It Takes a Village: Future Directions for Statistics Education Research

USCOTS Plenary Session
Saturday, May 21, 2011
1:45-2:45

Bob delMas
Department of
Educational Psychology

University of Minnesota

Overview

- How old is statistics education?
- How old is statistics education research?
- What are we researching?
- What do we know?
- Where should we go next?
- How should we get there?

How old is statistics education?

- Difficult to fix a date, but a relatively new field.
- 1949: Committee on Statistical Education formed by the International Statistical Institute (ISI).
  - Focus on the university training of statisticians
  - Gradually broadened its scope to all levels of statistical education
- 1982: First International Conferences on Teaching Statistics (ICOTS) held by the ISI.
- 1993: International Association for Statistical Education (IASE) formed.
  - IASE became the organizers of ICOTS.
- Recognition of statistics education as a discipline has come about in the last three decades.

How old is statistics education research?

- Again, difficult to fix a date
- Research studies on children's and adult's understanding of probability have been conducted for at least 50 years.
  - Lindman, H. & Edwards, W. (1961). Supplementary report: Unlearning the gambler's fallacy. *Journal of Experimental Psychology, 62*, 630.
  - Tversky, A. & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 2*, 105-110.
  - Fischbein, E. (1975). *The intuitive Sources of Probabilistic Thinking in Children*. Dordrecht, Holland: D. Reidell Publishing.

How old is statistics education research?

- 1980s: Articles on the teaching of statistics emerged
  - Kempthorne, O. (1980). The teaching of statistics: Content versus form. *The American Statistician, 34*, 17-21.
  - Holmes, P. & Turner, D. (1981). Teaching statistics to eleven-to-sixteen-year-olds. In A. Shulte (ed.), *Teaching Statistics and Probability*. 1981 Yearbook, NCTM.
  - Nisbett, R, Krantz, D., Jepson, C., & Kunda, Z. (1983). The Use of Statistical Heuristics in Everyday Inductive Reasoning. *Psychological Review, 90* (4), 339-363.
- 1993: Journal of Statistics Education (JSE)
- 2002: Statistics Education Research Journal (SERJ)

Kempthorne on Teaching Statistics

Aims of the discipline; statistics is not mathematics; the foundations are not in mathematics; mathematics should be the servant of statistics and not the master; … content should dominate form; the ultimate content must be philosophical; … in the beginning was not the "word" but a problem; data analysis without a problem is pure waste; the flooding of humanity with data and data analyses; association or correlation versus causation; mere observation can be totally misleading; …the dangers of rotten statistics and examples thereof.

What are we researching?

- SERJ Analysis: Andy Zieffler and Statistics Education Research Group at the University of Minnesota
  (Joan Garfield, Bob delMas, Audbjorg Bjornsdottir, Rebekah Isaak, Laura Le, Jiyoon Park)
- Detailed analysis of articles published in SERJ
  - 2002-2009
  - 64 Articles
- Methods: Content and Text analysis
  - Who is publishing research in SERJ?
  - What is being published and why?
  - How is the research being carried out?

Who is publishing?

- 7 different types of Departments
  - Education: First authors (53%); All authors (34%)
  - Statistics: First authors (17%); All authors (16%)
- 66% of articles have multiple authors
  - Increase from about 40% in 2003 to about 80% in 2009
  - 45% cross-departmental collaborations (47% with Education)
  - 19% international collaborations (primarily USA and Australia)

What is being researched?

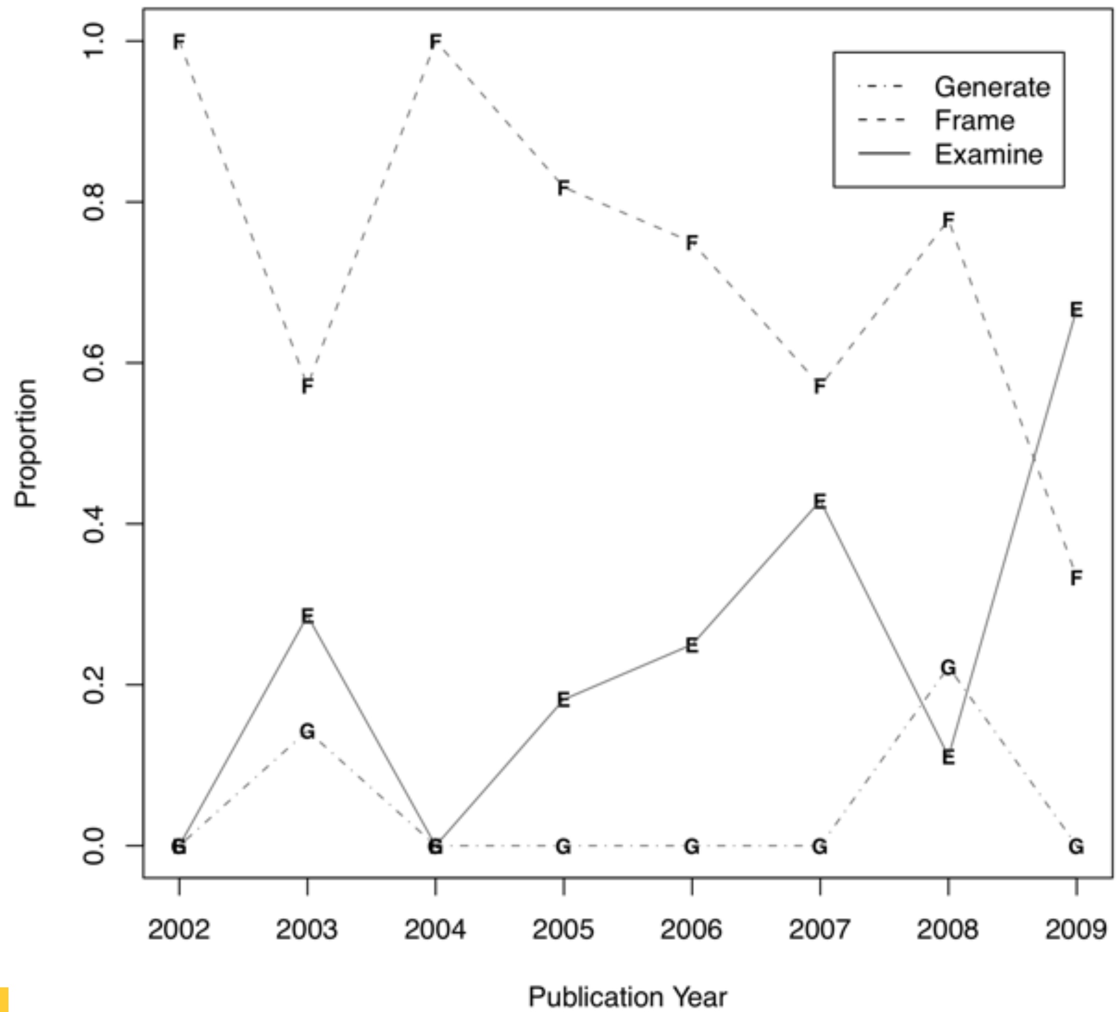
- Used articles that stated research question(s).
- **Reasoning/Understanding (28):** General or about a specific concept or topic.
- **Teaching and learning(8):** Questions and goals related to teaching methods and students' learning.
- **Affect (6):** Examination of students´ attitudes, anxiety, or motivation while studying statistics.
- **Technology (6):** Questions/goals related to the use of technology in the teaching of statistics

How is research being conducted?

Classification Framework: *Using Statistics Effectively in Mathematics Education Research* (SMER; ASA, 2007)

- ***Generate:*** Ideas about the phenomena of interest.
- ***Frame:*** Clarification of research goals, definitions of constructs, development of construct measurement.
- ***Examine:*** Systematic examination of a phenomena.
- ***Generalize:*** Generalize what has been found addressing questions of scale, refining theory, or reframing research.
- ***Extend:*** Generalizable outcomes extended in a variety of ways—synthesizing multiple studies, examining long-term effects, developing policies for effective implementation.

How is research being conducted?



Summary: What are we researching?

Some Strengths

- Statistics Education research is collaborative and interdisciplinary.
- Large diversity of disciplines, theoretical perspectives, research approaches and methods.
- Diversity in types of studies ranging from
  - Purely theoretical
  - Purely qualitative
  - Purely quantitative

Summary: What are we researching?

Some Weaknesses

- About one-third of research articles did not state a research question.
- Quantitative studies rarely cited results from qualitative studies.
- Need more research utilizing mixed-methods.
- Need more foundational studies to generate questions.
- Need more development of measures and instruments.

What do we know?

Garfield, J. & Ben-Zvi, D. (2007). How Students Learn Statistics Revisited: A Current Review of Research on Teaching and Learning Statistics. *International Statistical Review, 75*, 3, 372–396.

- Identified 15 areas or topics of research
- Summaries of research in each area
- Implications for instruction
- Implications for developing students' reasoning about distribution, center and reasoning

Studies of errors and misconceptions

- Misconceptions are widespread and persistent
- Similar at all age levels and difficult to change.
- Statistical reasoning is often inconsistent from item to item or topic to topic.
- Statistical training can lead to positive results, but no strong evidence that:
  - Results were sustained beyond the training sessions
  - Generalized beyond the specific types of problems used.

How do K-12 students learn statistics?

- Students find reasoning about data and chance difficult and complex.
- Developing understanding of distribution, center, variability, etc., can be done with:
  - Carefully designed sequences of activities
  - Using appropriate technological tools
  - Over substantial periods of time, revisiting ideas.
- Start with informal, intuitive ideas and move toward more formal understanding.
- Cover ideas of distribution, center and spread simultaneously, rather than as separate, isolated topics.

K-12 teacher knowledge of statistics

- Demonstrate difficulties in understanding and teaching core ideas of probability and statistics.
  - Limited or incorrect understanding of sample, distribution.
  - Mathematics teachers tend to teach computational methods.
- Carefully planned instruction with appropriate technology can improve both knowledge and teaching.
- Greater knowledge of statistics is associated with better teaching and learning outcomes for students.
- Need more research on effective ways to develop teacher knowledge of statistics, teaching statistics, and technology.

How do college students learn statistics?

- K-12 recommendations apply to developing statistical reasoning at the college level.
- "High ability" students can have relatively unsophisticated understanding of important ideas at end of a course.
- Cooperative learning approaches result in higher course performance.
- Focusing on structural features of context and providing meaningful, immediate feedback can improve students selection of appropriate procedures.
- Attitudes and anxiety not strong predictors of learning; Little change in attitudes from beginning to end of course.

Critique of college-level research

- Note the lack of high quality and consistent measures to assess student learning outcomes.
  - Lack evidence of validity and reliability
- Results of comparative studies often not generalizable.
  - Definitions: "traditional course" versus "active learning".
  - Localization of results (this institution; these students).
- Need for clear definitions of:
  - Research questions
  - Learning outcomes
  - Instructional methods

Where should we go next?

- Studies of teacher preparation, teacher knowledge of statistics, teacher knowledge for teaching.
- Studies that explore promising pedagogical approaches such as cooperative learning.
- Studies that extend promising approaches to broader contexts and populations of students.
- Studies of whether teaching informal ideas of statistics promotes better understanding of formal ideas.
- Studies of alternative approaches to developing statistical reasoning and thinking.

How should we get there?

It takes a village:

- Form interdisciplinary collaborations
- Develop statistics education researchers
- Develop graduate programs in statistics education
  - Coursework in learning theory and educational research
  - Greater understanding of mixed-methods approaches.
  - Coursework in Quantitative Methods in Educ. Research
  - Training in qualitative research and analysis (to generate questions).
  - Develop knowledge of modern measurement theory and methods to develop measures and instruments.

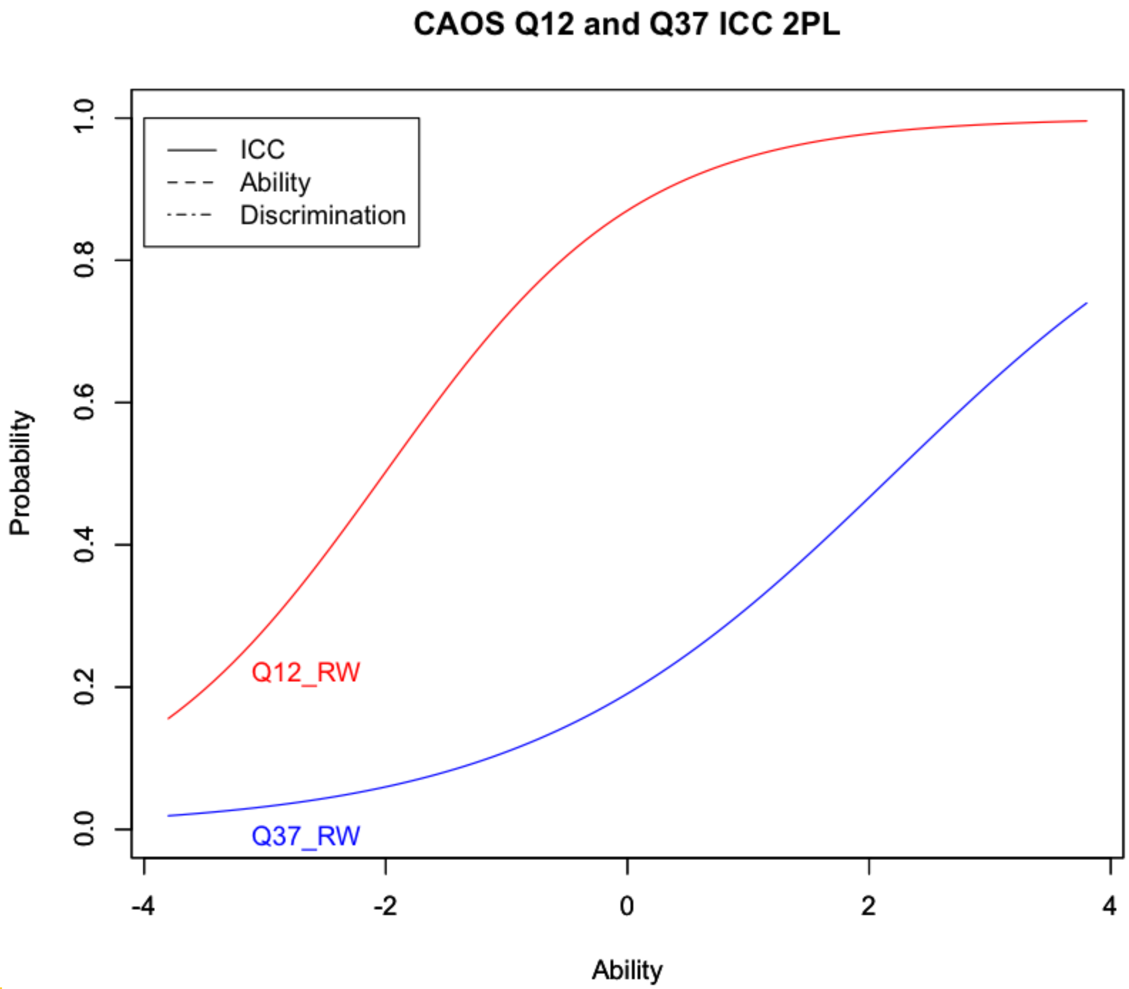
IRT: Item Response Theory

- Modern psychometric method for test development
- Framework for evaluating how well
  - The assessments works
  - The individual items work (or function)
- Focus is on the item (instead of the entire test)
- Probability of a correct response is a function of
  - Person's (latent) ability
  - Item parameters
    - Difficulty (location)
    - Discrimination (slope or correlation)

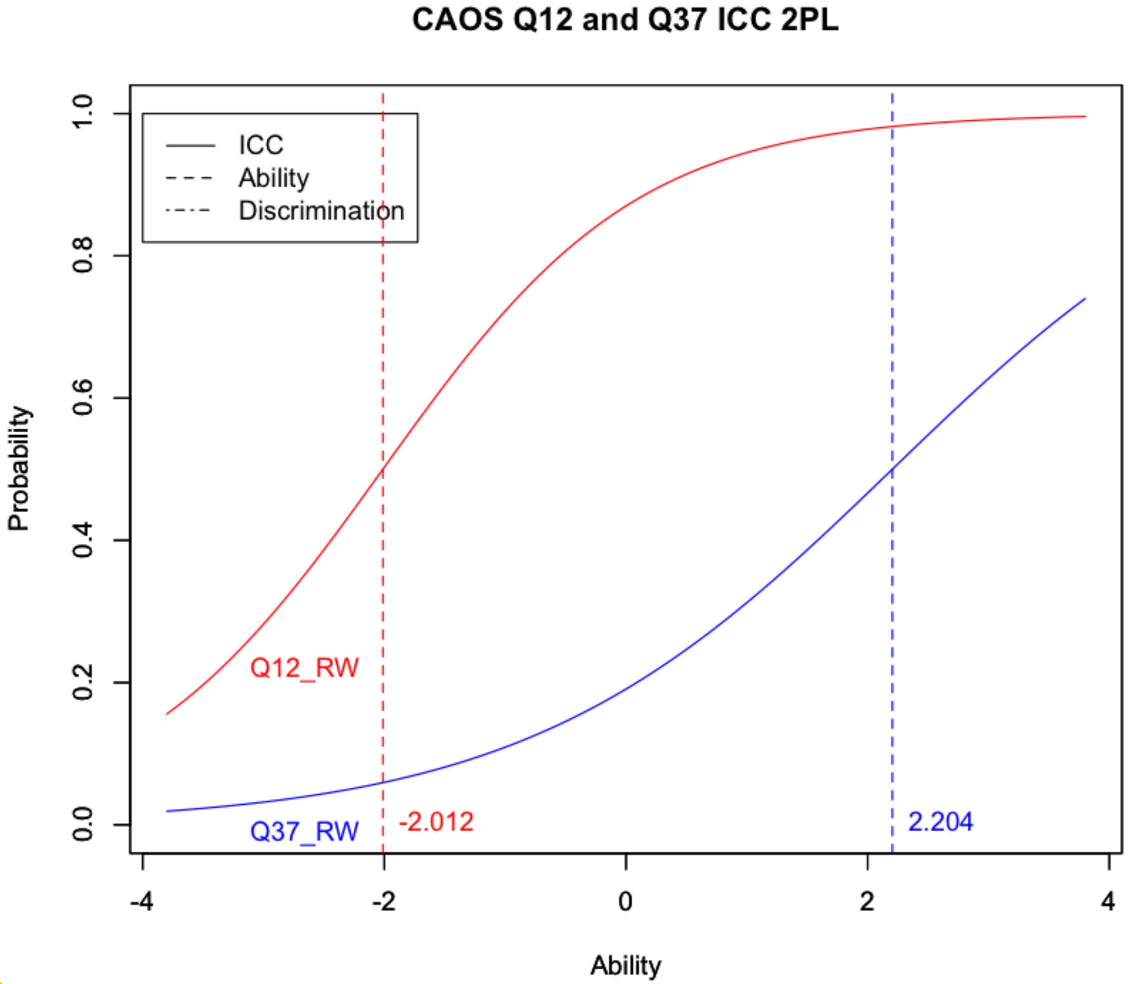
IRT Analysis of CAOS Test

- Leah McGuire – measurement expert (UC Berkeley - BEAR Center)
- Working with three U of MN Stat Ed faculty and students (Bob delMas, Jiyoon Park, Laura Ziegler)
- Part of StatWay project (Carnegie Foundation and Dana Center, UT Austin)
- N = 11,726 undergraduates

Item Characteristic Curves (ICC)

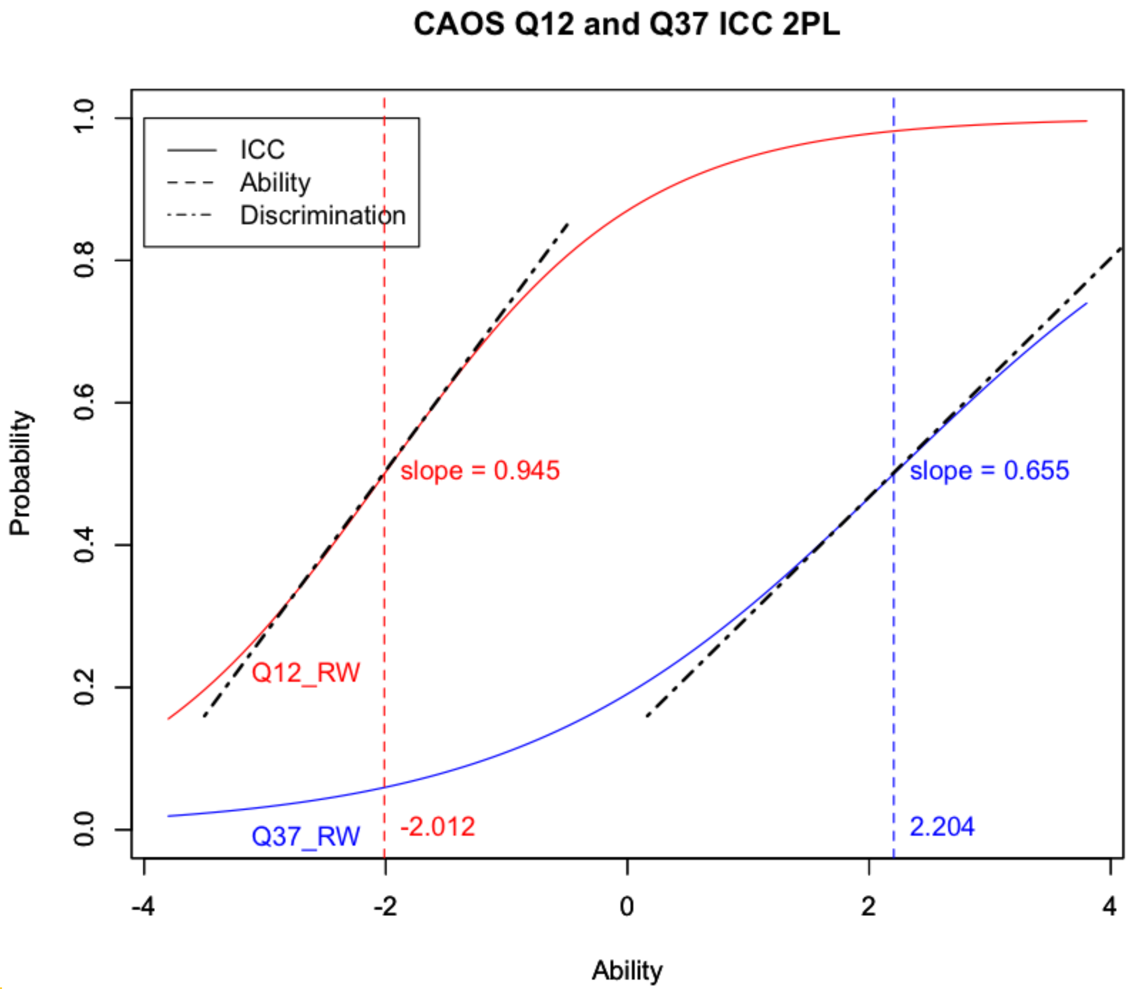


Item Difficulty (Student Ability Level)



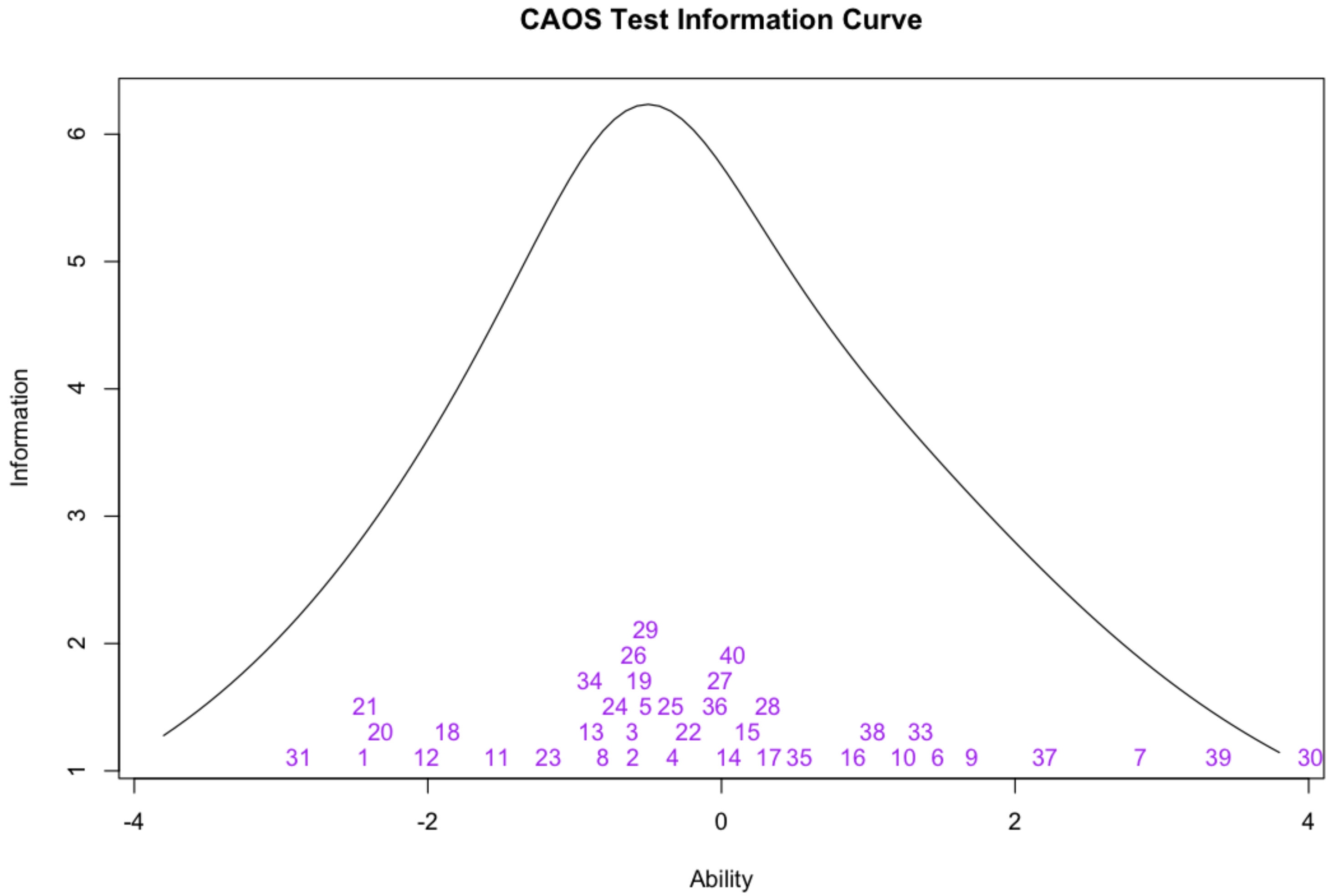
University of Minnesota

Item Discrimination



University of Minnesota

IRT Ability Levels and Item Difficulty



University of Minnesota

Example of Collaborative Research

- The nature of the introductory statistics course has been changing.
- Reflected in GAISE guidelines:
  - Less emphasis on computation and procedures
  - More emphasis on statistical thinking and reasoning
- Field of statistics has been changing:
- George Cobb (USCOTS 2005; TISE, 2007): Promotes teaching permutation-based methods
- Several projects to develop curriculums (Rossman & Chance - CSI project; Rob Gould - UCLA; Webster West & Roger Woodard – INCIST; Nathan Tintle – Hope College; Robin Lock & Family)

CATALST Project, Univ. of MN

- Collaborative project funded by NSF
- Curricular materials based on research in cognition and learning, instructional design principles, and new ideas about content for teaching intro stats
- Materials expose students to the power of statistics, real problems, and real, messy data
- Radical changes in content and pedagogy: No t-Tests; randomization and re-sampling approaches; MEAs
- Assess GAISE-recommended student outcomes

Teaching Experiment

- 2010/2011: Two-semester teaching experiment (Year 3 of grant)
- Goals:
  - Immerse students in statistical thinking
  - Change the pedagogy and content
  - Move to randomization/simulation approach to inference
  - Have students really "cook"

Preparation for the Teaching Experiment

- Reading, thinking, writing, adapting MEAs
- Planning and decisions about sequence of course content, software choice(s), etc.
- Conversations and working sessions with visiting scholars:
  - George Cobb
  - Cliff Konold
  - Richard Lesh
  - Robin Lock
  - Dennis Pearl
  - Allan Rossman
  - Beth Chance
  - John Holcomb
  - Chris Wild
  - Robert Gould
  - Nick Horton
  - Daniel Kaplan

Teaching Experiment: Semester 1

- Teaching CATALST in 1 section of undergraduate course (~30 students)
- Unit 1 was written (+ MEAs for Unit 2 and 3)
- Plans/Outline for Unit 2 and 3
- Plans for software (TinkerPlots™, R-Tools, and R)
- Weekly meetings to debrief and plan

Teaching Experiment: Semester 1

- Graduate student taught the course
- Andy observed almost every class (sometimes co-taught)
- 2-3 graduate students would also observe each class
- Weekly team meetings (2-3 hours) to debrief, plan and create materials, and conduct ongoing analysis of course assessment data
- Other weekly meetings (~2 hours) to develop assessments; scoring rubrics; talk about grading, etc.
- CATALST PIs and instructor had a monthly meeting to touch base and plan (Conference call)
- Student interviews after first unit

Ch-ch-ch-ch-Changes

- Team met in January to make changes based on what was learned during the semester (also met with 6 potential implementers)
- Re-sequencing of some topics (e.g., bootstrap)
- Course readings added (content) and removed (abstracts only)
- Assessments adapted as needed
- Group exams rather than individual

Teaching Experiment: Semester 2

- Teaching CATALST in 3 sections of undergraduate course (each with ~30 students)
- Also being taught in one course at North Carolina State University (Herle McGowan)
- Units 1 & 2 were written
- Plan/Outline for *new* Unit 3

Teaching Experiment: Semester 2

- Sections at Minnesota taught in active-learning classrooms
- Students brought their own laptops to class
- Development team met weekly
- Teaching team met weekly (Herle Skyped into the meeting)
- CATALST PIs and instructors had a monthly meeting

What We Have Learned

- *We can* teach students to "cook"
- Based on interview and assessment data, students seem to be thinking statistically (even after only 6 class periods!)
- *We can* change the content/pedagogy of the introductory college course
- *We can* use software at this level that is rooted in how students learn rather than purely analytical

Judging Randomness MEA

Students create rules in order to judge whether or not the shuffle feature on a particular Apple iPod appears to produce randomly generated playlists.

1) Students' initially describe characteristics of 25 song playlists that were randomly generated (Population: 8 Artists, 10 songs each)

Playlist Example

(Set 1) Track	Artist	Album	Year
The National Anthem	Radiohead	Kid A	2000
Hells Bells	AC/DC	Back In Black	1980
Immigrant Song	Led Zeppelin	Led Zeppelin III	1970
Me And The Major	Belle & Sebastian	If You're Feeling Sinister	1996
Back In Black	AC/DC	Back In Black	1980
Eyes For Windows	Damien Jurado	Rehearsals For Departure	1999
Trust Me	Janis Joplin	Pearl	1971
Gallows Pole	Led Zeppelin	Led Zeppelin III	1970
Ohio	Damien Jurado	Rehearsals For Departure	1999
Flat Chested Girl From Maynardville	Bobby Bare Jr.	Young Criminal's Starvation League	2002
Judy And The Dream Of Horses	Belle & Sebastian	If You're Feeling Sinister	1996
Buried Alive In The Blues	Janis Joplin	Pearl	1971
Me & Bobby McGee	Janis Joplin	Pearl	1971
Tornado	Damien Jurado	Rehearsals For Departure	1999
Mayfly	Belle & Sebastian	If You're Feeling Sinister	1996
How To Disappear Completely	Radiohead	Kid A	2000
That's The Way	Led Zeppelin	Led Zeppelin III	1970
My Baby	Janis Joplin	Pearl	1971
Stay In Texas	Bobby Bare Jr.	Young Criminal's Starvation League	2002
Idioteque	Radiohead	Kid A	2000

Judging Randomness MEA

Students create rules in order to judge whether or not the shuffle feature on a particular Apple iPod appears to produce randomly generated playlists.

1) Students' initially describe characteristics of 25 song playlists that were randomly generated (Population: 8 Artists, 10 songs each)
2) Generate rules for "not random"; test and revise their rules (model) using five additional playlists.
3) They apply their model to three particular playlists that have been submitted to Apple by an unhappy iPod owner.
4) They write a letter to the iPod owner explaining the use of their model and their final conclusion.

Example Student Report

Mr. Hoffman, our group has done some research on the recent complaint to Apple about the shuffle feature on the iPod. From the research, we have compiled a list of three rules to determine whether a playlist is randomly generated or not. One, if there are less than six artists represented, then the playlist is not random. Two, if there are repeated songs in the playlist, then it is not random. Thirdly, if an artist is repeated consecutively more than three times then the playlist is not random.

Example Student Report (cont.)

Based on our rules for the shuffle feature of the iPod, your original three playlists fall within these criteria of the rules so the playlists are considered to be randomly generated. There is expected to be some artists that are played more frequently than others because if the shuffle feature is truly random, there wouldn't necessarily be equal play among the artists. It doesn't matter if the artists are repeated because shuffle is random, not equal. Thank you for concern with our product.

Informal Statistical Inference Evident

- **NULL MODEL:** Characteristics of random lists
- **JUDGMENT CRITERIA:** What is "atypical" or "uncommon" for random lists (i.e., not likely)
- **COLLECT DATA** and **COMPUTE STATISTICS**
- **APPLY THE CRITERIA TO MAKE A JUDGMENT:** Is the data surprising given the characteristics of random lists?

NOTE: Written first day of class

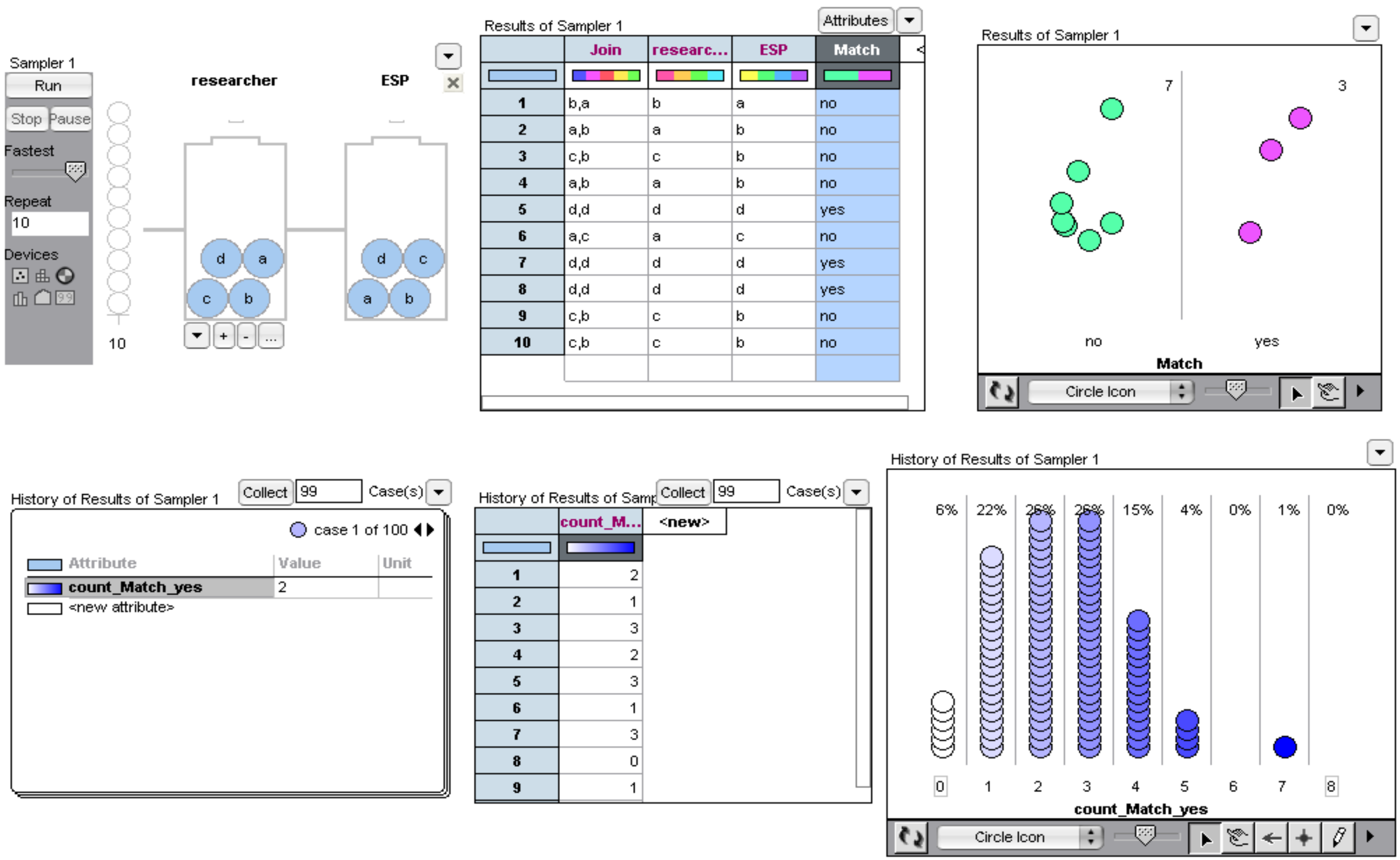
6th Week Interview Task
ESP Problem

Imagine a researcher is conducting an experiment to study the merits of ESP. The researcher has 10 cards, and each card has one of four different images on it. The researcher first shows the person being tested a sheet of paper that shows the four different possible images. The 10 cards are not shown to the person and they are placed face down in a stack on a table in front of the person.

Without showing it to the person claiming to have ESP, the researcher takes the first card on top of the stack, looks at the card, and concentrates on the image. The person is asked to identify which of the four images is on the card the researcher is holding. This is repeated for all 10 cards.

The person being tested correctly identifies the image on 5 of the 10 cards. What would you do to determine if this person really has ESP? Explain.

ESP Problem Student Model



Student Explanation ESP Problem

"So I want to count how many yeses per trial, so I created a new, I guess it's a new attribute that just counts yeses. And then I'm going to run 100 of these. …so I'm plotting for each trial how many yeses I got in each trial. And then I would again look and see how likely it is to get; what did he get, 5 out of 10? How likely it is to get 5 out of 10 if it's just by chance. And if it's, if it's a likely thing I would say he doesn't have ESP powers. But if it's not likely, then perhaps more research [is] necessary.

I would say they are kind of right on the border of what is, of what we would call statistically significant, I think. And it's not, it's certainly not very likely to get five, but it's possible. So I wouldn't say yes or no, that they have ESP powers. I would say it's, we would run more trials and see what happens."

Statistical Inferential Thinking

CREATE A NULL MODEL

GENERATE DATA BASED ON THE MODEL: So I want to count how many yeses per trial…I’m going to run 100 of these.

COMPARE DATA TO THE MODEL: And then…look and see how likely it is to get…5 out of 10 if it's just by chance.

MAKE JUDGMENT BASED ON CHANCE: And if it's…a likely thing I would say he doesn't have ESP powers. But if it's not likely, then perhaps more research [is] necessary.

SKEPTICISM: It's certainly not very likely to get five but it's possible. So I wouldn't say yes or no, that he has ESP powers.

To Bring About Change...

- It takes a village
  - A village of "great minds" to brainstorm, offer input, and give feedback (from the beginning)
  - A village of dedicated individuals
  - A village of collaborators who believe in the ideas enough to dare to teach the course
  - Conducting a collaborative teaching experiment requires time to discuss, argue, observe, and evaluate

Thank You for Your Attention

It Takes a Village: Future Directions for Statistics Education Research

Bob delMas
Department of
Educational Psychology
University of Minnesota
delma001@umn.edu

University of Minnesota