

The Language Muse™ Activity Palette: NLP-guided Curriculum Support for English Language Learners

Jill Burstein

Educational Testing Service

Joint workshop on NLP for Computer-Assisted
Language Learning & NLP for Language Acquisition
Umeå, Sweden

November 16, 2016

Acknowledgements

- Research presented in this paper was supported by ETS, and the Institute of Education Science, U.S. Department of Education, Award Number R305A140472. *Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the views of the IES or ETS.*
- ETS Research Team: S. Andreyev, Kietha Biggers, Kelsey Drier, Nitin Madnani, Dan McCaffrey & John Sabatini

Today's Discussion

- Motivation for language technology to support classroom teachers
- *Language Muse™ Activity Palette ("the Palette")*
 - Pedagogical alignment
 - Demo
- Small-Scale Pilot Instructional Intervention Study
- Future Work: Large-Scale Instructional Intervention Pilot Study
- Take-away: Contributions and Potential of NLP for Educational Technology = Impactful NLP

Motivation: U.S. K12 Literacy Issues

- English language learner (ELL) population
 - 4.5M ELLs enrolled in K12
 - 9.3% participating in ELL programs
- Common Core State Standards
 - Students required to understand increasingly more complex text with difficult English constructions
 - *Figurative language, complex sentences, discourse relations*
- Content-area teachers support reading, writing & language skills

Similar Literacy Issues in Europe

- **European Literacy Policy Network (ELINET)**
 - Collaboration between researchers, practitioners, professionals and volunteers in 28 countries
 - *"In Europe one in five 15-year-olds ... lack basic literacy skills."*
 - "Framework of Good Practice"
 - Children, Adolescents, and Adults
 - Focus on reading and writing
 - The Netherlands' "Reading and Writing Foundation"
 - *"Healthier Through Literacy"*
- **Common European Framework for Language Learning, Teaching, and Assessment**
 - 6 levels of proficiency and related assessments
 - Learning a foreign language for residents / promote multilingualism
- 5 | – Refugee foreign language learning

NLP Technology Solution

- Provide classroom activities centered around linguistic elements in content-area texts
- *The Palette*, a web-based tool
 - Various NLP algorithms used to automatically generate customizable language activities
 - Activities generated are grounded in language learning research

The *Language Muse Activity Palette* “The Palette”: Reading & Language Skills Development

“The *Palette*” Motivation & Pedagogy

- Motivation
 - U.S. K12 classrooms can have high % of English learners (ELs)
 - Content-area teachers are responsible for EL learning, but not sufficiently trained in reading or language learning
- Pedagogy-informed
 - Leverages empirical findings from 2nd language learning literature
 - Collects practitioner/educator knowledge to inform language activity types/design

The *Palette*: How it works

- Relies on **NLP backend** to identify different linguistic features (lexical entities, syntactic structures, rhetorical & discourse relations.)
- Upload classroom texts ➔ Generate activities for various aspects of content comprehension: vocabulary, grammatical structures, and discourse & text organization
- Administers activities to students
- Provides automated scoring and analytics for teachers based on student work

ELL Difficulties Informed Palette Activities

- Vocabulary: lack of familiarity with academic vocabulary, morphology, unfamiliar words
- Sentence-Based: complex sentences
- Phrase-Based: complex phrasal structures (NPs, VPs, PPs)
- Discourse-Based: discourse term meaning/function in text

NLP, Lexical Resources & Activity Types

Lexical Activity	NLP/ Lexical Resources
Academic Words	Coxhead (2000) Academic Word List + Wordnik API to generate definitions (https://www.wordnik.com/)
Key Concepts	Heuristic to identify frequent “key” words across paragraphs
Multiword Expressions	Rank-ratio-based collocation detection algorithm trained on Google Web1T n-gram corpus (Futagi et al, 2008)
Contractions	Regular expressions defined on constituency parses
Morphologically -complex Words	Rule-based morphological parser (discussed in Leacock & Chodorow, 2003)
Morphological Variants	ETS proprietary resource that identifies inflectional & derivational variants
Synonyms	WordNet (Fellbaum, 1998) + a Lin thesaurus (Lin, 1998) + SMT-based paraphrases (Bannard and Callison-Burch, 2005)

Lexical Activity	Lexical Resources
Cognates	<p>Manually-created dictionary of Spanish-English word pairs e.g., <i>retired/retirar, person/persona</i></p>
Antonyms	Wordnet (Fellbaum, 1998)
Homonyms	<p>Manually-created list (Burstein et al., 2004) e.g., <i>their, there, they're</i></p>

Sentence Activity	NLP: Regular expression on constituency parser
-------------------	--

Relative Clauses	<i>Fish are rich in protein and nutrients that</i> contribute to good health.
Multi-Clause Sentences	<i>The half of the moon that faces the sun is always illuminated</i> , but we see both the light and dark sides in different proportions as it orbits the Earth.
Complex Noun Phrases	<i>It is during the rapid eye movement stage</i> that most dreams occur.
Complex Verb Phrases	<i>During this time, women were never supposed to travel alone.</i>

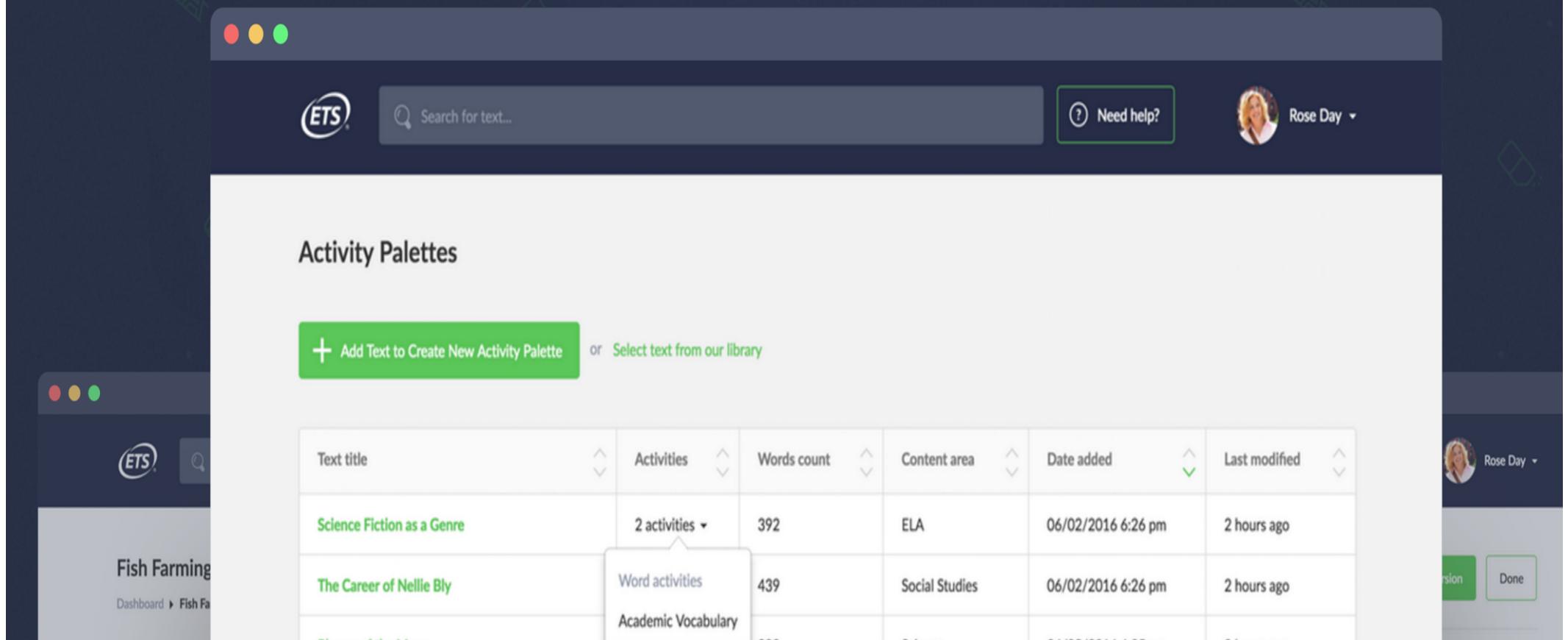
Discourse Activities	NLP/Lexical Resources
Compare-Contrast	Discourse term lexicon indicating a comparison or contrast between text segments, e.g., <i>"He was a wise and patient leader; <u>however</u>, his son . . ."</i>
Cause-Effect	Discourse term lexicon indicating a cause-effect relation between text segments, e.g., <i>"The discovery of fossils of tropical plants in Antarctica <u>led to</u> the hypothesis that . . ."</i>
<i>Pronoun Reference</i>	<i>Pronoun identification with maximum entropy tagger (MXPOST) part-of-speech tagger</i>

Palette Screenshots

Language support in content-area classrooms

The *Language Muse*™ engine generates activities intended to help English learners build the skills they need for deeper reading comprehension in Science, Social Studies and English Language Arts.

Get an invite



The screenshot shows a software application window titled "Activity Palettes". At the top, there is a toolbar with the ETS logo, a search bar, a "Need help?" button, and a user profile for "Rose Day". Below the toolbar, the main area is titled "Activity Palettes" and contains a green button with the text "Add Text to Create New Activity Palette" and the instruction "or Select text from our library". A table lists several activity palettes, including "Science Fiction as a Genre" and "The Career of Nellie Bly". The table columns are: Text title, Activities, Words count, Content area, Date added, and Last modified. The "Activities" column for "Science Fiction as a Genre" shows "2 activities". The "Activities" column for "The Career of Nellie Bly" shows "Word activities" and "Academic Vocabulary". The "Content area" for "The Career of Nellie Bly" is "Social Studies". The "Date added" and "Last modified" for both entries are "06/02/2016 6:26 pm" and "2 hours ago" respectively. The bottom of the window shows a navigation bar with "Dashboard" and "Fish Farming" buttons, and a status bar with "Done" and "Done" buttons.

Text title	Activities	Words count	Content area	Date added	Last modified
Science Fiction as a Genre	2 activities	392	ELA	06/02/2016 6:26 pm	2 hours ago
The Career of Nellie Bly	Word activities Academic Vocabulary	439	Social Studies	06/02/2016 6:26 pm	2 hours ago

Activity Palettes

 Add Text to Create New Activity Palette

or Select text from our library

Text title	Activities	Words count	Content area	Date added
Penny Debate	5 activities ▾	1097	Other	09/07/2016 2:05 pm
Penny	Sentence activities		Other	09/06/2016 2:04 pm
The Triangle Shirtwaist Factory Fire	Multiple Clause Sentences		Social Studies	06/30/2016 2:16 pm
Phases of the Moon	Finding Referential Terms in Relative Clauses		Science	04/09/2016 2:47 pm
The Career of Nellie Bly	Discourse Marker Vocabulary		Social Studies	04/01/2016 12:53 pm
Fish Farming	Paragraph activities		Science	03/29/2016 12:27 pm
Cave Paintings	Summary practice with paragraphs		Social Studies	12/14/2015 3:34 pm
Good Sleep for Good Health	Word activities		Science	10/29/2015 2:12 pm
	Academic Vocabulary		Other	09/25/2015 9:34 am
	 Add new activity			

Paragraph activities

Variant word forms in paragraphs 

Summary practice with paragraphs

Summary practice with paragraphs with key term hints

Synonyms in paragraphs 

Phrasal Verbs

Cognates (Spanish) 

Recommended activity: Contains 5 or more questions related to language skill 

Activity contains 1-4 questions related to language skill

Unable to generate activity related to language skill



Review sample activities from the list, then click **Create Activity** to generate activities based on your text

Sentence activities

Homonyms in sentences 

Activity

 Hide all sentences

Sentence 1



Hide

In the early 1900s, a time known as America's Progressive Era, 146 people **were killed** in the Triangle Shirtwaist Factory Fire in New York City.

Correct answer

- Now  Hide
- Past  Hide
- Future  Hide

[Add new answer](#)



Create another variant of sentence

Sentence 2



Hide

All but six bodies **were identified** in the days that followed the fire of March 25, 1911.

Directions

In each sentence below, decide whether the event is happening now, happened in the past, will happen in the future, or may happen in the future.

Activity

Sentence 1

In the early 1900s, a time known as America's Progressive Era, 146 people **were killed** in the Triangle Shirtwaist Factory Fire in New York City.

- Now
- Past
- Future

Sentence 2

All but six bodies **were identified** in the days that followed the fire of March 25, 1911.

- Now
- Past

1 In the early 1900s, a time known as America's Progressive Era, 146 **people** were killed in the Triangle Shirtwaist Factory Fire in New York City.

Put the correct answer in the box

peoples peopling peopled



2 Jewish and Italian immigrant girls and women made up the majority of those , with just 23 boys and men making up the rest.

Put the correct answer in the box

killing killed kills kill

3 Approximately half teenagers, and some were as young as ten years of age.

Synonyms in Isolation

Match the words with similar meanings.

 Hide instructions

1

America

US

Progressive

liberal

killed

dead

Shirtwaist

blouse

Match synonyms from the box

2

Factory

women

made

majority

Match synonyms from the box

people

mainly

plant

get



Text

 Print text

In the early 1900s, a time known as America's Progressive Era, 146 people were killed in the Triangle Shirtwaist Factory Fire in New York City. Jewish and Italian immigrant girls and women made up the majority of those killed, with just 23 boys and men making up the rest. Approximately half were teenagers, and some were as young as ten years of age. All but six bodies were identified in the days that followed the fire of March 25, 1911. One hundred years later, a researcher named Michael Hirsch identified those six individuals. Every year since the fire, those who died have been remembered with the reading of their names to the tolling of a bell.

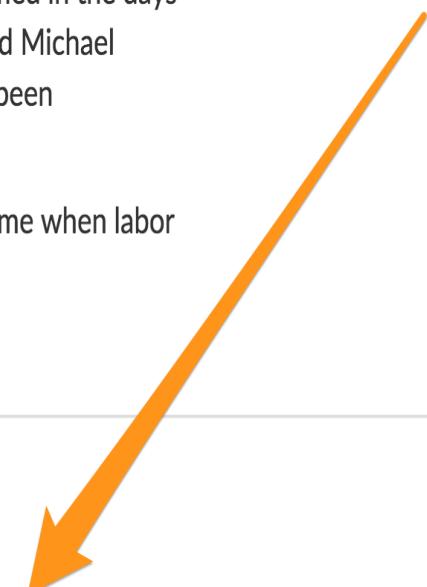
The Triangle Factory workers were sewing machine operators and cloth cutters at a time when labor unions were...

[Show more...](#)

Activities

 Add new

Paragraph activities



[Summary practice with paragraphs with key term hints](#)
Created: 8 days ago

[Cognates \(Spanish\)](#)
Created: 11 days ago

[Variant word forms in paragraphs](#)
Created: 12 days ago

[Synonyms in paragraphs](#)
Created: 2 months ago

[Variant word forms in paragraphs](#)
Created: 2 months ago

Last update: 06/30/2016 2:16 p

Description: [Add description](#)

Tags: 

[Edit text](#)



Library texts cannot be edited

[Palettes](#)[Assignments](#)[Classes](#)

Search...



Jill Burstein ▾

Classes

[New class](#)[Filters](#)[Show](#)[All accounts](#)**Mrs. B's Science Class**

Name	Student ID	Grade	Status
 J. Cassway	22345	6	Active account
 A. Burstein-Stern	12345	6	Active account
 A. O'Neill	12348	6	Active account
 A. Somasundaran	12349	6	Active account
 M. Burstein-Stern	12346	6	Active account
 M. Beigman	12356	6	Active account

[Add student](#)[Import file](#)[Print login credentials](#)

[Palettes](#)[Assignments](#)[Classes](#) Search...

Jill Burstein ▾

Assignments

[New assignment](#)

Filters Status Draft Active Archived

Assignment Name	Activities	Students	Due date	Created	Status	Type
Verb Practice #1: Working with Tenses	2	6 students assigned 2 turned in, 4 not turned in	11 days left (11/22/2016)	11/09/2016	active	online

[Show all](#)

Assignment Verb Practice #1: Working with Tenses

Assignments ▶ Assignment Verb Practice #1: Working with Tenses

Interactive preview

Actions ▾

Archive

Assignment Summary

Details	Activities	Turned in	Not turned in
Status	active	Verb Tense Practice	2
Type	online		4
Description	Helps student to work with tenses.		
Due Date	11/22/2016		
Sent			

6 students assigned

Name	Student ID	Grade	Submission
 J. Cassway	22345	6	11/09/2016
 A. Burstein-Stern	12345	6	Feedback received
 A. O'Neill	12348	6	-
 A. Somasundaran	12349	6	-
 M. Burstein-Stern	12346	6	-
 M. Beigman	12356	6	-

In each sentence below, decide whether the event is happening now, happened in the past, will happen in the future, or may happen in the future.

Overwrite score

1 Northern Lights **can be seen** in the northern or southern hemisphere, in an irregularly shaped oval centred over each magnetic pole.

- Now
- Past
- Future

2 The lights **are known** as 'Aurora borealis' in the north and 'Aurora australis' in the south.

- Now
- Past
- Future

3 Scientists **have learned** that in most instances northern and southern auroras are mirror-like images that occur at the same time, with similar shapes and colors.

- Now
- Past
- Future

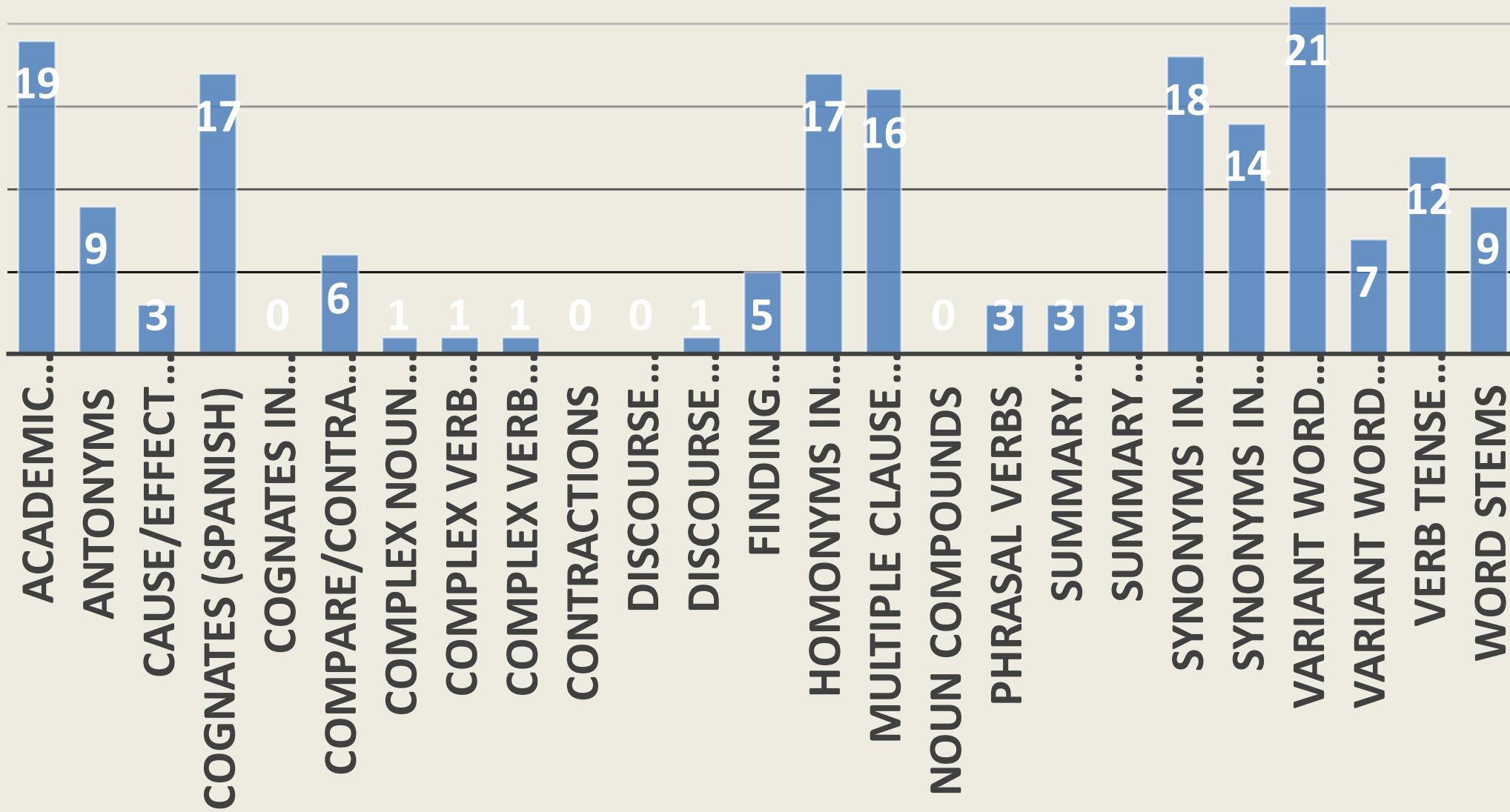
4 Because the phenomena occurs near the magnetic poles, northern lights **have been seen** as far south as New Orleans in the western hemisphere, while similar locations in the east never experience the mysterious lights.

- Now
- Past
- Future

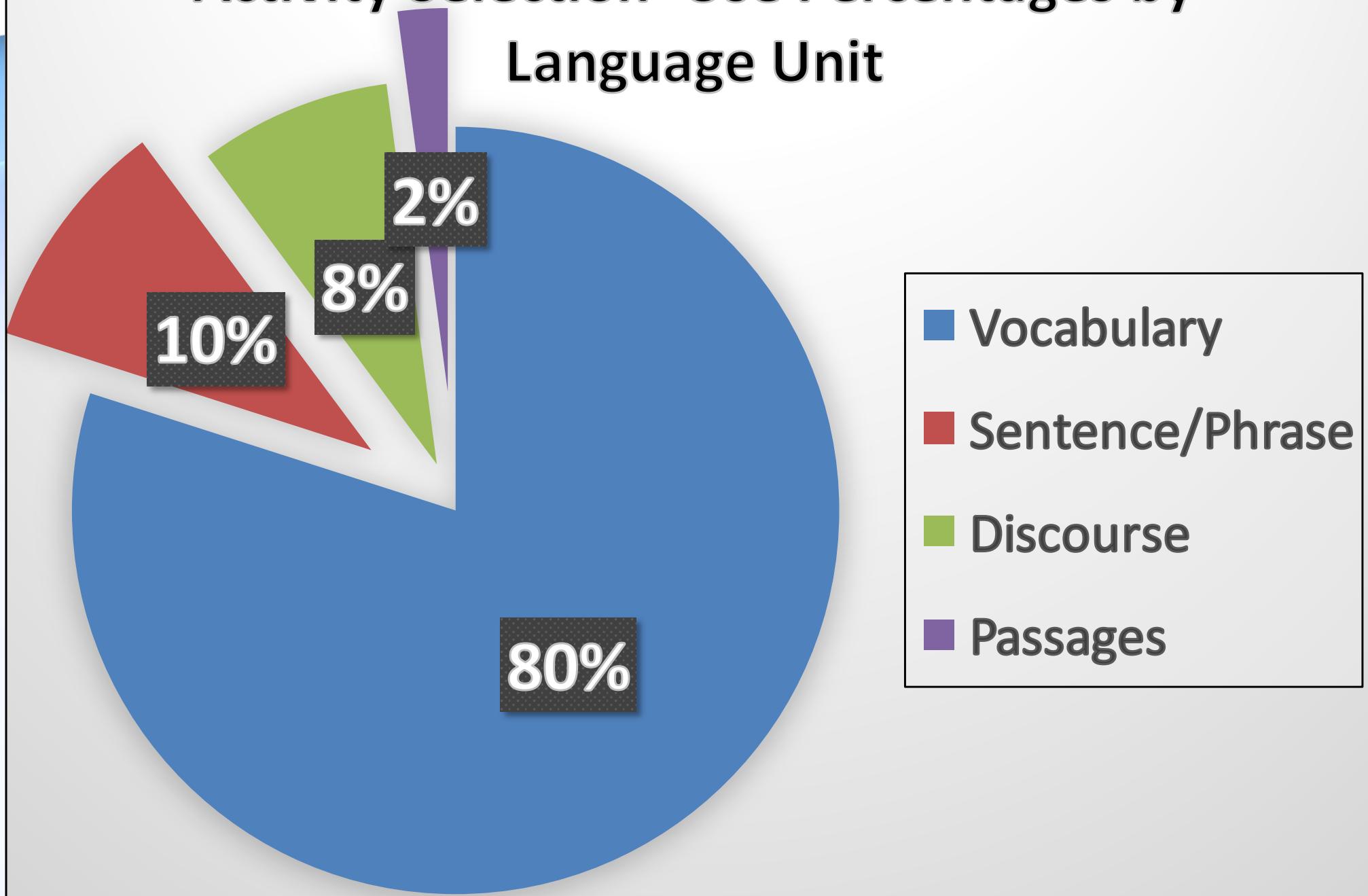
Spring 2016 Small-Scale Pilot Study

- 6-week instructional Intervention
 - Teachers integrated tool into curriculum
- 3 middle schools (2 in TX; 1 in NJ)
- 17 teachers
 - ELA
 - Social Studies
 - Science
- Teacher Perception Survey
- Pre-post test student reading and language assessment show **overall gains**
- Pre-, During, Post-Intervention Observations

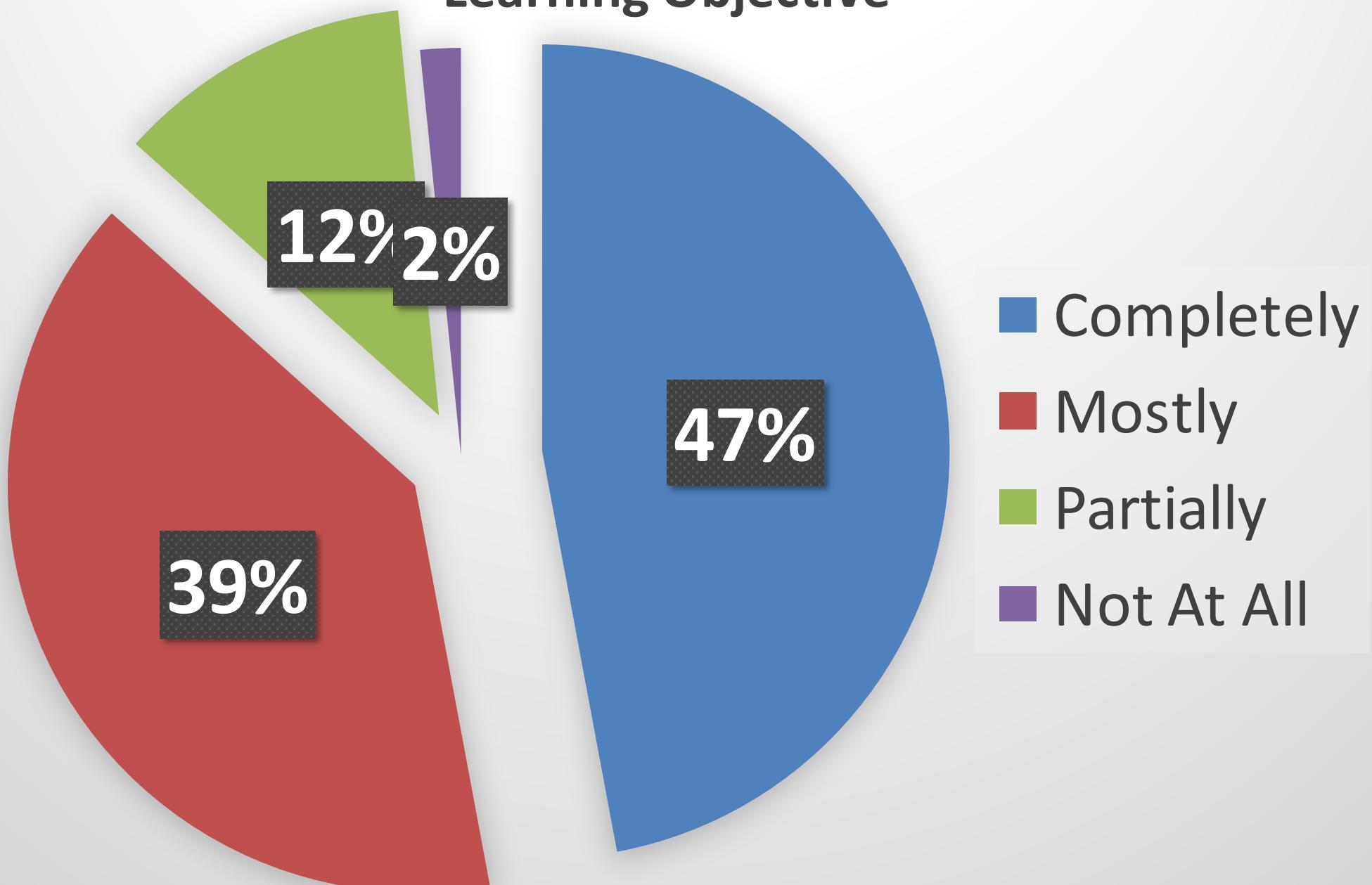
Activity Type Counts: Reported Use During Instructional Intervention



Activity Selection Use Percentages by Language Unit



Perception of Activity Fulfillment of Learning Objective



Pre-Post Test

- RISE Assessment
 - *Word Decoding*
 - *Vocabulary*
 - *Morphology*
 - *Sentence*
 - *Efficiency (*reading speed)*
 - *Reading Comprehension*
- Pre-test prior to classroom intervention
- *The Palette* classroom intervention
- Post-test after the intervention

Two Schools

- **School A**
 - Small private school in Texas
 - Primary first language of ELLs: Arabic
- **School B**
 - Large public middle school in New Jersey
 - Primary first language of ELLs: Spanish

RISE Scale Scores

- 250: Middle of scale.
- 270 or above: Proficient
- 230 or below: Warning level

Pre-Post Outcomes: Both Schools

Test	N	Mean, Pre	Mean, Post
DecWrd	169	255.21	253.47
Vocab	169	254.49	255.89
Morph	167	253.71	253.52
Sentences	63	251.21	247.6
Efficiency	59	250.31	252.8
RComp	41	246.5	247.11

School A vs School B

Test	Group	N	Mean, Pre	Mean, Post	Test	Group	N	Mean, Pre	Mean, Post
DecWrd	SchoolA, Grade 6	23	253.35	257.3	Sen	SchoolA, Grade 6	23	252.78	244.26
DecWrd	SchoolA, Grade 7	13	260	253.85		SchoolA, Grade 7	13	255.54	235.85
DecWrd	SchoolA, Grade 8	28	268.5	254.36		SchoolA, Grade 8	28	253.93	238.04
DecWrd	SchoolB, Grade 6	52	251.94	252.92		SchoolB, Grade 6	50	251.02	254.2
DecWrd	SchoolB, Grade 7	27	249.22	250.44		SchoolB, Grade 7	25	246.4	248.12
DecWrd	SchoolB, Grade 8	26	252.88	253.15		SchoolB, Grade 8	24	249.58	254

Motivation

Sample size across the tests

	DecWrd	Vocab	Morph	Sents	Effic	Rcomp
SchoolA, 6	23	23	23	23	23	23
SchoolA, 7	13	13	13	13	13	13
SchoolA, 8	28	28	28	28	28	28
SchoolB, 6	52	52	51	50	49	37
SchoolB, 7	27	27	26	25	24	23
SchoolB, 8	26	26	26	24	22	17

Observations: Palette in the Classroom

- **Content-area teachers**
 - Used *paragraph-level* activities
 - Focused on the entirety of the text by reading before doing the actual activity
 - Almost always used Palette activity as a *warm-up* in content area
- **English (ELA or ESL) teachers**
 - Used the activity as the *basis of the entire lesson*
 - Focused less on the text chosen
 - More likely to introduce the language feature and give detailed instructions about how to complete the activity
- **Activity Delivery: Most teachers used an interactive whiteboard**
 - Students responded on paper
 - Teachers would read aloud the question from white board
 - Students would respond at their desk, or walk up to the white board

Summary

- The Palette is a web-based, NLP-driven learning application that supports teachers and their ELLs
- A small-scale pilot suggests promise for classroom use
 - Positive gains at 6th grade in School A, but negative gains at 7th * 8th, strongly suggested that these are motivation based.
 - Significant, positive gains in School B across all three grades and across all six tests

Next Steps: Spring 2017 Randomized Control Large- Scale Instructional Intervention

- **Participants.** Schools. Approximately 20 middle school sites in districts in Connecticut, Georgia & New Jersey with higher ELL populations.
- Teachers. Expected 35 ESOL & content-area teachers across the districts/states.
- **Methods.** Teachers (1) receive 2-day professional development in ELL language & reading support, and ***the Palette version 2***, (2) administer a 6-week instructional intervention with the *Palette version 2* (3) are observed observed prior to, and during the intervention, & (4) participate in PLCs. Students will take a reading components proficiency (RISE) pre-posttest.

“Impactful NLP” Serving Language Learners

- US: Language learning resource need in U.S. K12
 - <http://www.dekalbschoolsga.org/english-learners-studies-program/>
 - *22K int'l students, 156 countries, 162 languages*
- MOOCs for Refugees (K12 – postsecondary)
 - edX (U.S.) + Kiron University (Germany)
 - Coursera + U.S. Department of State
 - Erasmus+ (*European Commission*)
 - <http://www.paper-airplanes.org/>
- “Palette” for languages other than English