Future of methods and measures in the field of education research

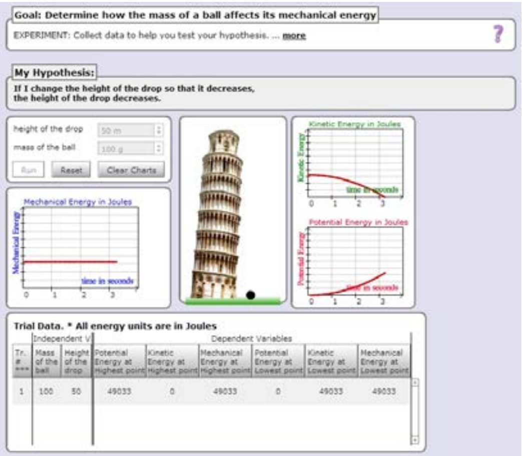
Ryan S. Baker

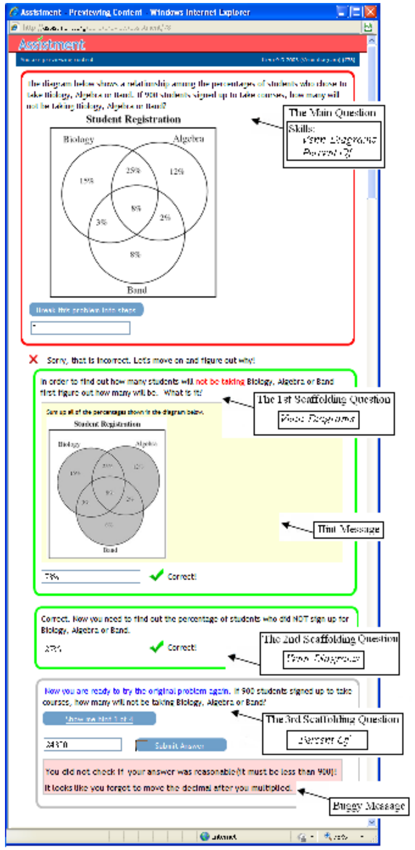
University of Pennsylvania

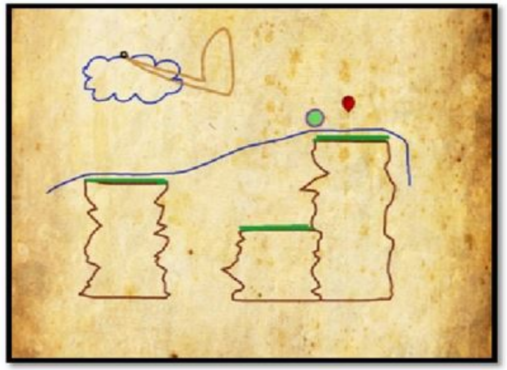
Increasing availability of process/ interaction data

Increasing availability of process/ interaction data

- Log data from intelligent tutoring, games, simulations, homework platforms
  - Already being used at scale prior to pandemic; increased significantly in 2020-2021

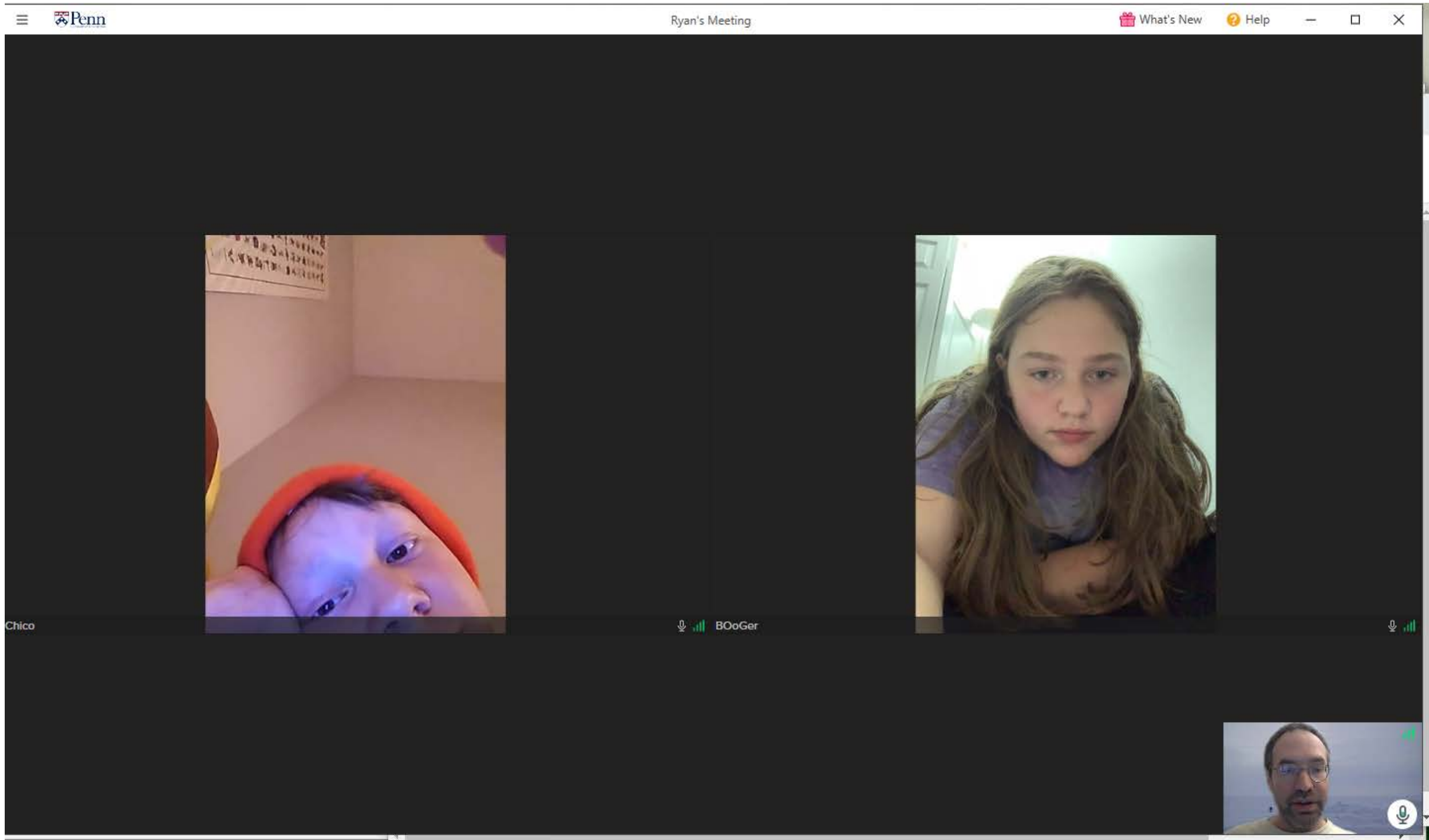






Increasing availability of process/ interaction data

- Video data/transcripts from Zoom hybrid class and tutoring sessions
  - Field is in the middle of a pivot to work more with this kind of data



Increasing availability of process/interaction data

- Classroom analytics from physical sensors of various types
  - Promising initial work, privacy and access concerns
  - Can integrate teacher or classroom observer data

From Dillenbourg&Jermann, 2010

Large-scale data

- Log data from systems used by hundreds of thousands or millions of students per year
  - ALEKS, MATHia, edX, Coursera, Inq-ITS, Duolingo, ASSISTments, Zearn, and many many more...
  - Available in many cases to researchers for measurement development

How can we use this data?

- Evidence Centered Design/Knowledge Engineering/Psychometric Paradigms
  - Intensive human effort to create and validate measurements that can be agreed upon
  - Produce highly interpretable and defensible measurements

- Educational Data Mining/Learning Analytics
  - Algorithms used to create measurements; still a considerable amount of human effort in setup and validation
  - More feasible for measuring constructs that are hard to explain or articulate, or where a construct can manifest in several distinct ways

Educational Data Mining/Learning Analytics

- Classic Machine Learning
  - Relatively interpretable algorithms
  - Relatively straightforward behavior
- Deep Learning
  - Generally more accurate prediction/inference
  - Relatively inscrutable algorithms
  - Relatively high frequency of unexpected behavior (Yeung & Yeung, 2018; Lee et al., 2021)

“Supervised” Learning

- Algorithm needs a set of examples – “training labels” to learn a model that can be used with new data
- The training labels can come from
  - Traditional tests
  - Classroom observations
  - Video data coding
  - Text replays
  - Self-report
  - Other outcome data

Stuff We Can Infer: Learning

- Has the student learned the current skill? (Corbett & Anderson, 1995; Pavlik, Cen, & Koedinger, 2009; Khajah et al., 2016; Pelanek, 2016; Wilson et al., 2016; Ekanadham & Karklin, 2017; Choffin et al., 2019; Pandey & Karypis, 2019; Scruggs et al., 2020)

- Where in the learning sequence is the student? (Desmarais & Pu, 2006; Adjei, Botelho, & Heffernan, 2016; Han et al., 2017)

- Is the student wheel-spinning: making no or minimal progress? (Beck & Gong, 2013; Matsuda et al., 2017; Botelho et al., 2019; Corbeil et al., 2020; Wang et al., 2020)

Stuff We Can Infer: Complex Learning

- Is the student learning to solve complex problems that require inquiry? (Sao Pedro et al., 2013; Baker & Clarke-Midura, 2013; Perez et al., 2017; Teig et al., 2020)
- Is the student developing rich conceptual understanding in domains such as physics and computational thinking? (Shute & Ventura, 2013; Rowe et al., 2015, 2019)

Stuff We Can Infer: Robust Learning, Meta-Cognition, Self-Regulation

- Will the student remember what they learned? (Jastrzembski et al., 2006; Pavlik et al., 2008; Wang & Beck, 2012; Choffin et al., 2019; Matayoshi et al., 2020)
- How confident is the student? (Litman et al., 2006; McQuiggan, Mott, & Lester, 2008; Arroyo et al., 2009)
- Is the student asking for help when they need it? (Aleven et al., 2004, 2006; Almeda et al., 2017; Price, 2018)
- Is the student persisting in the face of challenge? (Ventura et al., 2012; Erickson et al. 2018; Kai et al., 2018)

Stuff We Can Infer: Disengaged Behaviors

- Gaming the System (Baker et al., 2004, 2008, 2010; Walonoski & Heffernan, 2006; Beal, Qu, & Lee, 2007; Paquette et al., 2019; Mogessie et al., 2020)
- Carelessness (San Pedro et al., 2011; Hershkovitz et al., 2011)

Stuff We Can Infer: Affect

- Boredom
- Frustration
- Confusion
- Engaged Concentration/Flow
- Curiosity
- Excitement
- Situational Interest
- Joy/Delight

- (D’Mello et al., 2008; Mavrikis, 2008; Arroyo et al., 2009; Conati & Maclaren, 2009; Lee et al., 2011; Sabourin et al., 2011; Baker et al., 2012, 2014; Paquette et al., 2014, 2015; Pardos et al., 2014; Kai et al., 2015 ; Hutt et al., 2019)

No physical sensors needed

- Often feasible to infer these constructs solely from student interaction with the learning system
- Although using physical or physiological sensors, where feasible, increases model quality
(Kai et al., 2015; Bosch et al., 2015; Henderson et al., 2020, 2021)

Algorithmic Bias

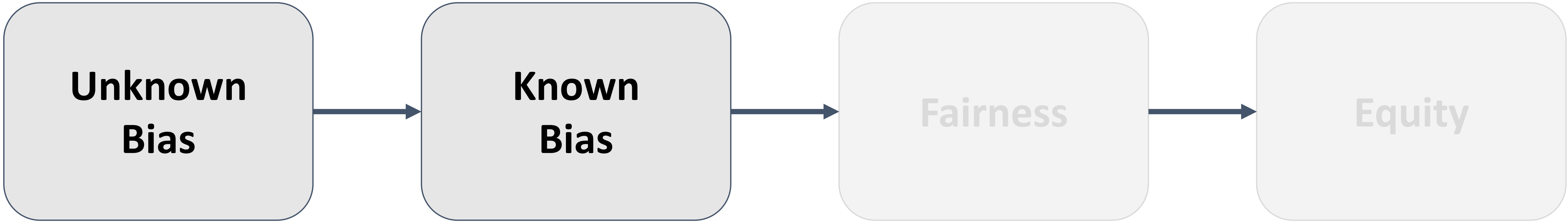
- Biased computer systems “*systematically* and *unfairly discriminate* against individuals or groups of individuals in favor of others.”

  (Friedman & Nissenbaum, 1996)

- Cases where an algorithm’s performance is substantially better or worse across different groups of learners

  (Baker & Hawn, 2021)

Path towards resolving algorithmic bias for these measurements



Convenience Sampling: A Common Problem

- Model performs less well for group(s) less represented in data used to develop model
  - Suburban middle-class students ≠ Urban lower-income students
  - Even a “complete” data set may not be enough if a group is rarely seen in the data set

What do we know about bias impacting learners in common demographic categories?

- Most research on algorithms in education does not even mention learner demographics, much less investigate impacts for learners in different demographics (Paquette et al., 2020)

Relatively well-documented algorithmic biases in education (review in Baker & Hawn, 2021)

- Race and Ethnicity (Anderson et al., 2019; Hu & Rangwala, 2020; Lee & Kizilcec, 2020; Yu et al., 2020)
- National Origin (Bridgeman et al., 2009, 2012; Ogan et al., 2015)
- Gender (Kai et al., 2017; Anderson et al., 2019; Christie et al., 2019; Gardner et al., 2019; Hu & Rangwala, 2020; Lee & Kizilcec, 2020; Riazy et al., 2020; Yu et al., 2020)

What do we know about bias impacting other groups? (review in Baker & Hawn, 2021)

- Insufficient research -- there needs to be more
- The field doesn’t even know about all the groups that are impacted
    - One study on second-language learners (Naismith et al., 2018)
    - Two studies on learners with disabilities (Loukina et al., 2018; Riazy et al., 2020)
    - Two studies on urban/rural/suburban differences (Ocumpaugh et al., 2014; Samei et al., 2015)
    - Two studies on parental educational background (Kai et al., 2017; Yu et al., 2020)
    - Two studies on socioeconomic status (Yudelson et al., 2014; Yu et al., 2020)
    - One study on children in military families (Baker et al., 2020)

    - **Many differences and groups have not been studied at all**

Not limited solely to machine learning

- Occasionally, people believe that algorithmic bias can be avoided by simply avoiding machine learning and using simple rubrics developed by hand

- A lot of the worst cases of algorithmic bias involve rubrics developed by hand

- British Algorithmic Grading Scandal of 2020

Not limited solely to machine learning

- Example in predictive analytics in education

- “Chicago model” (Allensworth & Easton, 2007) is a very straightfoward set of indicators for predicting dropout, developed by hand
- It predicts well in general (Bowers, Sprott, & Taff, 2013)
- But not nearly as well as modern machine learning approaches, when applied to diverse sample of districts (Coleman et al., 2019)
- Performance of Chicago Model is very uneven across demographic groups (Coleman, 2021)

The good news

- Increasing tools and metrics for assessing and reducing algorithmic bias, coming both from EDM/LA and Fair AI communities (see review in Kizilcec & Lee, 2020)

- Increasing awareness and concern about these issues in the community
  - Group working session as plenary last week at EDM2021

- There are methods for addressing algorithmic bias, key challenges are
  - Availability of demographic variables (privacy concerns)
  - Incentives for researchers and developers to do the extra work

What would help

- Policy that better balances privacy with equity – right now the scales have tipped heavily towards privacy
  - Research infrastructures can enable work to fix algorithmic bias while keeping demographic data non-viewable
- Funders explicitly requiring attention to algorithmic bias and equity in research
- School districts and education agencies and clearinghouses requiring evidence around algorithmic bias and equity

Thank you!

Penn Center for Learning Analytics
University of Pennsylvania

twitter.com/BakerEDMLab

Baker EDM Lab

"Big Data and Education", running on edX now
All published papers available online – Google "Ryan Baker"

Algorithmic Bias review -- Go to Google Scholar, type in "Baker algorithmic bias in education"