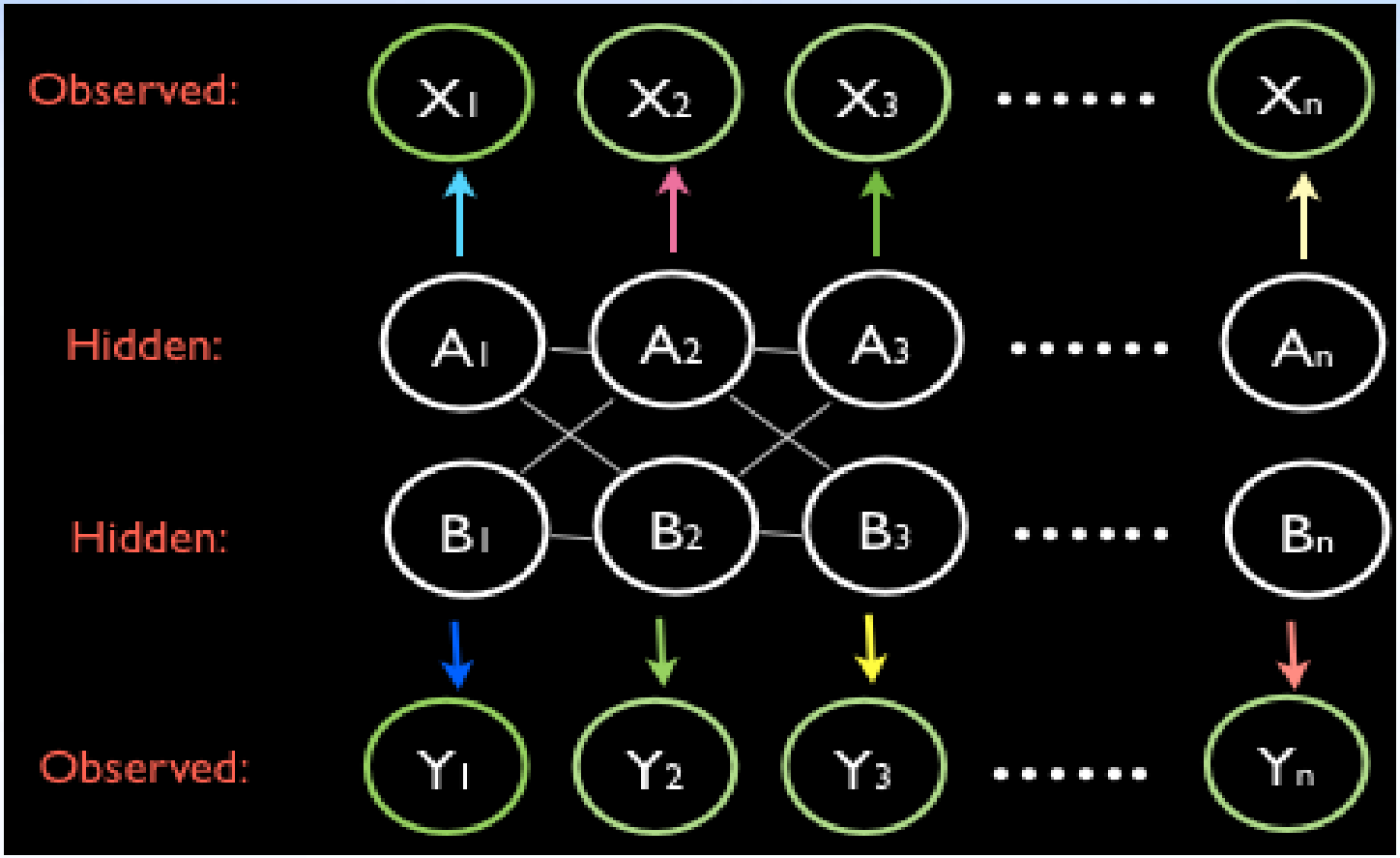
AN ASSESSMENT WITH CPS TASKS



MODELING THE INTERACTIONS

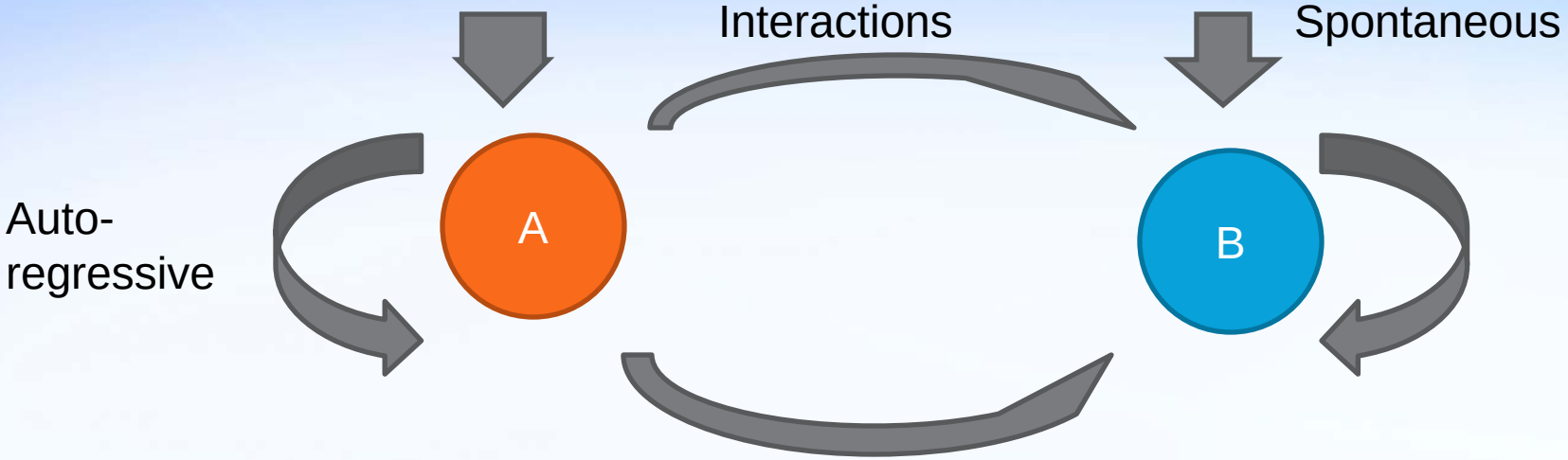
Listening. Learning. Leading.®

Hao and von Davier (2013)



4/2/2015

MODELING THE INTERACTIONS



Blundell, Heller, & Beck (2012)
Halpin & DeBoeck (2013)
von Davier & Halpin (2013)

PSYCHOMETRIC CONSIDERATIONS: MULTIDIMENSIONALITY & DEPENDENCE

- **Multidimensionality of skills and team members**
  - Multidimensional skills
  - Multiple time series (people)
- **Dependence**
- **Local independence assumption from IRT models doesn't hold**
  - Over items (task & time dependence)
  - Over people (within group interdependence & across groups dependence & learning*)

- **Local Independence Assumption:**

For all items (i=1,…I) & test takers (j=1,…J) and a vector of latent skills θ

$$p(U = u_{ij} \mid \theta_j) = \prod_{j=1}^{n} \prod_{i=1}^{I} p(u_{ij} \mid \theta_j)$$

DIFFERENT TYPES OF DEPENDENCE MAY LEAD TO DIFFERENT MODELS

- **Static Models that account for test takers dependence** (team/group dependence)
  - Multilevel modeling
  - Social Networks Analyses
  - Nonparametric exploratory models (hierarchical segmentation and small-world networks)
  - Latent class analysis/Neural networks

- **Static Models that account for items-within-task dependence**

  Bayes nets

  MIRT, Testlet Model, Bi-Factor Models

MODELING STRATEGIES FOR THE PROCESS DATA

Models that account for multidimensionality, item dependence over time, and test-takers interactions over

Continuous Process

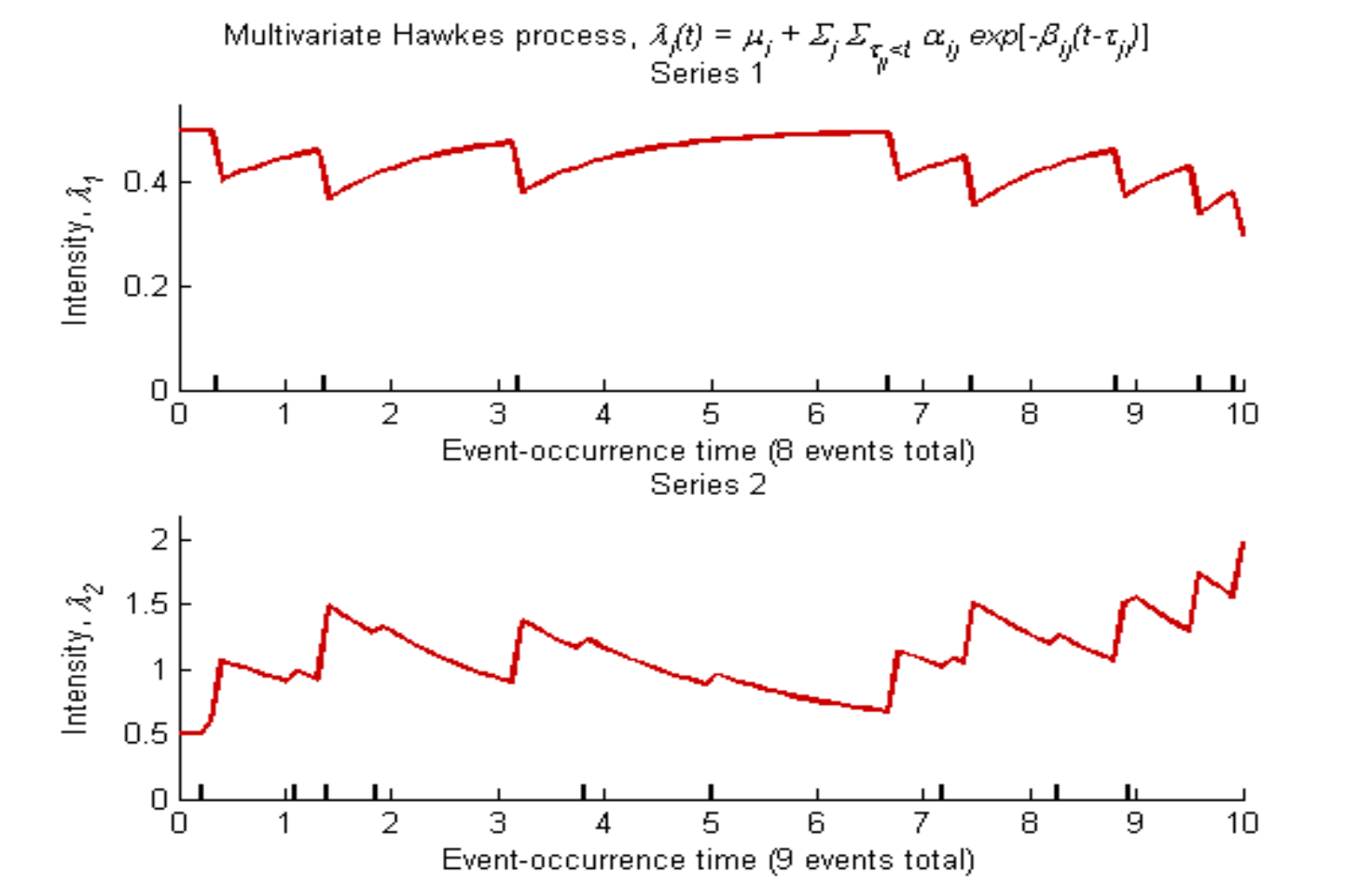
- Dynamic factor analysis
- Dynamic (linear) models
- Differential equation models
- Intra-variability models
- Machine learning methods*

Discrete Process

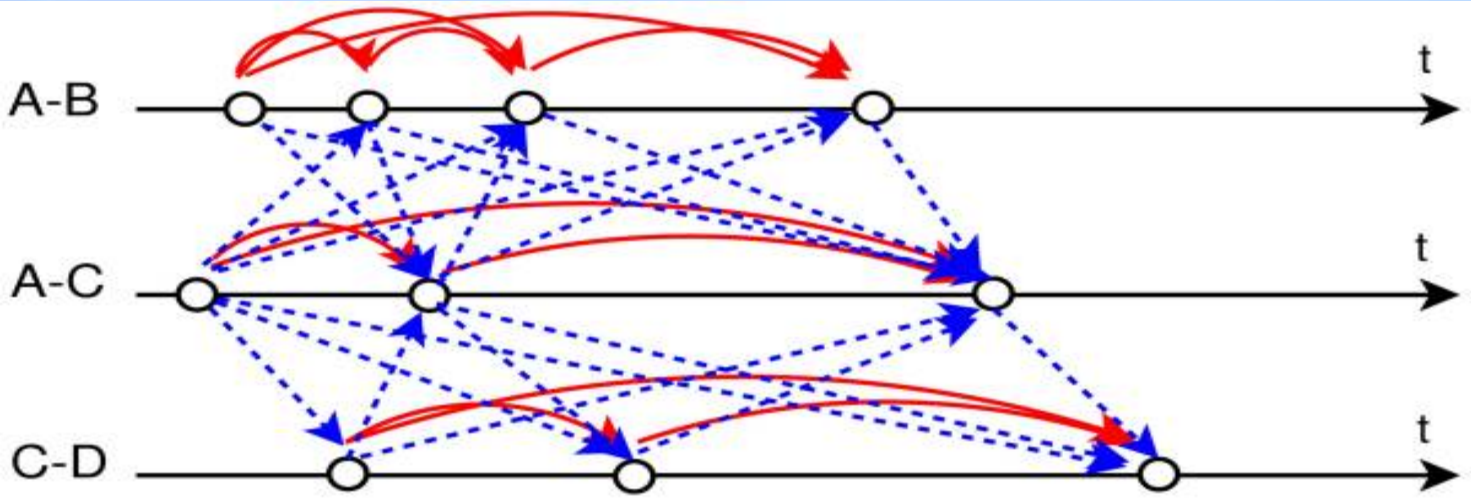
- Intra-variability models
- Point processes (stochastic processes for discrete events)
- Hidden Markov models/ Markov Decision Process/Dynamic Bayes nets*
- Develop New Models??

EXAMPLE OF SIMULATED A BIVARIATE HAWKES PROCESS

HTTP://WWW.MATHWORKS.COM/MATLABCENTRAL/FILEEXCHANGE/17983-SIMULATE-A-HAWKES-PROCESS/CONTENT/HAWKES/HTML/HAWKESDEMO.HTML



OVERLAPPING TEAMS



A, B, C are test takers assigned to different teams (A-B, A-C, C-D). Mixture Hawkes Models (Li & Zha, 2014)

INDICES THAT CAPITALIZE ON INTERDEPENDENCE ESTIMATION

- **Indices that depend on the covariance matrix**

- Mutual information function
- Entropy
- Kullback-Leibler divergence
- Conditional Information Function

- **Estimation issues**

- Identifiability (parameters cannot be distinguished/uniquely estimated)
- Ill-conditioning of likelihood (due to co-linearity/rank reduction in the covariance matrix/eigenvalues close to 0, functional form, etc)
- Quadratic growth of algorithm

ADAPTIVE TESTING & COLLABORATION

- **CPS as an adaptive test**

- The interactions of team members influence the next outcome
- Learning during CPS
- Adaptive features in the CPS

- **Adaptive/MST assignment to teams**

- Use a MC test scores to assign people to teams (attempted in the Tetralogue)

- **Adaptive testing & team assembly**

- Similar algorithms (numerical optimization and heuristics) can be used to develop tests & teams

MULTIMODAL AFFECT DETECTION IN COLLABORATION

- MMD represent a different type of data that can be used to augment the traditional logfile data

- MMD are multivariate process data

- Continuous (eye gaze)
- Discrete (position, time of interactions)

- Useful for measurement
- Useful for validity

THANK YOU!

Contact information

A. A. von Davier: avondavier@ets.org

ADDITIONAL SLIDES
A MORE DETAILED DESCRIPTION OF MODELS

NONPARAMETRIC EXPLORATORY APPROACHES

The use of hierarchical segmentation, a computational modeling technique, to detect periods of time in which multiple time series show similar patterns of variability.

The use of small-world networks to visualize and describe transitions in the variable of interest in specific time intervals, say, daily

Exploratory methods can be useful when theoretical information is not available to specify a model as well as when important assumptions about the data, variables, or models (e.g., Gaussian distributions, stationarity) are not met.

Exploratory methods can provide descriptive information that can be used to formulate hypotheses and to specify other confirmatory models

MULTILEVEL MODELING

Multilevel models take into account clustering in the data such team members in a team or as repeated observations within individuals.

Multilevel models are appropriate for the examination of changes in means over time as well as influences from covariates (both time-varying and invariant).

They are less appropriate for the modeling of short-term fluctuations, the factorial structure of multivariate data, or the temporal organization of dyadic interrelations, as well as in the case of small samples or single-case studies.

BAYESIAN INFERENCE NETWORKS

It starts by characterizing aspects of students' knowledge and skill in terms of a possibly vector-valued student model variable, θ, and aspects of their behavior in terms of possibly vector-valued observable variables, X.

Conditional probability distributions P(X|θ), obtained through theory, expert opinion, empirical data, or some combination of these, characterize how performance depends on knowledge and skill in task situations.

Letting the prior probability distribution P(θ) denote the assessor's belief about a student's at a given point in time, observing X leads to an updated posterior probability distribution P(θ|X) by Bayes' theorem.

AN ANALYSIS OF INTRA-INDIVIDUAL VARIABILITY

Ferrer, et al. (2012) considered variability at the individual level and then examine the extent to which the patterns of variability for both individuals overlap in time. They focused on relative constancy in emotional relationships in married couples.

The data (e.g., daily time series of positive and negative affect for couples) can be arranged such that, on any given day, an individual's overall affect state can be described using Cartesian coordinates.

This representation is informative but when repeated over time, the individual's affective positions become a signal indicative of emotional fluctuations and daily transitions.

This signal can then be subjected to algorithmic aggregations and segmentations providing the identification of periods within the signal that remain stable and periods that reveal salient transitions.

DYNAMIC FACTOR ANALYSIS

- Dynamic factor analysis (DFA) is appropriate for modeling dyadic interactions, especially when multivariate time series are available (Molenaar, 1985; Ferrer & Nesselroade, 2003).
- DFA combines factor analysis with time series such that it accounts for the autocorrelations among the multiple variables over time (Browne & Nesselroade, 2005; Ferrer & Zhang, 2009).
- DFA focuses on time-related influences between the two individuals, where one could identify possible influences from one person's action at a given time to the other person's action at the next occasion. These techniques are particularly useful when the data show fluctuations, without trends or other forms of nonstationarity.

DIFFERENTIAL EQUATION MODELS

DEM have been used to model interactions in dyads (Felmlee, 2006) such as turn-taking in conversations (Buder, 1991; Newtson, 1993) and the development of various types of social relationships.

These models are appropriate for data that are continuous and with many measurement occasions.

MARKOV MODELS

A Markov chain can describe the way that samples taken at consecutive time intervals follow a representative path.

A *mixed Markov model is a mixture* of a finite number of Markov chains. Each sample must be a member of one of the prescribed paths.

Mixed Markov models require that each sample be a member of one of the prescribed paths described by the transition probabilities. This may be overrestrictive in longitudinal data analyses because individuals may change over time, and such models do not allow individuals to move between latent classes over time (Vermunt, 2007).

The *hidden Markov model (HMM)* allows latent transitions. The response probabilities and transition probabilities for the Markov chains in the model can change over time.

Hence we can model the way the rate of change for each subject also changes over time as different temporal variables affect the way in which subjects respond to stimuli.

HMM INSTANTIATIONS IN CPS (E.G., SOLLER & STEVENS, 2007)

Observations might be classifications of different student problem-solving strategies with state transitions describing the likelihoods of transitioning from one general problem-solving strategy to another (e.g., on the next problem set or during the next term).

In a collaborative distance learning environment, observations might be sequences of online chat between students, and state transitions might describe the communicative roles of students (e.g., facilitator, critic, peer tutor) or the effectiveness of the information sharing and knowledge construction.

THE HAWKES PROCESS & CPS

von Davier & Halpin (2013)

Blundell, Heller, & Beck (2012)

Halpin & De Boeck (2013)

Models for the timing of discrete events, readily extended to multivariate scenario.

- **e.g., Multiple team members.**

Recent advances allow for scalable EM estimation.

- **Halpin & De Boeck (2013); Halpin (2014);; see also Veen & Schoenberg (2008)**

Bayesian approaches also available; Rasmussen (2012).

POINT PROCESS MODELS

Point process models are useful for describing phenomena occurring at random locations and/or times.

Some important models are

Poisson processes, renewal processes, Hawkes processes, and Markovian point processes.

The relationship between time series and point processes is worth noting. Many datasets that are traditionally viewed as realizations of (marked) point processes could in principle also be regarded as time series, and vice versa.

SCIENCE ASSESSMENT
WHAT IS IT?

An assessment of science & collaborative skills

- **Based on collaborative features identified from the PISA framework & ICLS literature (Liu, von Davier, Hao, Kyllonen, Zapata-Rivera, 2014)**

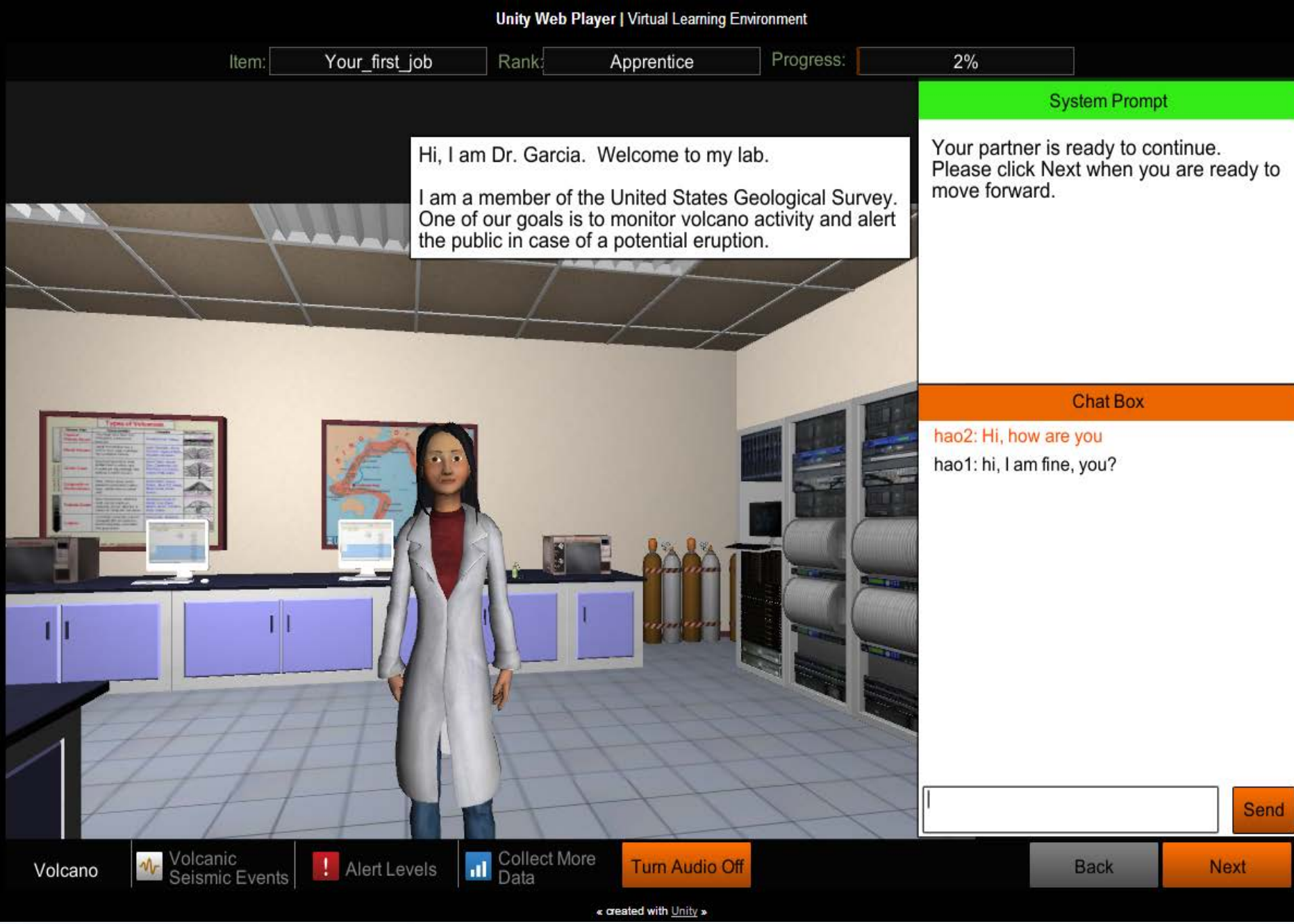
A traditional science test (general science knowledge)

A collaborative task in a game environment

- **Scripted avatar teacher teaching volcano lesson, scripted student colleague learning along with you**
- **Chat window with live student colleague**

Covariates

- **Personality, other background information**



VIDEO EXAMPLE

Collaborative Science Assessment—Humans & Avatars

http://ets-research.org/jhao/video/tetralogueVideoAudio.mp4

COLLABORATION LEADS TO BETTER PERFORMANCE

(HAO, LIU, VON DAVIER, 2014)