

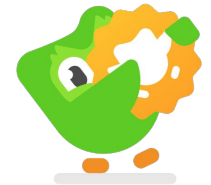
# Introduction to Writing Analytics: Fundamental Principles and Data Exploration

Jill Burstein

Duolingo

T3 @ LASI21

June 21-25, 2021



duolingo

# Tutorial Goals

- ✓ Participants walk away with a **fundamental understanding** of writing analytics
- ✓ Participants have sufficient **knowledge** to develop writing analytics research questions to pursue research projects
- ✓ Participants have an **interest** to continue to develop skills to conduct writing analytics research

# Tutorial Outline

- ✓ **Welcome (5 minutes)**
- ✓ **What is Writing Analytics? (10 minutes)**
  - Working Definition & Broad Implications
  - Writing Frameworks & Cross-Disciplinary Interaction
  - Research Workflow
- Example Writing Analytics Research Projects (40 minutes)**
- ✓ Writing Mentor -- Google Docs Add-on for writing support  
IES WAVES -- US Department of Education, IES-funded research project
- ✓ **BREAK (5 minutes)**
- ✓ **Data Exploration Activity: What can we explore/learn, and how can writing analytics contribute to education technology innovation? (20 minutes)**
  - Review Writing Mentor Process and Product feature data
  - Review `IES-WAVES` Data: Attitudes, Genre & College Retention
  - Example Explorations/Visualizations

# What is Writing Analytics?

Working Definition, Writing Framework/Cross-Disciplinary Interaction, and Data Considerations & Research Workflow

# Working Definition & Broad Implications

**Definition:** Writing analytics focuses on the **collection**, **measurement** and **analysis** of student writing to investigate patterns in writing process and product data to provide insights about: 1) student writing achievement at a specific point in time, or over a period of time, and 2) relationships between writing and broader academic outcomes.

**Practical Implications in Education:** Yield insights that inform student support mechanisms (such as early warning flags), and improvements in educational settings for writing (such as, writing instruction curriculum, and automated personalized instruction).

# Sociocognitive Writing Framework & Cross-Disciplinary Interaction

# Socio-cognitive framework

Framework for understanding relationships between process and product feature data & areas of writing achievement

## Socio-cognitive writing achievement framework components

- *Writing domain knowledge* (e.g., vocabulary choice)
- *General skills* (e.g., critical thinking, reading skill)
- Content knowledge
- *Intrapersonal* factors (e.g., motivation)
- *Interpersonal* factors (e.g., collaboration)
- Neurological factors (e.g., task attendance, vision)

# Cross-Disciplinary Interactions

- Writing studies
- Disciplinary constructs (e.g., STEM content knowledge)
- Learning analytics
- Educational measurement
- Computer & Data science
- Natural language processing (NLP)
- Automated writing evaluation (AWE)



# Research Workflow

Research Questions, Data  
Considerations, & Research Methods

**Motivation:** What kinds research questions are we trying to answer?

## Example Questions

- ✓ Is there a relationship between writing domain knowledge and student interest?
- ✓ Are there relationships between writing features and student outcomes, such as course grades?
- ✓ Are students using different kinds of language when they write in different academic genres (such as, persuasive vs informational writing)?

# Data Collection Resources: Where will I get my data?

- **Institutional Digital Resources**
  - Learning Management Systems (e.g.,
  - Canvas)
- **Student Institutional Data**
  - Student Background & Demographics
  - Enrollment, Grades, Test Scores
- **Commercial Writing-Support Applications**
  - Grammarly
  - Turnitin
  - Writing Mentor
- **Publicly-Available Data**
  - [IES Writing Achievement Data](#) (ETS Github)
  - Michigan Corpora of Upper Level Student Papers ([MICUSP](#))

# Data Considerations

- **Data Privacy**
  - Personally-identifiable information (or, PII)
  - Subject consent for data use
  - Data security
- **File Format**
  - Plain text
  - Text encodings (e.g., UTF-8)
- **Data Format**
  - .csv (e.g., comma-separated, structured data)
  - JSON
  - XML
- **Data Preparation**
  - “As is” (e.g., .csv cells)
  - Chunking/tokenizing: partitioning data into meaningful units (e.g., words, sentences) based on task (e.g., discourse units)
- **Data Feature Collection & Generation**
  - Automated writing evaluation components
  - Process Data
- **Data Storage**
  - Local machine (e.g., laptop)
  - Cloud storage (e.g., for massive data sets)

# Research Methods

- **Study Purpose**
  - *Descriptive*: describe writing characteristics
  - *Predictive*: predict outcomes
  - *Prescriptive*: inform educational processes or systems (e.g., writing instruction curriculum)
- **Method Types**
  - *Exploratory data mining*, e.g., event log data from writing apps
  - *Case studies*, e.g., data from small pilot studies
  - *Experimental methods*, such as quasi-experimental or randomized-control studies

# Data Types

Writing Process & Product Data

# Data Types: Writing Process and Product Data

## Writing Process Data

- Time-stamp data

- App navigation (e.g., feature use)

- Keystroke log data

## Writing Product Data

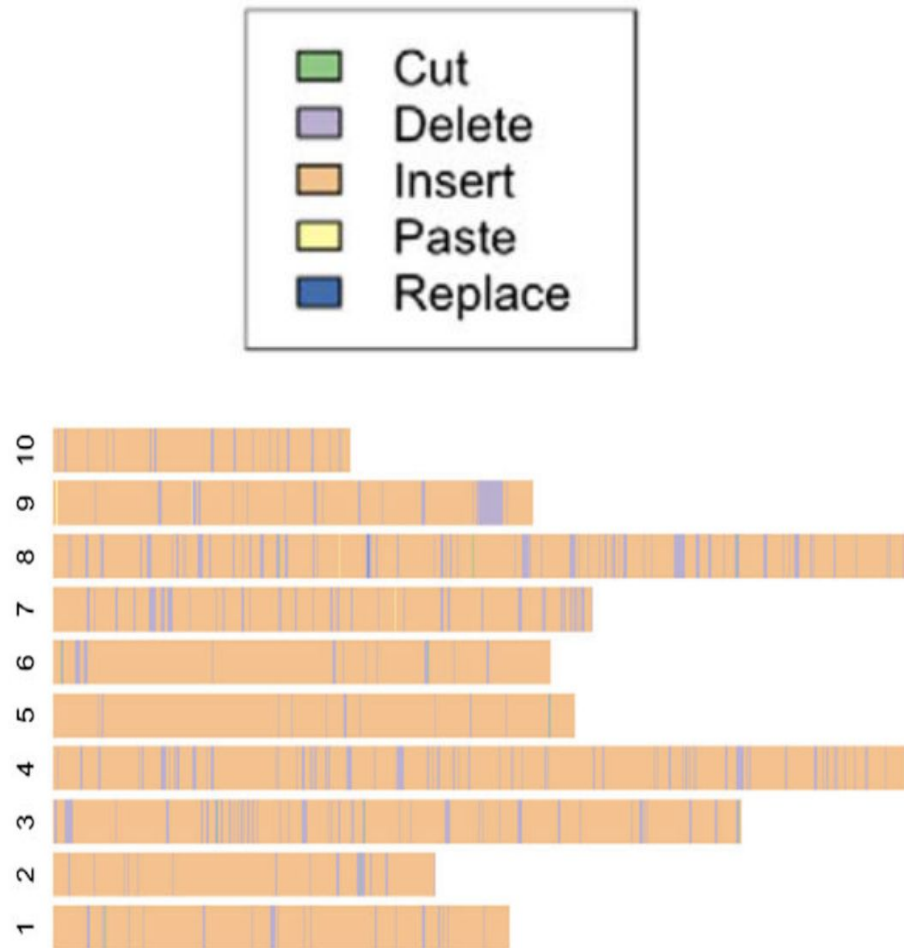
- Writing assessment responses

- Coursework assignments (K-12 through postsecondary)

- Writing app submissions

# Keystroke log analysis records

(Zhu et al, 2019)





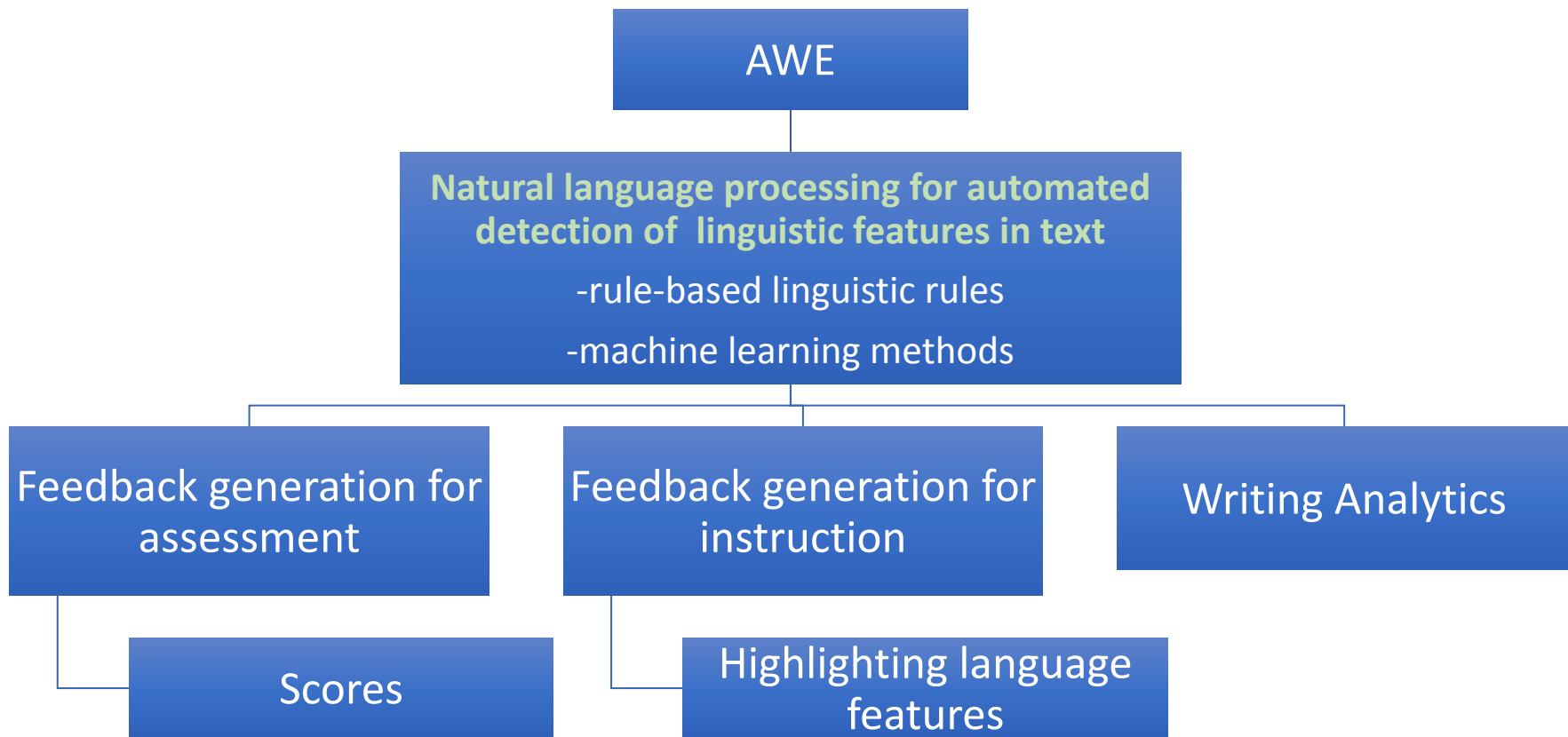
## Enhancing Interest and Performance With a Utility Value Intervention

A 2012 study conducted <sup>12</sup>by Hulleman, Godes, Hendricks, and Harackiewicz successfully proves how a utility value intervention can enhance interest and performance in many types of college students. A utility value, in this research, is about how how a task is “useful or relevant for other tasks or aspects of an individual’s life” (2). This research was conducted using two separate studies: studies 1 and 2 conducted in a laboratory and a classroom, respectively.

In study 1, 170 students from the University of Wisconsin, Madison (mostly over 90% Caucasian, but about equal in gender) were given simple math equations (two-digit multiplication) to solve using a new method they had just been taught by an audio demonstration given by the experimenters. The researchers’ hypothesis was that “participants in the relevance condition would be more interested in the math activity at the end of the session than those in the control condition. [They] also expected these effects would be moderated by participants’ performance expectation, such that students with low performance expectation would benefit more from the intervention than those with high performance expectations.” also, “utility value perceptions might also be associated with performance” (4).

The method was such that after “a measure of initial interest in math” (4) the participants were guided through the new method of completing the mathematical tasks as well as how the study would be conducted. After three minutes of practicing the mathematical technique on a problem

# Automated Writing Evaluation (AWE)



What is AWE Technology?

duolingo

# Automated Writing Evaluation Feature Types

- **Grammar, Word Usage & Mechanics Errors**
  - Subject-Verb Agreement: *the motel are ...*
  - Confused Word: *affect vs. effect*
  - Spelling errors
- **Discourse Structure**
  - Thesis statements, main points, supporting details, conclusion statements
- **Discourse coherence**
  - Measures for flow of ideas
  - Transition terms usage
- **Sentence types**
  - Variety of sentence structures
- **Vocabulary usage**
  - Vocabulary sophistication
  - Metaphor vs literal vocabulary usage
  - Personal reflection vocabulary

# Example Writing Analytics Research Projects

Writing Mentor Google Docs Add-on for writing support (Zhu , Burstein et al, 2021)

Writing Achievement Research (McCaffrey, Burstein et al, 2021)

# Writing Mentor<sup>®</sup>: User Behavior Modeling

# Research Questions

Can we identify user-in-the-wild types from WM patterns of use from a “proxy” middle/high school (MHS) data?

What relationships exist between WM MHS user types and changes in writing construct-relevant measures as captured by automated writing evaluation (AWE) features?

How does WM MHS user type performance differ with regard to writing quality (i.e., human rater scores)?

Are there user-in-the-wild type differences related to writing revision in WM?

# Writing data selection criteria

## Middle/High School (MHS) Data Set (N=1,857)

- WM Extended Writing mode submissions (~2.5 years)
- Submission header required:  
*Mr, Mrs, Ms, or Miss,*  
(suggesting *teacher name* )

## Selection Criterion for User Behavior Modeling (N=861)

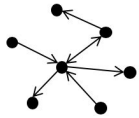
- Evidence of at least one WM function in event log
- Document had at least two submissions
- Submission  $\geq$  10 words



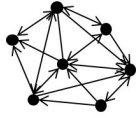
# Modeling Features & Analysis Measures

Modeling: WM Event Log Features & Social Network Analysis Modeling Features  
Analysis: Automated Writing Evaluation Measures, Human Essay Scores, & Revision Measures

# User behavior modeling: Process & social network analysis features



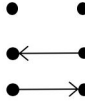
Sparse Network



Dense Network



Reciprocal Link



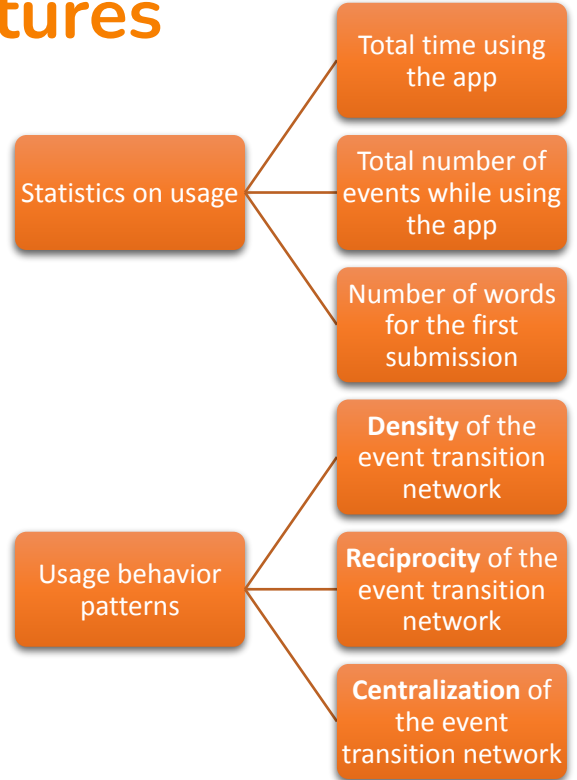
Non-Reciprocal Link



High Centralization Network



Low Centralization Network



# Human Rater Essay Score Features

**Data:** 670 first & last “pairs” of MHS essay submissions

**Human Raters:** 4 raters assigned 3 trait scores to each essay

**Rubric:** Purpose (0-4), Development (0-4) & Conventions scores (0-3)

# Revision (Editing) Features

## Revision Detection

- Identifies editing between pairs of submissions for a single document (e.g., *first-to-last*)
- Detects *Insertion*, *Deletion* & *Substitution* edit types

## Revision Types

- Total # of edits ( $\# \text{ of insertions} + \# \text{ of deletions} + \# \text{ of substitutions}$ )
- Total # of edit types: (a) # of insertions, (b) # of deletions, & (c) # of substitutions
- Normalized edit types: (a) # of insertions/# of edits, (b) # of deletions/# of edits, & (c) # of substitutions/# of edits
- # of edit types with *one-word revision* (e.g., “teh” -> “the”)
- # of edit types with *multi-word revision* (e.g., “The vase is...” – “A red vase is”)

# Findings

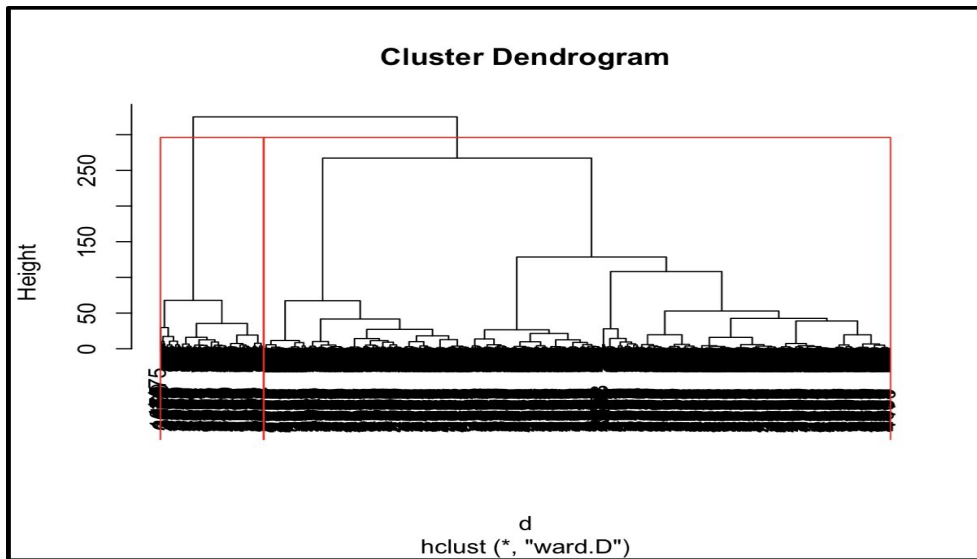
User Types, Patterns of Use, Writing Change, & Writing Proficiency

# User Types

Can we identify user-in-the-wild types from WM patterns of use from a “proxy” middle/high school (MHS) data?

# Dendrogram from Hierarchical Clustering Analysis

- Hierarchical Clustering using *Ward's* minimum variance method
- Euclidean distance



CLUSTER	N	TOTAL TIME	EVENT COUNT	WORD COUN T	DENSITY	RECIPROCITY	CENTRALIZATION
Cluster 1 (Average User Document)	739	393.53	15.52	615.65	0.01	0.14	0.08
Cluster 2 (Serious User Document)	122	1575.66	49.70	553.40	0.02	0.22	0.1
Total/Average	861	561.03	20.37	606.83	0.01	0.15	0.09

Cluster Comparisons for 6 Features Used in the Cluster Analysis 



# User Types & Writing Change

1. What relationships exist between WM MHS user types and changes in writing construct measures as captured by AWE features?
2. Are there user-in-the-wild type differences related to writing revision in WM?



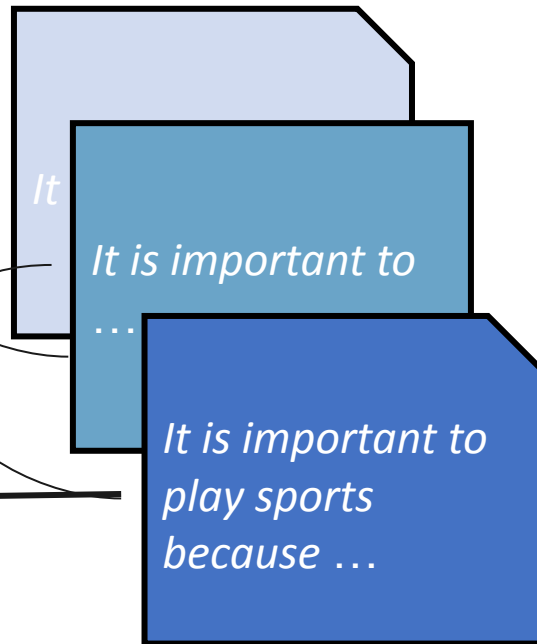
Should Kids Play Sports



File Edit View Insert Format Tools Adc

### NLP Tools can identify writing features

1. Grammar errors counts
2. Complex sentences
3. Vocabulary usage
4. Transition term use



19	6	1	0	210	0.55238095	0.09047619
11	0	0	0	52	0.80769231	0.21153846
103	38	1	1	1663	0.30727601	0.06193626

# Automated writing evaluation features

## ARGUMENTATION

- Features related to argumentation, e.g., presence of claims and sources

## ORGANIZATION & DEVELOPMENT

- Features related to organization & development, e.g., presence of discourse units (such as thesis), discourse coherence quality measure

## ENGLISH CONVENTIONS

- Features related to knowledge of conventions, e.g., grammar errors, use of contractions

## UTILITY VALUE

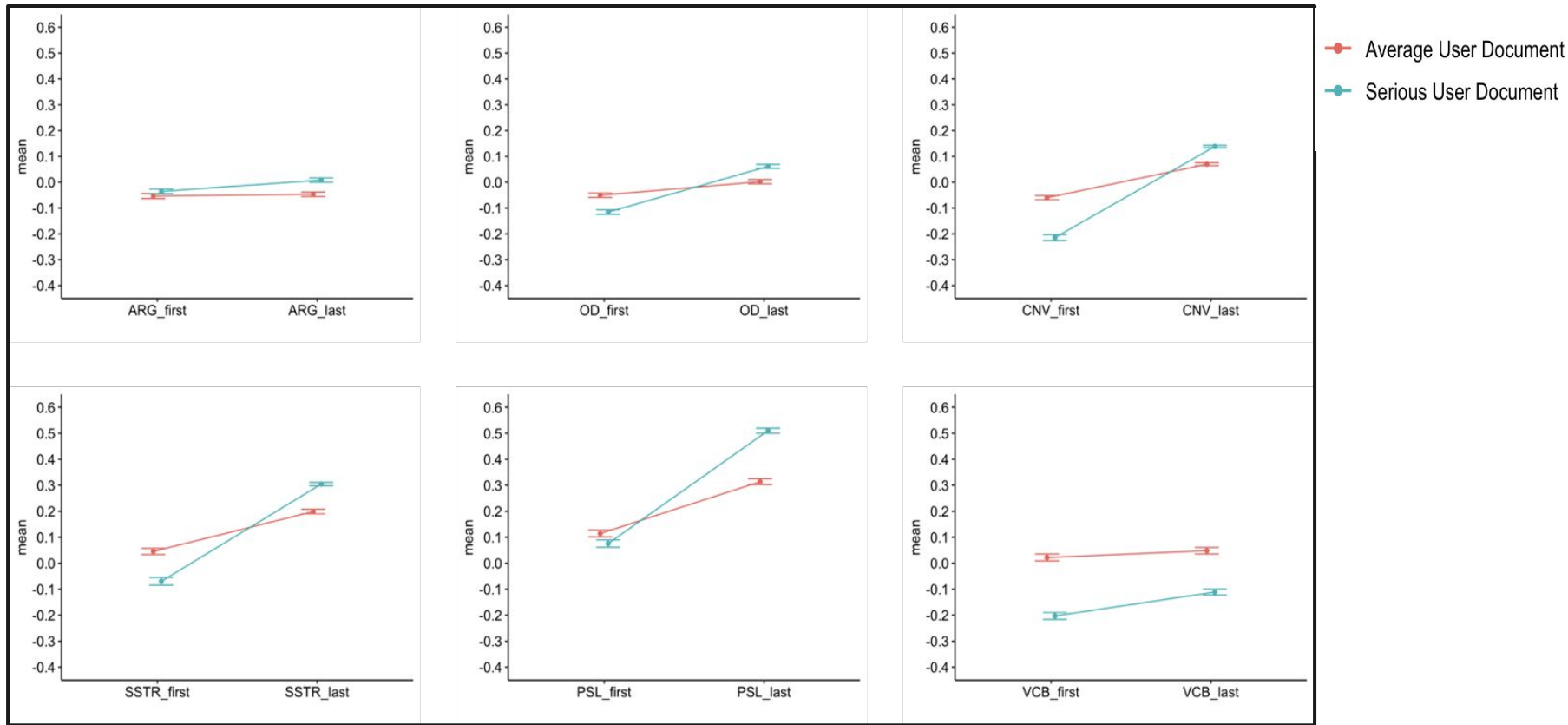
- Features related to personal reflection, e.g., use of 1<sup>st</sup> person pronouns

## SENTENCE STRUCTURE

- Features related to phrasal and sentence structure, e.g., sentence variety, complex noun phrases

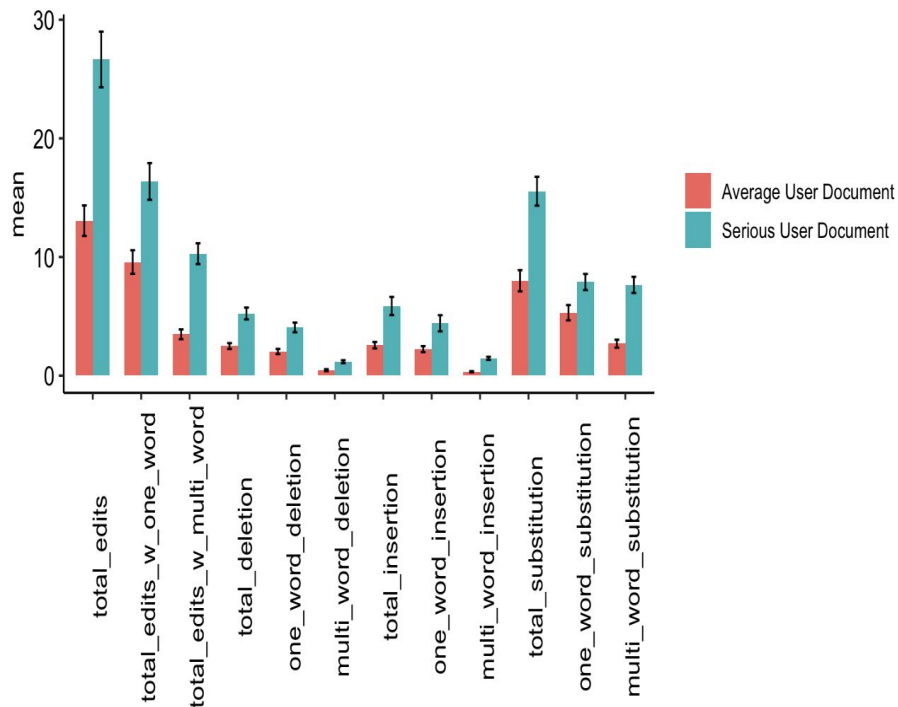
## VOCABULARY

- Features related to vocabulary usage, e.g., metaphor use, rare words



Statistically significant differences for *Serious Users* on the AWE features (at .05 level) were found for OD, CNV, SSTR, and PSL, but not for ARG and VCB.

# Revision & User Type



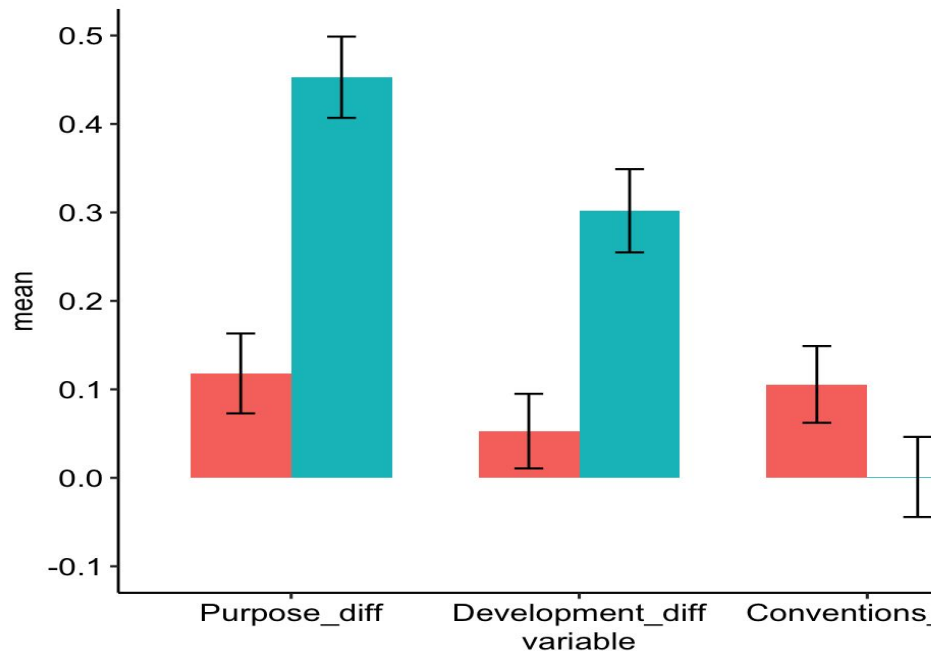
- Measured by differences between *first* and *last* submissions
- *Serious* Users tend to make more changes, overall and across edit types
- Significant differences on the *edit type* features between two clusters for
  - *Total\_edits* ( $t(50) = 2.29, p = .03$ )
  - *Total\_deletion* ( $t(49) = 2.18, p = .03$ )
  - *Total\_substitution* ( $t(55) = 2.40, p = .02$ )

# User Type & Writing Quality

How does WM MHS user type performance differ with regard to writing quality (i.e., human rater scores)?

# Trait Score Mean Differences Comparisons

- There were significant differences on the score differences between the two clusters for **Purpose** ( $t(70) = 2.55, p = .01$ ), and marginally significant for **Development** ( $t(66) = 1.87, p = .06$ ); but not for **Conventions** ( $t(69) = 0.82, p = .41$ ).
- # of Average User Document = 322 (red)
- # of Serious User Document = 53 (blue)
- Total N users= 375



# What did we learn about middle/high school Writing Mentor users from the writing analytics study?

- Two user types: *Serious* & *Average*
- User Types & Writing Quality
  - User types associated with differential gains in writing quality
    - i.e., *Serious* Users associated with greater gains
- User Types & Writing Revision
  - Differences in how *Serious* Users navigate and revise in WM as compared to *Average* Users.
- Implications for system development & personalized learning for writing



# Exploratory Writing Achievement & College Retention

Foundational Research & Implications for College  
Retention Analytics (Funded by the US Dept of Education,  
IES)



# Student Writing, Writing Attitudes & Broader Success

## Outcomes

### Motivation

- Low U.S. 4-year college retention is a national concern
- Across US institutions, ~65% of first-time students in 2012 graduated in 6 years
- Writing skills are critical to college success
- Nearly 75% of grade 8 and 12 students are below *Proficient* on NAEP writing
- No literature explicitly examines relationship between writing achievement and university retention

### Goals

- Explore the relationship between writing and retention in university using AWE to assess writing
- Identify writing achievement factors that may be for analytics to identify students at risk for dropout
- Generate hypotheses for future studies about essential writing skills for postsecondary success

# Research Questions

*What are the relationships between writing attitudes and AWE features, & college success factors?*

What are the relationships between student writing data and indicators of college retention (survival)?

# Intrapersonal factors & success

*What are the relationships between writing attitudes and AWE features, & college success factors?*

# Writing Attitudes & College Success

## Data

- Response data to a 50-item writing attitudes survey
  - *Writing Goals (Mastery, Performance, Avoidance)*
  - *Writing Confidence*
  - *Writing Beliefs (Importance of Content & Conventions)*
  - *Writing Affect (Feelings)*
- Institutional outcomes data: Cumulative GPA

## Participants

566 students from six 4-year institutions

# Writing Confidence & Success Indicators

Higher Cumulative GPAs are significantly correlated with writing confidence

- \* $p < 0.05$
- \*\* $p < 0.01$

Academic Performance Indicators	n	Confidence
<i>Academic Measures</i>		
Study Semester Cumulative GPA	563	.043
Study Semester Plus One Semester Cumulative GPA	526	.069
<b>Study Semester Plus Two Semesters Cumulative GPA</b>	<b>456</b>	<b>.138**</b>
<b>Study Semester Plus Three Semesters Cumulative GPA</b>	<b>364</b>	<b>.117*</b>
Study Semester Plus Four Semesters Cumulative GPA	215	.101

# AWE Features & Attitudes

## Data

- 997 coursework writing assignments
- 366 students completed standardized writing assessment
- AWE Feature Composite Scores

## Participants

418 students from six 4-year institutions

## Intercorrelations Between Writing Motivation Subconstructs & AWE Components

AWE Features	Mastery	Avoidance	Confidence	Content (Belief)	Conventions (Belief)
Argumentation	-.00	-.05	.03	.00	-.03
Organization and Development	-.02	-.07	.00	.00	-.07
Personal Reflection	.012	-.06	-.01	<b>.12*</b>	-.023
Sentence Structure	-.04	.03	-.05	-.02	.07
Vocabulary	.00	<b>-.12*</b>	<b>.15**</b>	-.04	<b>-.12*</b>
Convention	<b>.15**</b>	<b>-.14*</b>	<b>.18**</b>	<b>.15**</b>	<b>-.12*</b>

\*  $p < .05$ .

\*\*  $p < .01$ .



# AWE & College Retention

*What are the relationships between student writing data and indicators of college retention (survival)?*

# AWE Features and College Retention

## Data

- 997 coursework writing assignments
- 366 students completed standardized argumentative writing assessment
- AWE Feature Composite Scores
- Institutional data: Enrollment status between 3 - 5 semesters after study semester

## Participants

418 students from six 4-year institutions

# Survival Analysis

## Analysis

Random effects Cox  
proportional hazards  
regression used to model  
dropout as a function of the  
AWE feature composite score

## Results

Suggests *vocabulary usage* might  
identify students *at risk* of  
dropping out

- ***Personal Reflection Language* feature** predicted increased dropout for coursework ( $p < .05$ ) and assessment ( $p < .10$ ) writing
- ***Vocabulary* feature** predicted decreased dropout ( $p < .10$ ) in the standardized assessment

# What did we learn about relationships between writing achievement & college retention?

- **New Insights**

- Student writing attitudes and writing characteristics
- Student retention and writing characteristics

- **Practical Implications**

- *Writing Attitudes & Success*
  - Stakeholder understanding of student writing attitudes could inform individualized instruction and support for students early on
- *AWE & Retention*
  - AWE integration into a learning management system could provide *not only* personalized learning feedback for writing, but also retention analytics for stakeholders

# Data Exploration Activities

# IES writing achievement study data

<https://github.com/EducationalTestingService/ies-writing-achievement-study-data>

# Data Exploration Activity

<b>signs</b>	Inform/Explore	43
<b>try</b>	Inform/Explore	111
<b>call</b>	Inform/Explore	61
<b>national</b>	Inform/Explore	101
<b>hotline</b>	Inform/Explore	18
<b>sloppy</b>	Persuade	190
<b>people</b>	Persuade	2017
<b>what</b>	Persuade	925
<b>is</b>	Persuade	4949
<b>your</b>	Persuade	388

## Research Questions:

What can we learn about vocabulary usage by writing genre? And, by subject area?

## Data

1. Coursework writing assignment plain text documents
2. Human-assigned genre labels: a) inform/explore, b) persuade, c) reflect

## Data Preparation

1. Word-tokenize text (identify all words)
2. For all documents in a genre, provide word counts in.csv for analysis
3. Create visualizations to examine word use by genre and subject area.

## Example Jupyter Notebook for Analysis

1. Open [Anaconda](#) to access Jupyter Notebook
2. Create New or Open Existing Notebooks

**Inform/Explore**



Persuade

[illegible]

**Inform/Explore:**

# Biology



**Inform/Explore:**

# **Business**

# Rough Insights?

Writing in different genres might offer students opportunities to...

Writing in different subject areas might offer students opportunities to ...

# Moving forward: Tools and Resources

What tools are needed? What are some available tools? Moving forward!

# Which tools do I need?

- 1) What are my **writing analytics** research questions?
- 2) What **kinds of data** do I have?
  - a) Continuous numerical (number values, e.g., GPA)
  - b) Categorical (names or classifications, e.g., undergraduate vs. rising\_senior, novice vs. expert)
- 3) What **kinds of analyses** do I need to conduct?
- 4) How can I create **visualizations** to illustrate my findings?

# Example Data Exploration & Visualization Tools

## No programming

- Microsoft Excel or Google Sheets -- Tools for numerical and categorical data analysis & visualizations
- TAACO - Cohesion Indices (Crossley et al, 2016)
- TAALES - Lexical Sophistication (Kyle & Crossley, 2015)

## Programming

- Python libraries
  - Pandas
    - Open source data analysis and manipulation tool, built on top of the Python programming language.
  - Visualization libraries
    - Matplotlib, Seaborn, Plotly
- R language
- Jupyter Notebook (Python & R)
  - Interactive data science environment
  - Easy access through Anaconda



ing some text:

## Suggested Resources

[Python for Everybody  
Specialization - Coursera](#)

[Data Visualization with  
Python - Coursera](#)

[Natural Language Tool Kit  
\(NLTK\)](#)

[spaCy](#)

```
"""At eight o'clock on Thursday mo  
t feel very good."""
```

```
nk.word_tokenize(sentence)
```

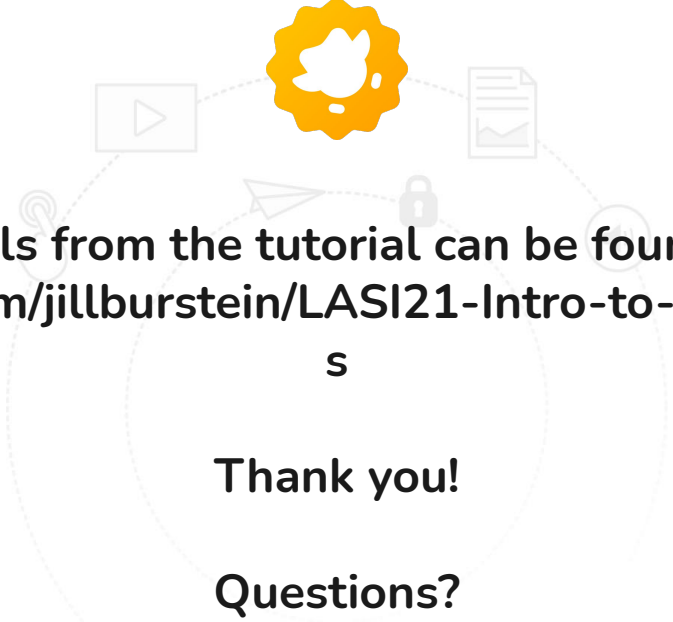
```
"o'clock", 'on', 'Thursday', 'mor  
"n't", 'feel', 'very', 'good', '
```

```
nk.pos_tag(tokens)
```

```
('eight', 'CD'), ('o'clock', 'JJ')  
IP'), ('morning', 'NN')]
```

# References

- Burstein, J. Elliot, N., Beigman Klebanov, B., Madnani, N., Napolitano, D., Schwartz, M., Houghton, P. & Molloy, H. (2018). [Writing Mentor: Writing Progress Using Self-Regulated Writing Support](#). *Journal of Writing Analytics*. Vol. 2: 285-313.
- Ling, Guangming, Elliot, N., Burstein, J. C., McCaffrey, D. F., MacArthur, C. A., Holtzman, S. (2021). [Writing motivation: A validation study of self-judgement and performance](#). *Assessing Writing*, 48.
- McCaffrey, D., Holtzman, S., Burstein, J., and Beigman Klebanov, B. (to appear). What can we learn about college retention from student writing? In *Companion Proceedings in the 11<sup>th</sup> International Conference on Learning Analytics & Knowledge (LAK21)*.
- Zhu, M., Zhang, M., & Deane, P. (2019). Analysis of keystroke sequences in writing logs. *ETS Research Report Series*, 2019(1), 1-16.
- Zhu, M., Burstein, J., Hao, J., & Livne, O. (2021) Uncovering Patterns of Use in the Writing Mentor® App through Cluster Analysis, Presented at the Advancing Digital Instruction and Assessment with Natural Language Processing & Learning Analytics, Coordinated Symposium to be held virtually at the 2021 Annual Meeting of the National Council on Measurement in Education (NCME), June 2021.



Materials from the tutorial can be found here:  
<https://github.com/jillburstein/LASI21-Intro-to-Writing-Analytic>  
s

Thank you!

Questions?

[jill@duolingo.com](mailto:jill@duolingo.com)

