

# A complex learning progression structure for a critical statistical thinking construct

Amy E. Arneson<sup>1</sup>, James Mason<sup>1</sup>, Diah Wihardini<sup>2</sup>, & Mark Wilson<sup>1</sup>

<sup>1</sup>Graduate School of Education, University of California, Berkeley

<sup>2</sup>Bina Nusantara University, Jakarta

NCME 2018 Annual Meeting  
15 April 2018



NATIONAL  
MATH + SCIENCE  
INITIATIVE

# OUTLINE



- Overview of college-ready statistical thinking
- The *formal inference* construct definition
- Sample tasks and responses
- Preliminary results
- Discussion and next steps

# INTRODUCTION

College-Ready Statistical Thinking



# STATISTICAL THINKING IN K-12 EDUCATION

## College-Ready Statistical Thinking

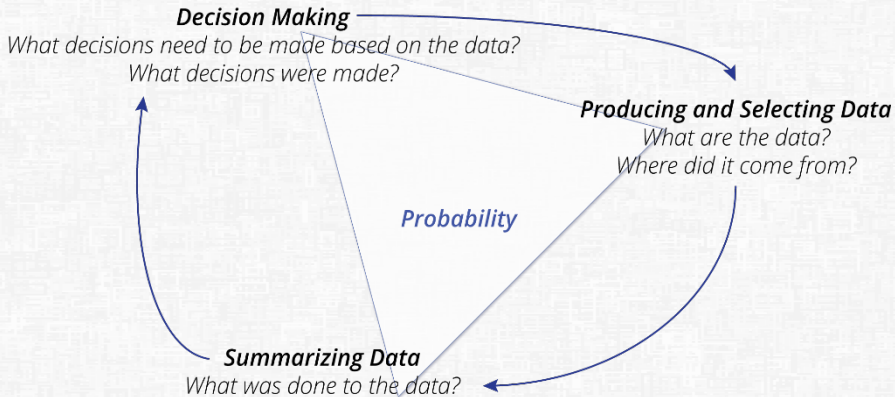


"Citizen" Statistical Literacy  
*Critical Questioning (Watson, 2006)*



Data Modeling & Statistical Reasoning  
*Data Modeling (Lehrer et al., 2014)*

# A FRAMEWORK FOR COLLEGE-READY STATISTICAL THINKING



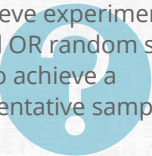
# COLLEGE-READY STATISTICAL THINKING

Critical questions  $\Rightarrow$  Constructs

## Producing and Selecting Data

Are the data the right data?

Was random assignment used to achieve experimental control OR random sampling used to achieve a representative sample?



.

.

.

## Summarizing Data

What do the provided summary statistics show and hide about the data?

What does the display show and hide about the data?



.

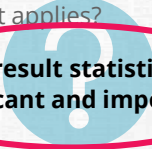
.

.

## Decision Making

Does the claim overreach at all – in terms of (1) causality and (2) the population to which it applies?

**Is the result statistically significant and important?**



.

.

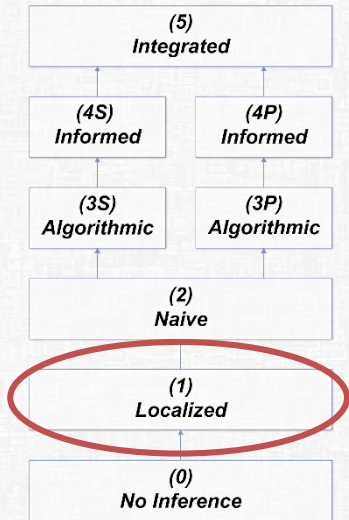
.

# THE CONSTRUCT MAP

Formal Inference (Fol, Fol-S, Fol-P)



# FORMAL INFERENCE (FOI) LEVELS 0 & 1



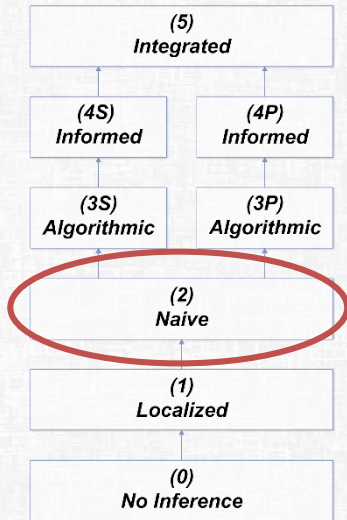
(LEVEL 1) Localized

- Over-influenced by the details of the current data collection.
- “Significance” is ignored.
- No interpretation in terms of a population.

(LEVEL 0) No inference



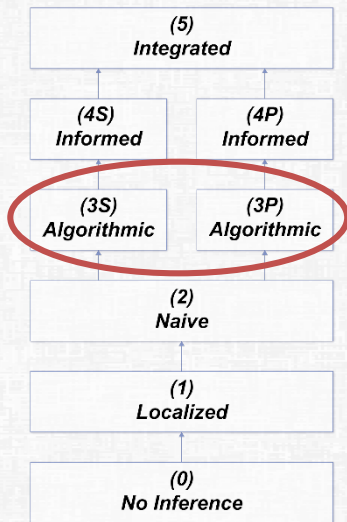
# FORMAL INFERENCE (FOI) LEVEL 2



## (LEVEL 2) Naive

- Does acknowledge random sample variability.
- Determines significance on the basis of personal ideas about magnitude.

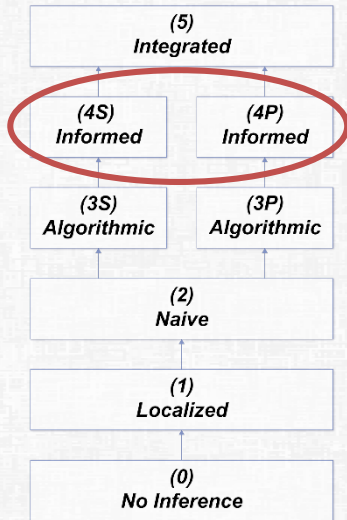
# FORMAL INFERENCE (FOI) LEVELS 3P & 3S



## (LEVEL 3) Algorithmic

- Calculates and reports inference statistics (S) and effect sizes (P).
- Statistical tests may be inappropriately applied.
- Interpretations of calculated values may not be correct.

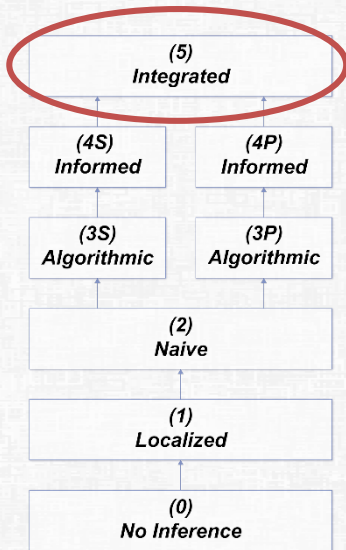
# FORMAL INFERENCE (FOI) LEVELS 4P & 4S



## (LEVEL 4) Informed

- Interprets inference statistics (S) and effect sizes (P) correctly.
- Understands the underlying logic of the process.
- Pays attention to the appropriateness of the procedure.

# FORMAL INFERENCE (FOI) LEVEL 5



## (LEVEL 5) Integrated

- Interprets statistically significant results in terms of an effect size.
- This is the basic task in understanding the limitations of statistical significance.

# THE ITEMS DESIGN

Statistical Thinking Tasks



# ITEM DEVELOPMENT PRINCIPLES

## Item clusters

- Statistical thinking is complex
- Stimulus material is cognitively demanding

## Advisory panel review

- Relevance to first-year college coursework is the priority
  - K-12 STEM specialists
  - University instructors
- Critical thinking > jargon/technical language

# SAMPLE ITEM CLUSTER

## Miles per Gallon / Liters per 100 km

### Stimulus TEXT

An automotive industry lobbyist wishes to claim that cars have been getting more energy efficient over time, and thus the federal government should relax some of the environmental regulations they impose in the industry.

To back up this claim, they use Environmental Protection Agency (EPA) fuel economy data to compare the "city" miles per gallon (MPG) ratings of random samples of cars from 2012, 2013, and 2014. Using statistical software, they found the mean MPG along with 95% confidence intervals. These are reported in the following table.

### Stimulus DISPLAY

	Mean MPG	95% confidence interval
2012	19.41	(18.01, 20.81)
2013	20.73	(19.40, 22.06)
2014	22.03	(20.84, 23.22)

### Items on MULTIPLE DIMENSIONS

[1] The lobbyist makes the claim "There was a statistically significant improvement in fuel efficiency from 2012 to 2013." Do you agree with this statement? Why or why not?

[2] The lobbyist also makes the claim "There was a statistically significant improvement in fuel efficiency from 2012 to 2014." Do you agree with this statement? Why or why not?

[3] Is the observed improvement from 2012 to 2014 in fuel economy large enough to be of importance to an average driver? Explain why or why not.

[4] The lobbyist claims that because these data show an improvement, the government can relax some environmental regulations. Are you convinced that regulations can be relaxed? Why or why not?

FoI-S

FoI-S

FoI-P

LDC

## SAMPLE ITEM – ZOOM IN

Miles per Gallon [1] (Fol-S)

[1] The lobbyist makes the claim

*“There was a statistically significant improvement in fuel efficiency from 2012 to 2013.”*

Do you agree with this statement? Why or why not?

	Mean MPG	95% confidence interval
2012	19.41	(18.01, 20.81)
2013	20.73	(19.40, 22.06)
2014	22.03	(20.84, 23.22)



# THE OUTCOME SPACE

Mapping Responses to Fol



# SAMPLE RESPONSE

to Miles per Gallon/Liters per 100 km [1]

*I do not fully agree with this statement. I agree with the fact that there was an improvement from 2012 to 2013, but there was no significant statistical improvement.*

*It is not a significant improvement because the percentage decrease was very small.*

*[irrelevant content removed]*

*The interval shows that the max value in 2013 could be higher than 12, and the min value in 2012 could be less 11.39. In this case, there would be no improvements in efficiency.*

What's happening?

- Conflation of **statistical** and **practical** significance. BUT information on BOTH sides?
- Discussion of practical significance is not using technical terminology, should we interpret it as practical significance?
- Misinterpretation of a confidence interval.
- Incorrect procedure applied.

# SAMPLE RESPONSE

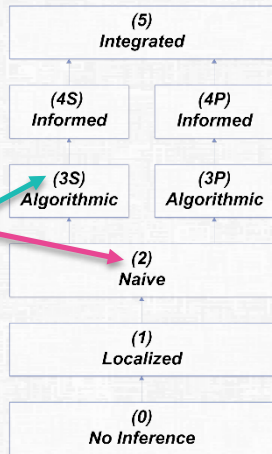
to Miles per Gallon/Liters per 100 km [1]

*I do not fully agree with this statement. I agree with the fact that there was an improvement from 2012 to 2013, but there was no significant statistical improvement.*

*It is not a significant improvement because the percentage decrease was very small.*

*[irrelevant content removed]*

*The interval shows that the max value in 2013 could be higher than 12, and the min value in 2012 could be less 11.39. In this case, there would be no improvements in efficiency.*



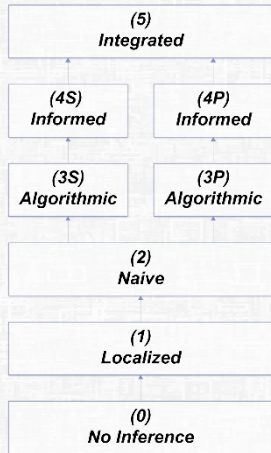
# THE MEASUREMENT MODEL

Preliminary Results

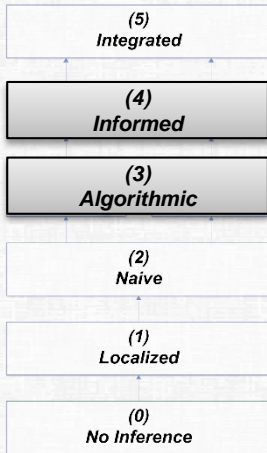


# WHAT IS THE DIMENSIONALITY OF FOI?

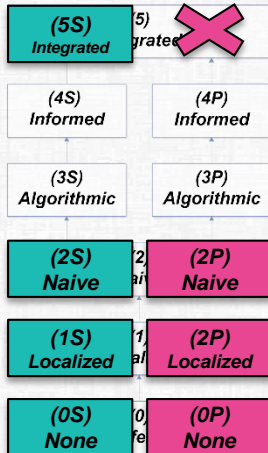
## UNIDIMENSIONAL complex



## UNIDIMENSIONAL linear



## 2 DIMENSIONAL complex



# (VERY) PRELIMINARY RESULTS - DATA

## 21 Polytomously Scored Items

41 Respondents

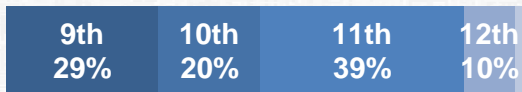
Dimension



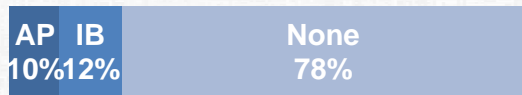
Format



Grade Level



Advanced Statistics Coursework

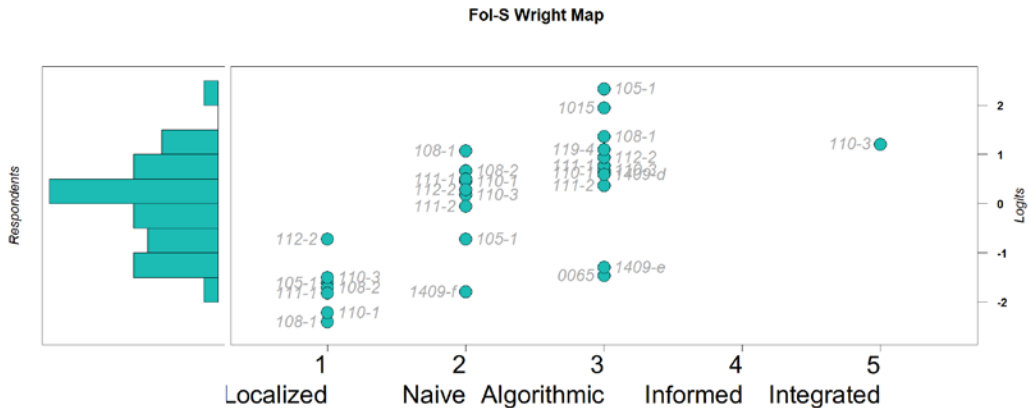


Test Form



# PRELIMINARY RESULTS (2-DIMENSIONAL ANALYSIS)

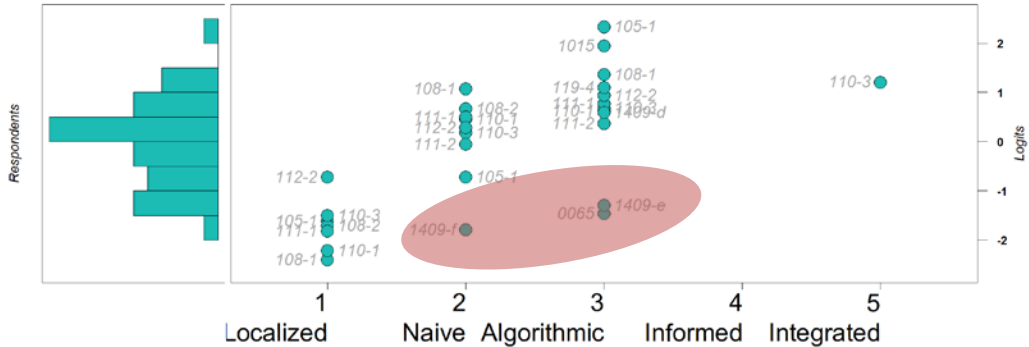
## Fol-Statistical



# PRELIMINARY RESULTS (2-DIMENSIONAL ANALYSIS)

## Fol-Statistical

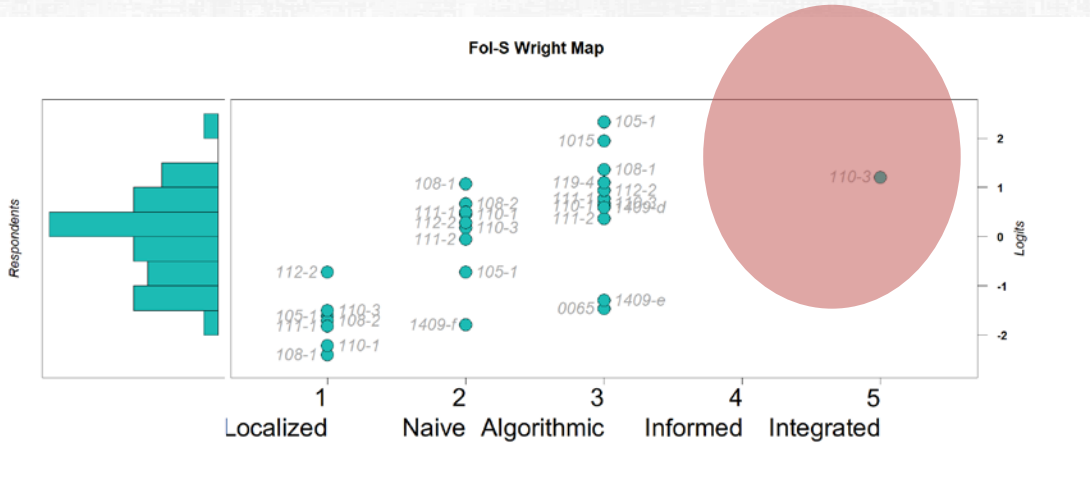
Fol-S Wright Map





# PRELIMINARY RESULTS (2-DIMENSIONAL ANALYSIS)

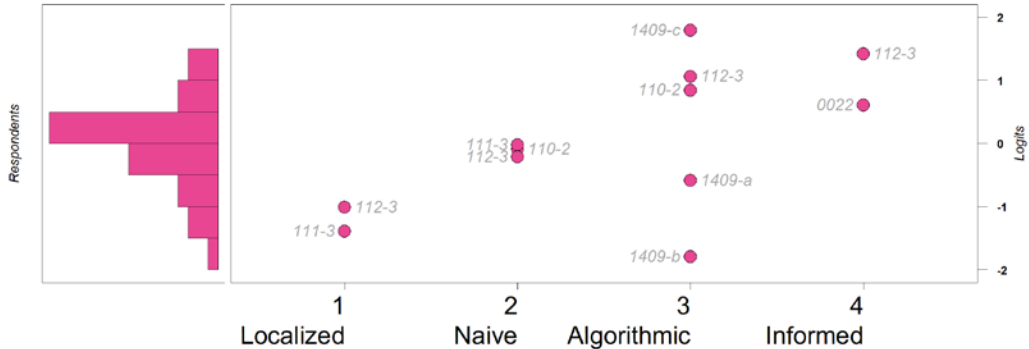
Fol-Statistical



# PRELIMINARY RESULTS (2-DIMENSIONAL ANALYSIS)

## Fol-Practical

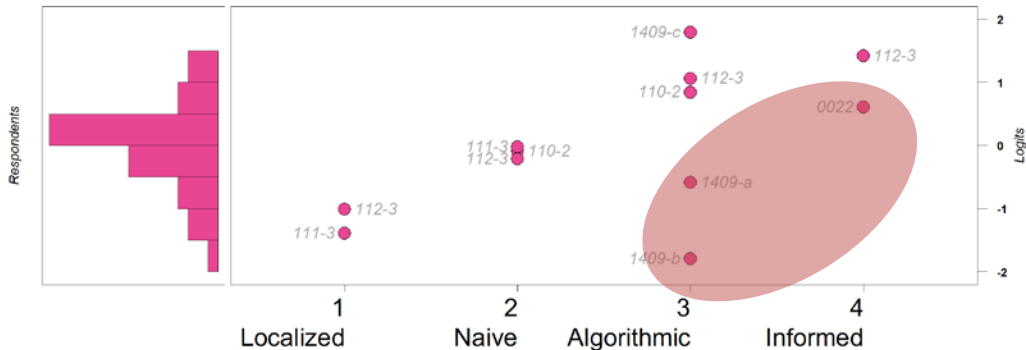
Fol-P Wright Map



# PRELIMINARY RESULTS (2-DIMENSIONAL ANALYSIS)

## Fol-Practical

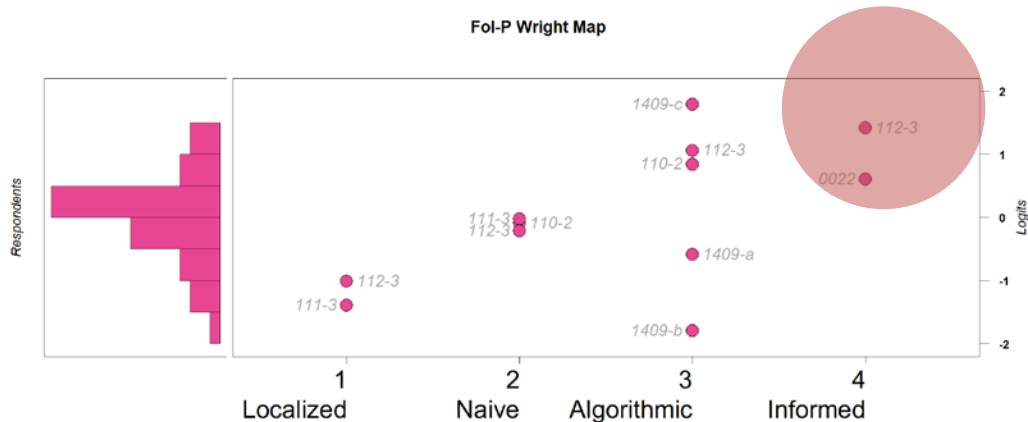
Fol-P Wright Map



# PRELIMINARY RESULTS (2-DIMENSIONAL ANALYSIS)

## Fol-Practical

Fol-P Wright Map



**AND?**

Learnings and Next Steps



# NEXT STEPS

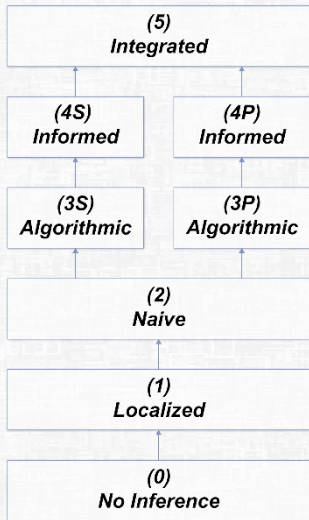
## More data!

- Test dimensionality hypotheses
- Complex modeling (e.g. ordered partition models)
- Investigate cluster (testlet) effects
- Cutoff points for “college-ready”

## More questions!

- What are the relationships among the S & P levels (across the branches)?
- How should we design reports for different audiences?

# TAKEAWAYS



- Measuring critical thinking is complex
  - Stick to principles!
  - Stick to theory!

# ACKNOWLEDGEMENTS

## Advisory Panel

Shruti Bathia	Tzur Karelitz
Alexander Blum	Maureen Lahiff
Jonathan Bostic	Linda Morrell
Rich Brown	Rich Patz
Michelle Carney	Julia Phelan
David Conley	Laura Pryor
Karen Draney	Weeraphat Suksiri
Perman Gochyyev	Josh Sussman
Emiliano Gomez	Yukie Toyama
Jerred Jolin	Danhui Zhang

## BEAR IT

Ana Maria Albornoz Reitze  
Rebecca Freund  
Shazi Khan  
Chanel Quisol  
David Torres Irribarra

***This work was supported by the National Math and Science Initiative (NMSI).***





**Thank you!**

[aea@berkeley.edu](mailto:aea@berkeley.edu)