

# Designing Formative Studies

(and a little bit about pilot studies)

Todd Kulesza

# Summative vs Formative

Summative studies examine **outcomes**.

Formative studies **inform design**.

# Where they differ

Summative	Formative
Answers about outcomes	Clues about behavior/ needs

# Where they differ

Summative	Formative
Answers about outcomes	Clues about behavior/ needs
Large[r] participant pool	Small[er] participant pool

# Where they differ

Summative	Formative
Answers about outcomes	Clues about behavior/ needs
Large[r] participant pool	Small[er] participant pool
Often analyzes aggregates	Often focuses on specific participants

# An example

**Eventual goal:** help people make  
intelligent assistants smarter

# An example

**Eventual goal:** help people make intelligent assistants smarter

**Roadblock:** how do people communicate with intelligent assistants?

# Todo list

What do we want to **learn**?



# Todo list

What do we want to **learn**?



How can we **collect** data?

# Todo list

What do we want to **learn**?



How can we **collect** data?



How should we **analyze** participant data?

# Todo list

What do we want to **learn**?



How can we **collect** data?



How should we **analyze** participant data?



What are the design **implications**?

# What do we want to learn?

Clearly frame the... {

# What do we want to learn?

Clearly frame the... { problem



# What do we want to learn?

Clearly frame the... { problem domain



# What do we want to learn?

Clearly frame the... { problem  
domain  
population



# Framing example

We want to help *people*  
*improve intelligent assistants.*



# Framing example

We want to help *people*  
*improve intelligent assistants.*

Vague!



Vague!



Vague!



# Framing example

We want to help *people*  
*improve intelligent assistants.*

Vague!



Vague!



Vague!



What's a popular type  
of assistant?

# Framing example

We want to help *people*  
*improve intelligent assistants.*

Vague!

Vague!

Vague!

What's a popular type  
of assistant?



Why do they matter?

# Framing example

We want to help *people*  
*improve intelligent assistants.*

Vague!

Vague!

Vague!

What's a popular type  
of assistant?



Why do they matter?



Who uses these  
assistants?

# Framing example

We want to help *people*  
*improve intelligent assistants.*

Vague!

Vague!

Vague!

What's a popular type  
of assistant?



Why do they matter?



How can they be  
improved?



Who uses these  
assistants?

# Qualitative analysis

Time	Text	Code
1	So just to make sure for box E these are already, is there something to be fixed in box E because there are already grades? <i>Proctor: I can't tell you that.</i> Okay, Sum of AF12>7 AF12?	
2	Let's look at the formula [clicks the evaluates formula for the total points column. Clicks trace dependents for total points] Where's E12 at? E12/E6. Total points possible, okay.	
3	[examines the grade table near the top] Okay, let's evaluate the letter grades [traces dependents for letter grades. Scrolls right to look at the labs attended. Evaluates formula] Z12=3.	
4	Why does that arrow stop at five and not at total labs? [referring to the trace dependents of the first letter grade G12. Traces dependents on GPA]	

# Our domain and population

# Our domain and population

**What:** Naive Bayes text classifiers



# Our domain and population

**What:** Naive Bayes text classifiers

**Why:** Save time & effort in qualitative analysis

# Our domain and population

**What:** Naive Bayes text classifiers

**Why:** Save time & effort in qualitative analysis

**Who:** Social science researchers

# Our domain and population

**What:** Naive Bayes text classifiers

**Why:** Save time & effort in qualitative analysis

**Who:** Social science researchers

**How:** Explaining specific reasons

# Our research questions

# Our research questions

- | How do end users think text-classifying intelligent assistants make their decisions?

# Our research questions

1 How do end users think text-classifying intelligent assistants make their decisions?

2 Can such assistants [intelligibly] explain their reasoning to end users?

# Our research questions

- 1 How do end users think text-classifying intelligent assistants make their decisions?
- 2 Can such assistants [intelligibly] explain their reasoning to end users?
- 3 How do end users “naturally” try to fix an assistant’s reasoning?

# Natural programming

Human-centered approach



# Natural programming

Human-centered approach

Study **how** people perform tasks

# Natural programming

Human-centered approach

Study **how** people perform tasks

Build tools **around** existing behavior

# Collecting data

Surveys

Interviews

Low-fi prototypes

# Collecting data

Surveys ← Only answers #1

Interviews

Low-fi prototypes

# Collecting data

Surveys

Only answers #1

Interviews

Answers #1 &  
*maybe part of #3*



Low-fi prototypes

# Collecting data

Surveys

Only answers #1

Interviews

Answers #1 &  
*maybe part of #3*

Low-fi prototypes



Ding! Ding! Ding!

(often paired with think-aloud study)

# Design considerations

Need to measure effects of explanations

Small sample size

Avoid influencing participant reasoning

Open-ended feedback

# Design considerations

Two treatments: one with explanations of assistant's reasoning, one without

Small sample size

Avoid influencing participant reasoning

Open-ended feedback



# Design considerations

Two treatments: one with explanations of assistant's reasoning, one without

Within-subjects design

Avoid influencing participant reasoning

Open-ended feedback

# Design considerations

Two treatments: one with explanations of assistant's reasoning, one without

Within-subjects design

Treatment lacking explanations needs to be first

Open-ended feedback

# Design considerations

Two treatments: one with explanations of assistant's reasoning, one without

Within-subjects design

Treatment lacking explanations needs to be first

Paper prototype, audio/video recording

# Our paper prototype

How do end users explain  
reasoning to an intelligent assistant?

# Our paper prototype

How do end users explain reasoning to an intelligent assistant?

Can an assistant explain its logic to end users?

# Our paper prototype

1	Okay, Sum of AF12 > 7 AF12?  They didn't answer whether AF12 > AF12, So no info was gained or lost.	None  OK!
2	lets look at the formula [clicks the evaluates formula for the total points <u>column.</u>  ~ looking for info.	None  Seeking info.

# Our paper prototype

80	[looks at waive instructions]	None
<u>This was coded as None because:</u>		ok!
<ul style="list-style-type: none"><li>• Actions following "Info Gained" are almost certainly "None"</li><li>• Segments beginning with "looks at" are probably "None"</li></ul>		
<u>This was not coded as Seeking Info because:</u>		
<ul style="list-style-type: none"><li>• Segments following "Info Gained" are probably not "Seeking Info"</li><li>• Segments containing "instruction" are probably not "Seeking Info"</li></ul>		
<u>This was not coded as Info Gained because:</u>		
<ul style="list-style-type: none"><li>• Actions are probably not "Info Gained"</li><li>• Segments lacking "OK" are probably not "Info Gained"</li></ul>		
<u>This was not coded as Info Lost because:</u>		
<ul style="list-style-type: none"><li>• Since one of the next 3 segments is "Info Gained", this is probably not "Info Lost"</li><li>• Segments beginning with a verb are almost certainly not "Info Lost"</li></ul>		Need a "seeking info" before "Info Lost"

# Pilots

**Pilot studies** debug your experiment

Materials and procedures may  
change

Software parameters may require  
tuning



# Our pilot results

Problem

Adaptation

# Our pilot results

Problem	Adaptation
Written feedback was unclear	Proctor needs to ask questions during the study

# Our pilot results

Problem	Adaptation
Written feedback was unclear	Proctor needs to ask questions during the study
Verbal feedback tapered off	Proctor needs to prod participants

# Our pilot results

Problem	Adaptation
Written feedback was unclear	Proctor needs to ask questions during the study
Verbal feedback tapered off	Proctor needs to prod participants
Participants appeared annoyed after experiment	Shorten study duration

# Analysis

More often **qualitative** than **quantitative**

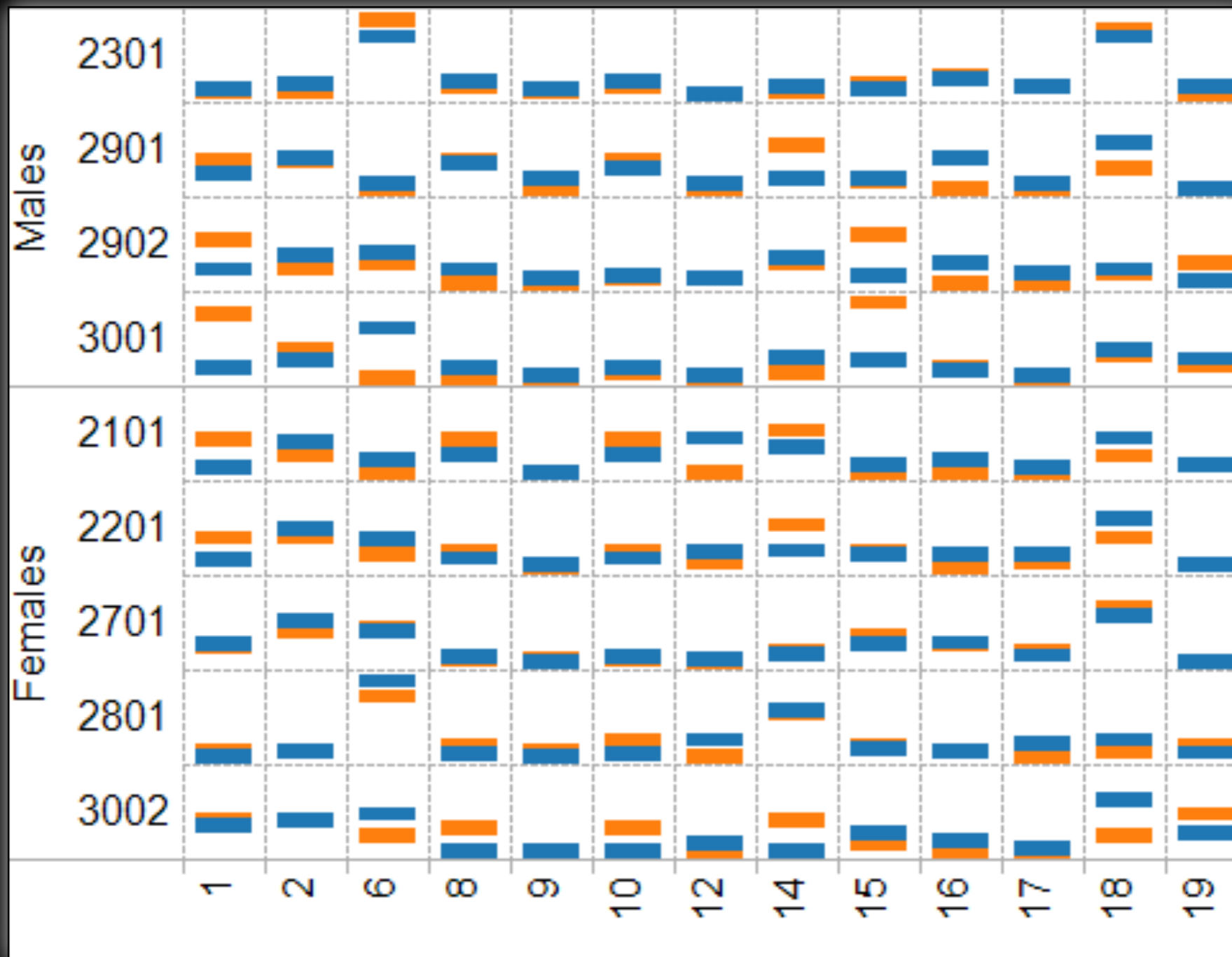
Focus on **individual participants**  
instead of aggregates

Look for ties to existing **theories**

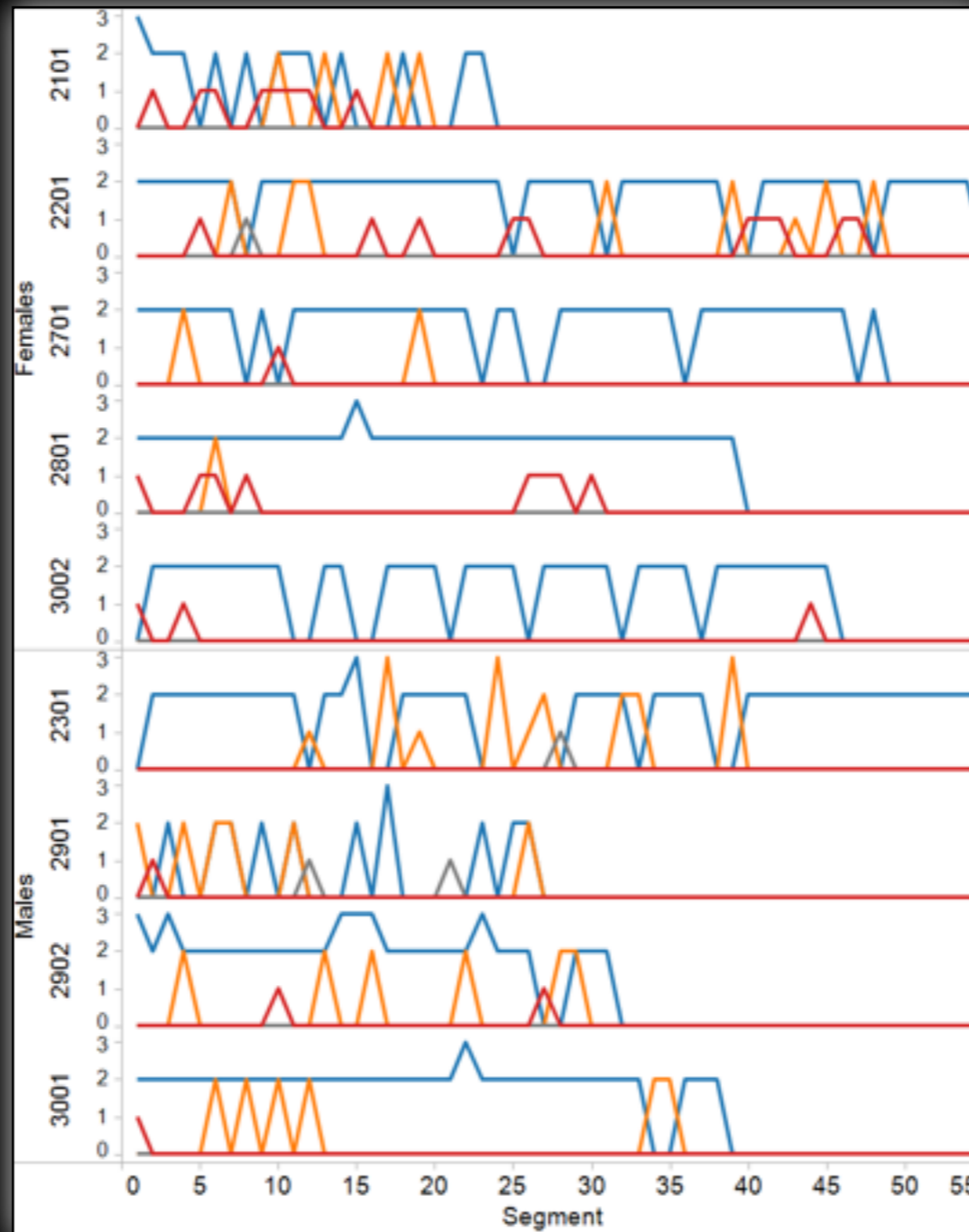
# Our analysis methods

Subject	Word single	Word multiple	Word present	Punctuation	Relationships	KB-English	Probabilities
2101	●	● ■	●		● ■	●	■
2201	●	■	●	●	■	● ■	
2301		● ■	●		■	■	■
2701	● ■		● ■	●	■	●	
2801	●	■	● ■			■	■
2901	●	■	● ■				■
2902	● ■		● ■	● ■	■	■	
3001	● ■		● ■	● ■	■	■	
3002			● ■		■		■

# Our analysis methods

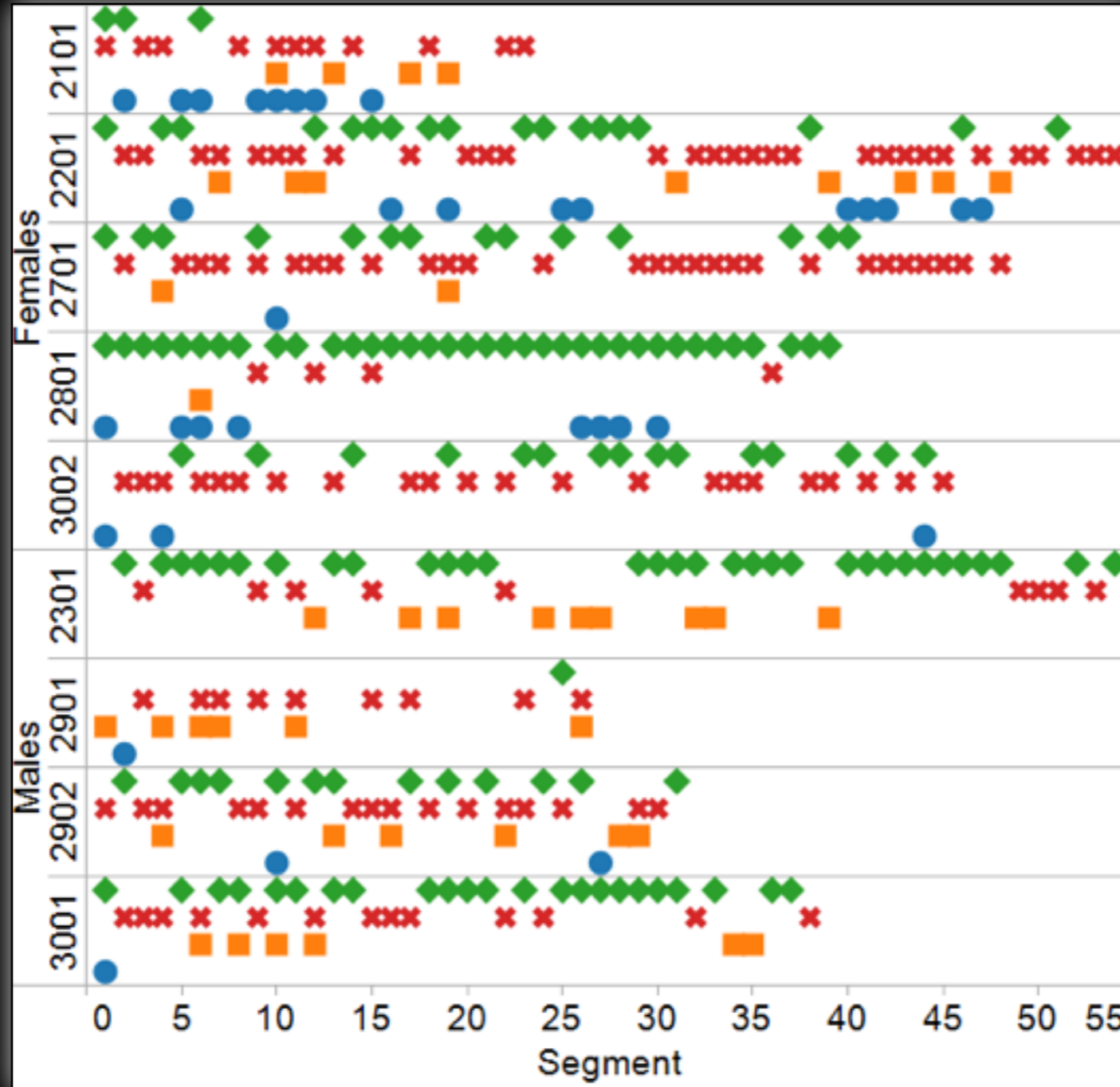


# Our analysis methods

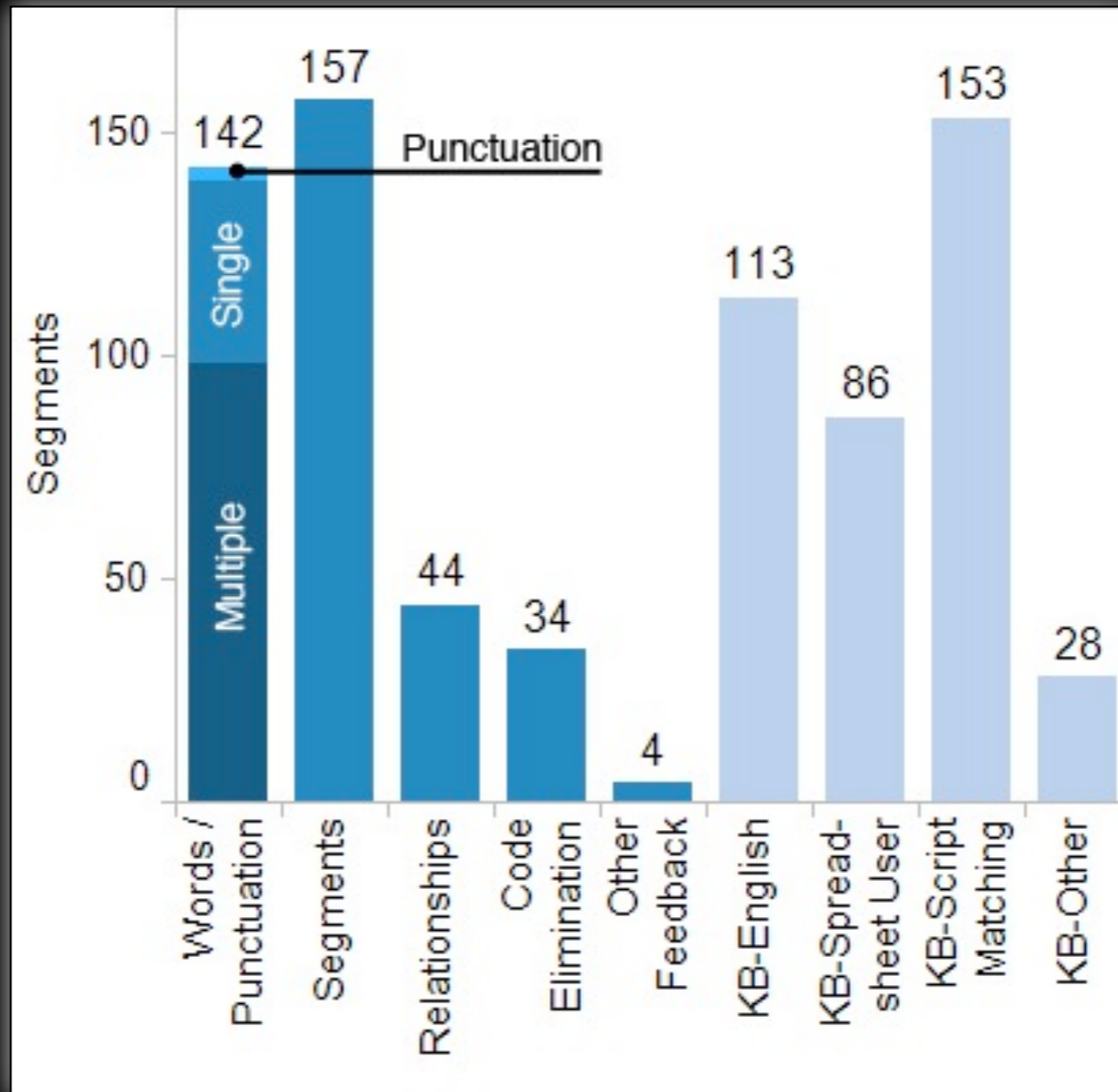




# Our analysis methods



# Our analysis methods



# Design implications

How can we support observed behaviors?

Priorities?

Motivation for other research?

How would you craft a formative study for your term project?