Listening. Learning. Leading.®

NLP & the Literacy Challenge: Supporting Reading and Writing Proficiency

Jill Burstein
Educational Testing Service

The Seventeenth Annual Maryland Assessment Conference
November 2, 2017

Acknowledgements

Research presented in this paper was supported by ETS, and the Institute of Education Science, U.S. Department of Education, Award Numbers R305A140472 and R305A160115. *Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the views of the IES or ETS.*

Discussion

Social Challenge: Improve Literacy!

- Low literacy affects critical practical and social aspects of social participation
  - Employability
  - Social factors: Self-esteem & self-confidence
- Global issue

Body of Knowledge: NLP & Literacy Support

- History
- State-of-the-art

Motivation: How Can Innovative NLP Contribute to Improving Literacy?

- *Language Muse*™ Activity Palette
- Writing Mentor™ application, Google Docs Add-on
- NLP, Writing Analytics, and Success Indicators

Evidence of Global Literacy Challenge

U.S. K12

- *NCES Nation's Report Card (2011,2015)*
  - **Writing:**73% -- *Below Basic* (21%) or *Basic* (52%) Proficiency in 12th Grade (2011)
  - **Reading:** 63% -- *Below Basic* (28%) or *Basic* (35%) Proficiency in 12th Grade (2015)
  - *~51% increase in ELLs in K12 since 1997-98.*

U.S. 4-year post-secondary institutions (Complete College America, 2012)

- More than 20% placed in developmental courses
- 1/3 or fewer of students in remediation graduate in 6 years

***European Union**: (EU Commission "High Level Group Report ", 2012)*

- "One in five 15-year-olds in the EU still has insufficient reading skills." (PISA 2012 findings)
- "In 2011, across Europe ~73M low-educated adults …many of whom …have literacy problems .."

Literacy Challenge Factors & Education Policy in U.S.

In K-12

- ELL populations
  - In 2013-14
    - ~4.5M ELLs enrolled in K12
    - 9.3% participating in ELL programs
- Under Common Core State Standards, **content-area teachers** support reading, writing & language skills

In 2- & 4-year postsecondary institutions

- Roughly *18 million students enrolled* in postsecondary education
- About *1.7M lack prerequisite skills* to complete college
- Projected enrollment increases over the next 10 years

NLP & Social Challenges

Biomedical Informatics

Author profiling in health forum analysis

Mental Health/Clinical Psych

Violence risk, suicide assessment from text

Negative Societal Issues

Analysis of web ads soliciting sex

"Bad Policing"

Education

Automated writing evaluation for instruction & assessment

How can NLP support the literacy challenge? History, Status Quo, and What's New

Project Essay Grade (PEG) (Page, 1966) for essay scoring for classroom writing assignments

- Transformation of essay length
- Some syntactic analysis

Writer's Workbench (Cherry et al, 1982) for editing support

- Editing tool for students
- Diction, style, spelling
- Detection of topic sentences (discourse)

Intelligent Essay Assessor™ (Landauer, et al 1998) for large-scale high-stakes exams

- Vocabulary usage (Latent Semantic Analy
- Style, mechanics measures

E-rater® (Burstein et al, 1998; Attali & Burstein, 2006) for large-scale high-stakes exams

- Vocabulary usage
- English conventions
- Discourse, argument structure & coherence
- Sentence complexity

A lot has happened since the 90's

Reading

- **Readability measures**: 1) Flesch-Kincaid to compute document readability in MS Office's MS Word, 2) Lexile® is used to assign textbook readability by large publishers (such as, McGraw-Hill Education)
- **Language and reading skills development** with automated item generation: *Language Muse*™ Activity Palette, (ETS)

Writing

- **Automated Writing Evaluation (AWE) feedback:** *Criterion*® (ETS), *Writing Mentor* (ETS), Write-to-Learn™(Pearson), Turnitin® Revision Assistant, Grammarly®
- **Peer Review:** SWoRD: MyReviewers, Turnitin Feedback Studio

Speaking

- Automated scoring of spontaneous speech: ETS' SpeechRater™
- Automated evaluation of multimodal input (e.g., video for interviews)
- Automated evaluation of speech for reading apps

NLP & the Literacy Challenge @ETS

Automated language activity generation

- The *Language Muse™ Activity Palette*

Automated writing evaluation

- Writing Mentor™ app, a Google Docs Add-on
- Writing analytics to examine college success

The *Language Muse Activity Palette:* Addressing Reading & Language Skills Development

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

Content-area teachers support reading, writing & language skills

ESL instructors need to tailor instruction to ELLs

NLP Technology Solution

What if we had technology for curriculum support to provide language-based classroom activities centered around linguistic elements in content-area texts?

The Language Muse Activity Palette

- NLP algorithms used to automatically generate customizable language activities
- Activities generated are grounded in language learning research
- Freely avaliable

"The *Palette*": Motivation & Pedagogy

Motivation

- U.S. K12 classrooms have high % of English learners (ELs)
- Content-area teachers responsible for EL learning, but not sufficiently trained
- ESL instructors can use support too

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

NLP backend identifies different linguistic features (lexical entities, syntactic structures, rhetorical & discourse relations.)

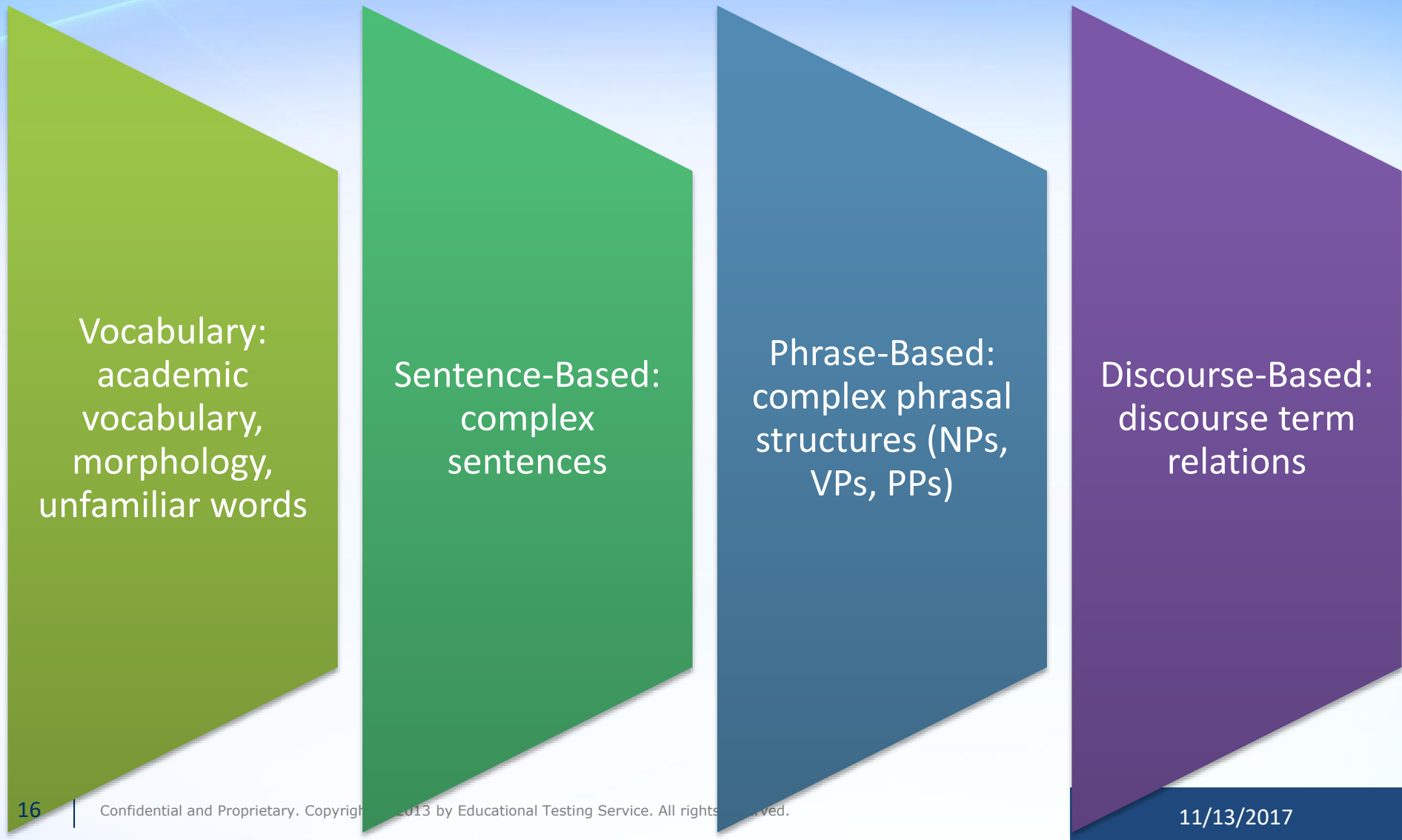
Instructors upload classroom texts

The *Palette* generates activities to support content comprehension: vocabulary, syntactic structures, and discourse & text organization

Instructor administers activities to students

The *Palette* provides automated scoring and analytics for teachers based on student work

Palette Activities Informed by ELL Difficulties



Quick Tour!

https://languagemuse.org/

SPRING 2017 RANDOMIZED CONTROL TRIAL INTERVENTION: YEAR 3/PART 1

Research Questions

What is the usability and feasibility of implementation by teachers in classroom settings? Specifically, can the teachers use the training, use the tools, and implement as intended?

How are ESL teachers using the Palette to support ELLs?

What are the student learning outcomes, especially with regard to English learner gains in content learning, and reading and language skills development?

Study Participants

School Sites

6th-12th grade classrooms across 3 high schools & 10 middle schools in CT, GA, & NJ

Students

185 Total:

78 Control; 107 Experimental

L1's: Spanish, Arabic, Haitian-Creole, Hindi/Urdu, Punjabi, Chinese & Portuguese.

Teachers

14 Total:

7 Control; 7 Experimental

2 ELA, 2 ELA/ESL, 10 ESL

Training & Data Collection

Two-day *Palette* training

Data Collection

Study Instruments

Palette Data Usage Logs

Teacher pre-background & tool post-perception survey

Student pre-post reading components assessment (RISE)

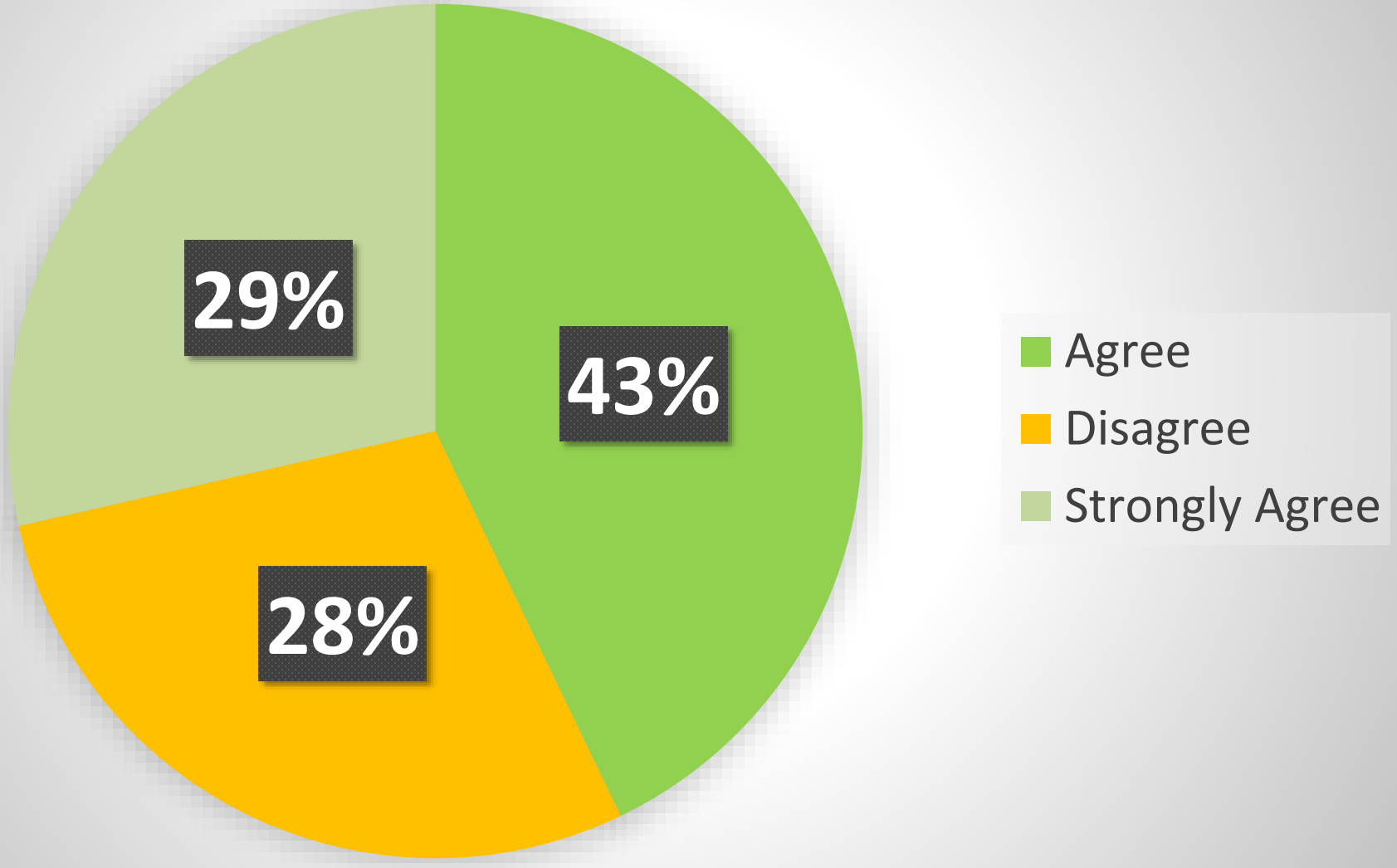
Classroom fidelity observations

Teacher reflection logs: Tool use details during intervention

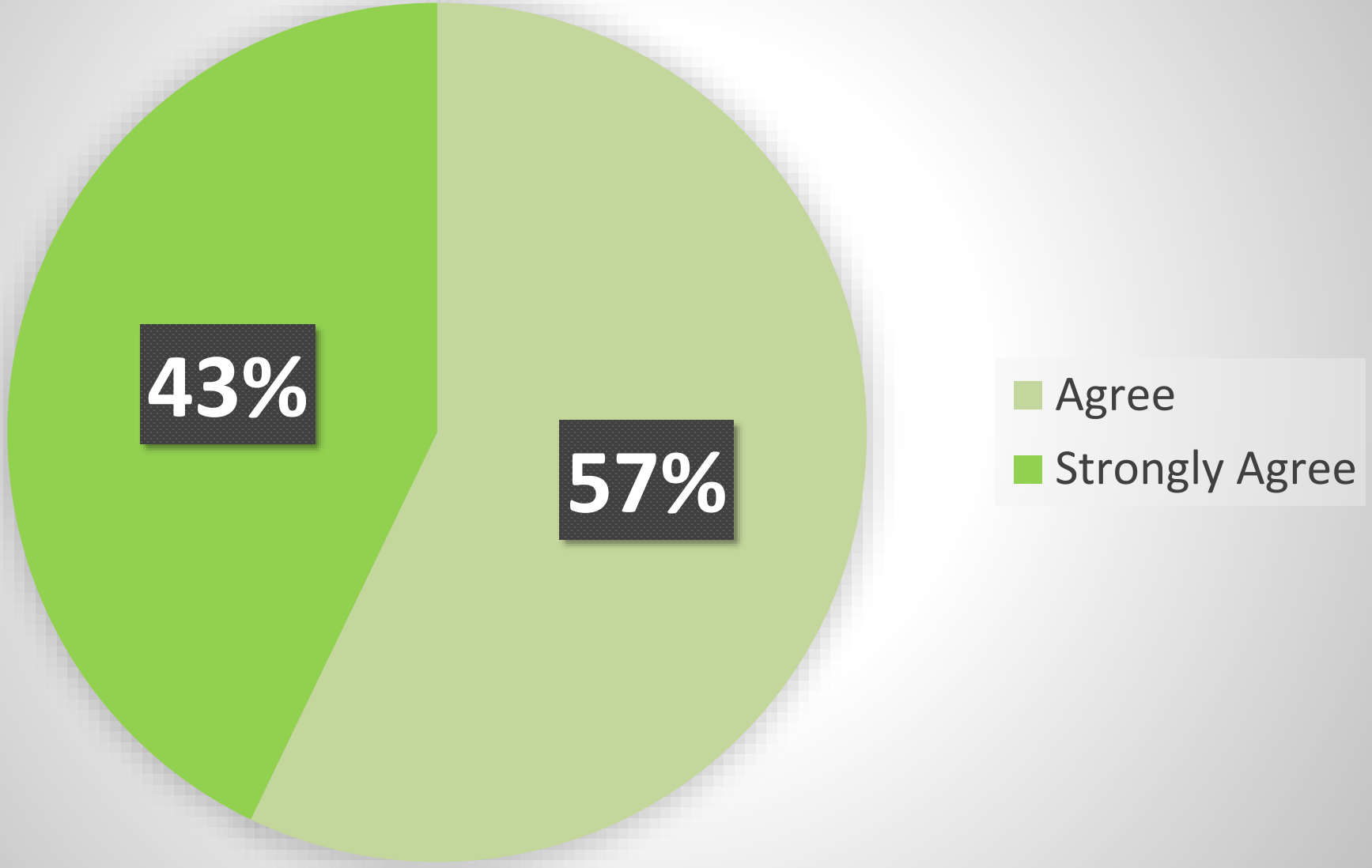
Teacher Perception Survey: Teacher Experience & Percieved Learning Support

TEACHER EXPERIENCE

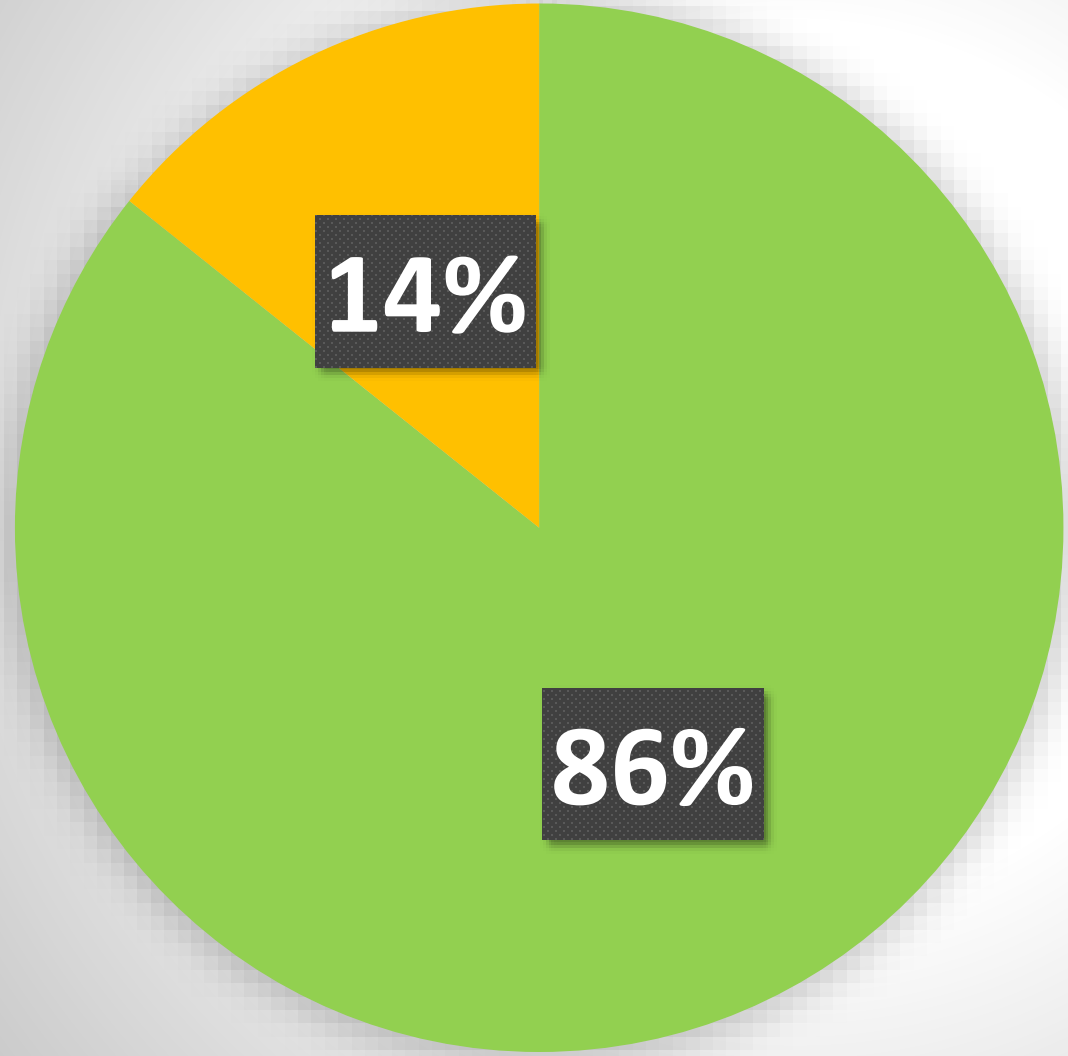
I enjoyed creating & customizing activities



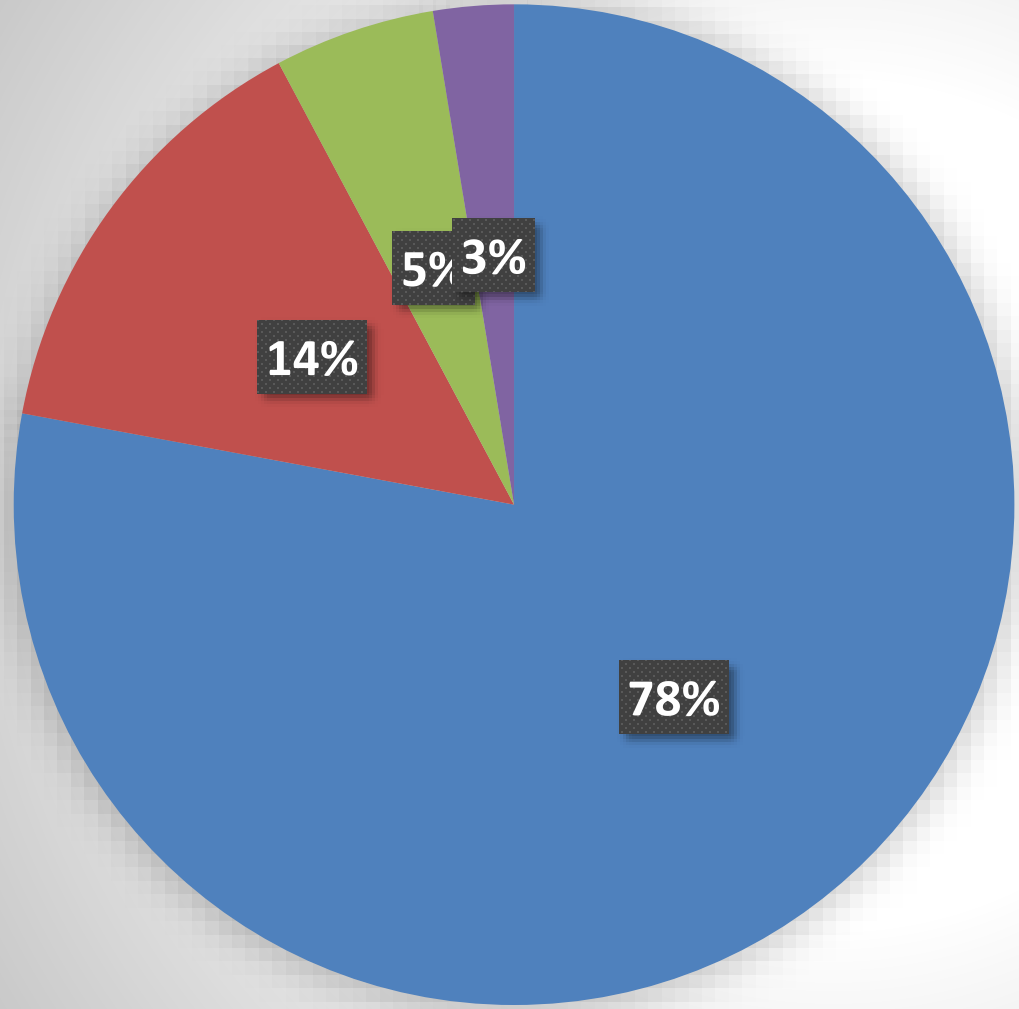
I felt confident administering activities



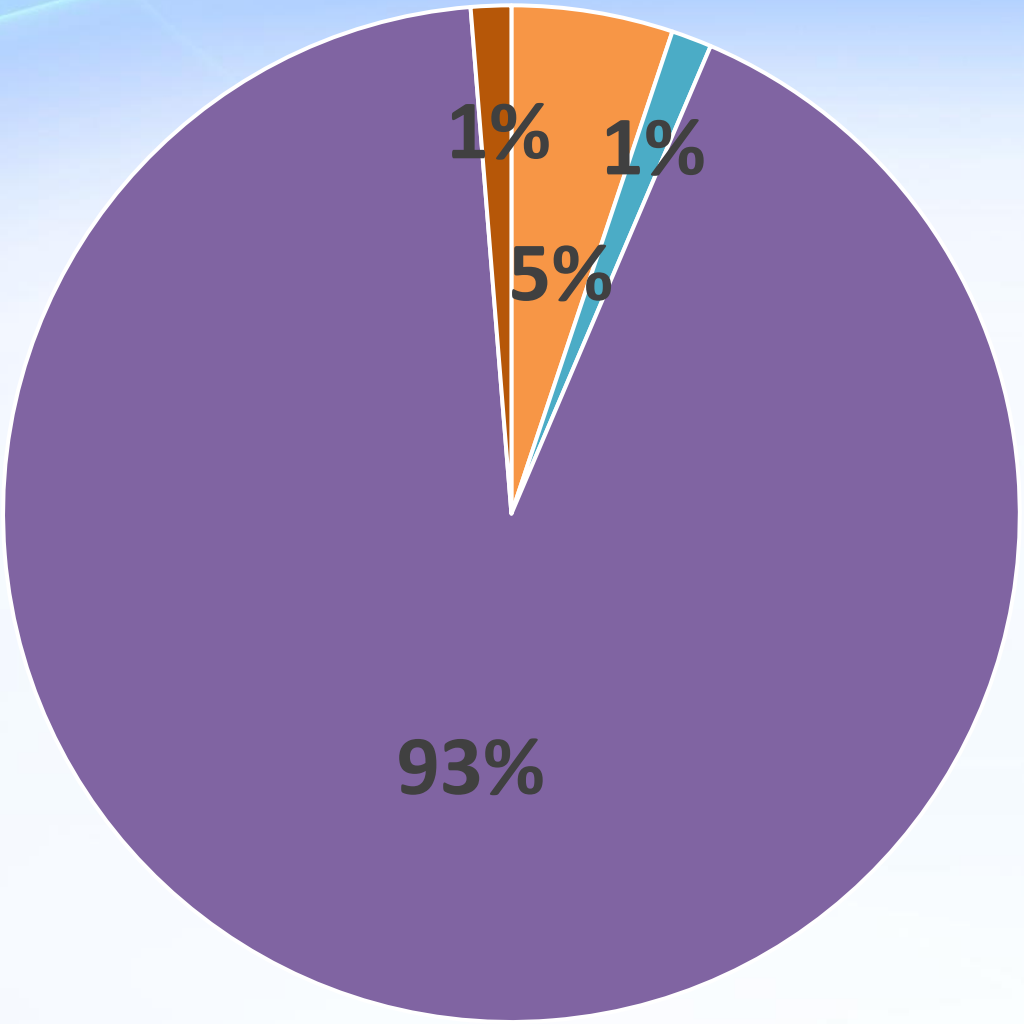
Weekly Use During Intervention



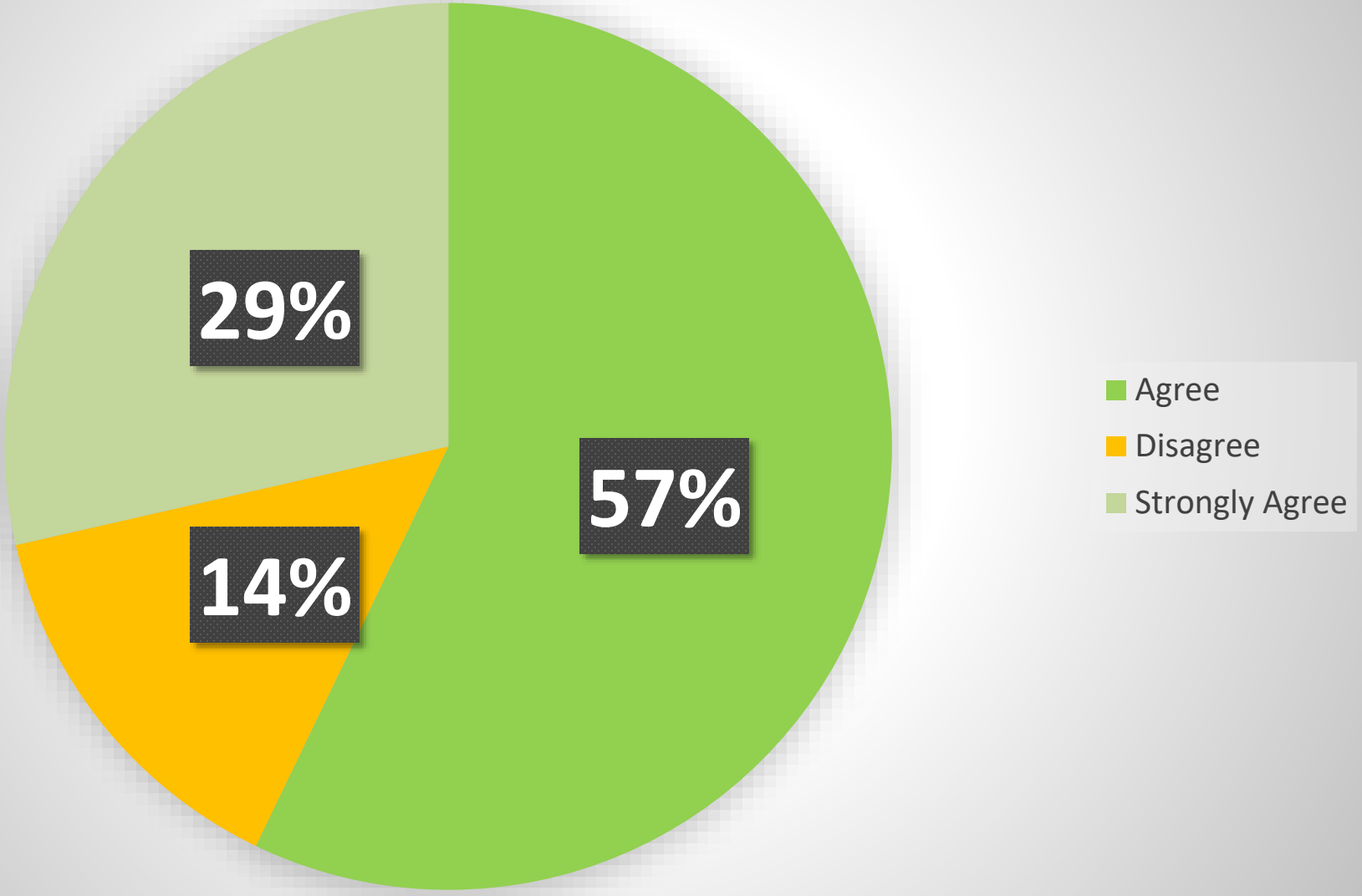
Time Spent Preparing Activity



Time Spent Scoring Activities

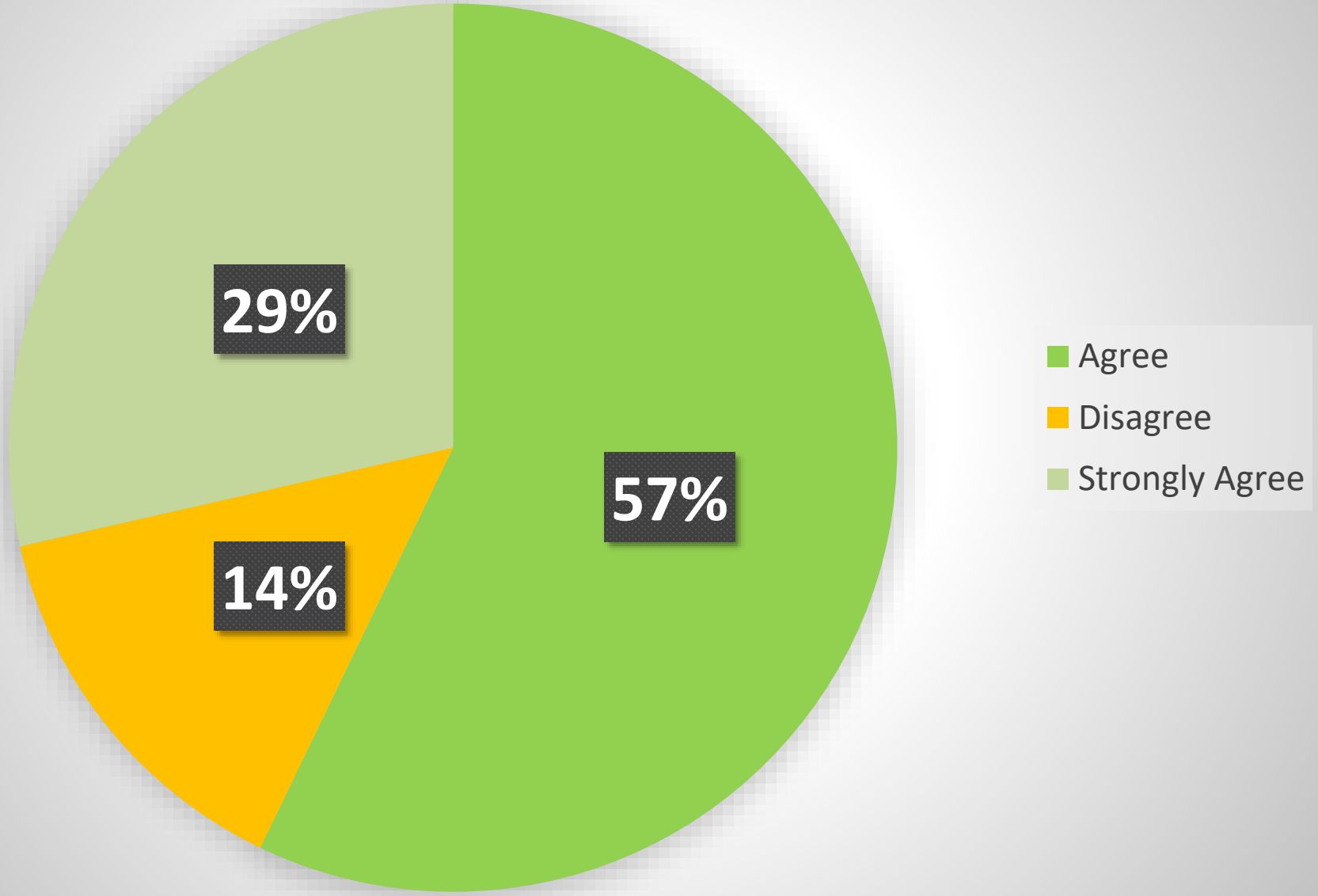


Automated scoring saves me time

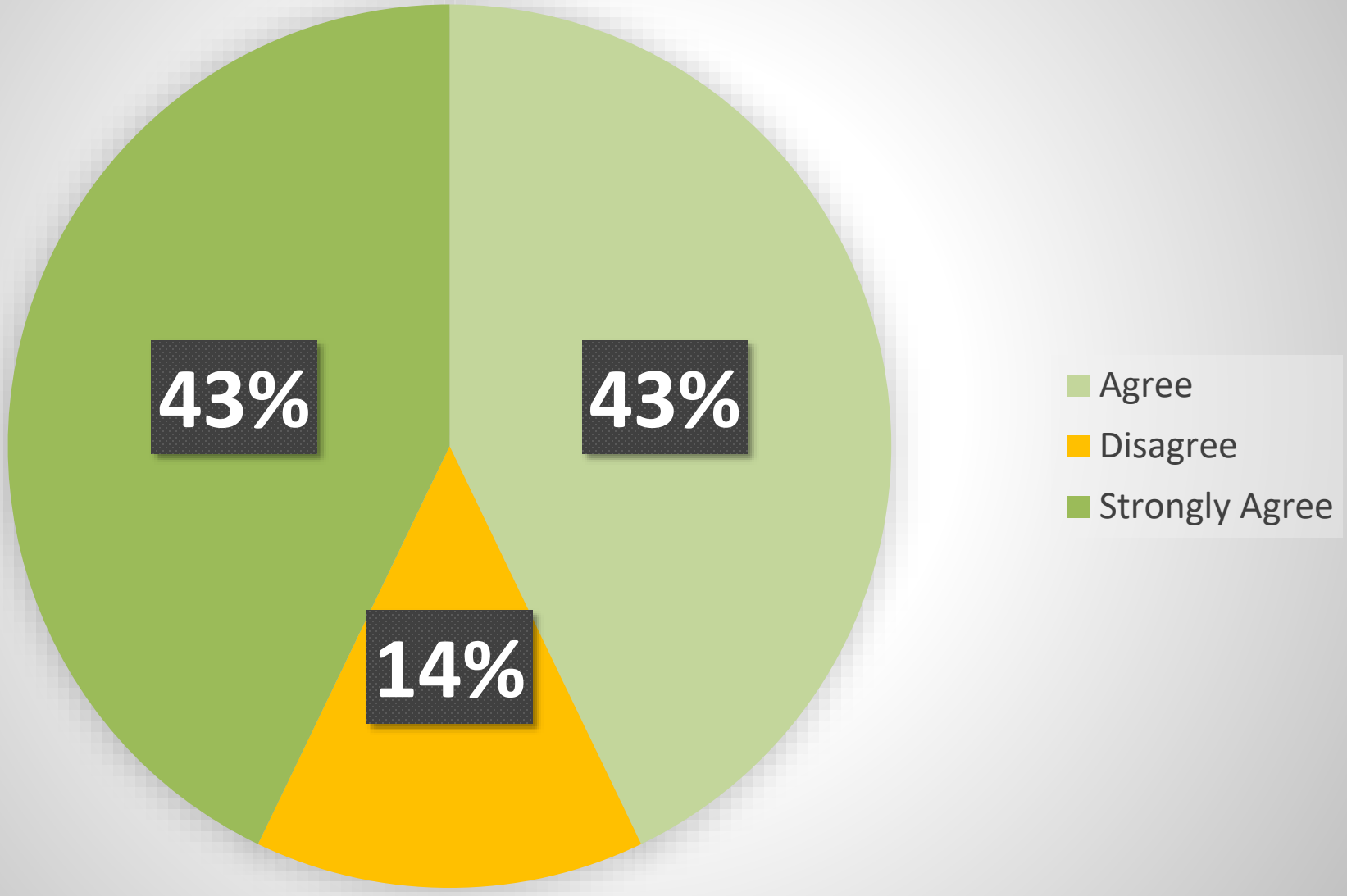


PERCEIVED LEARNING SUPPORT

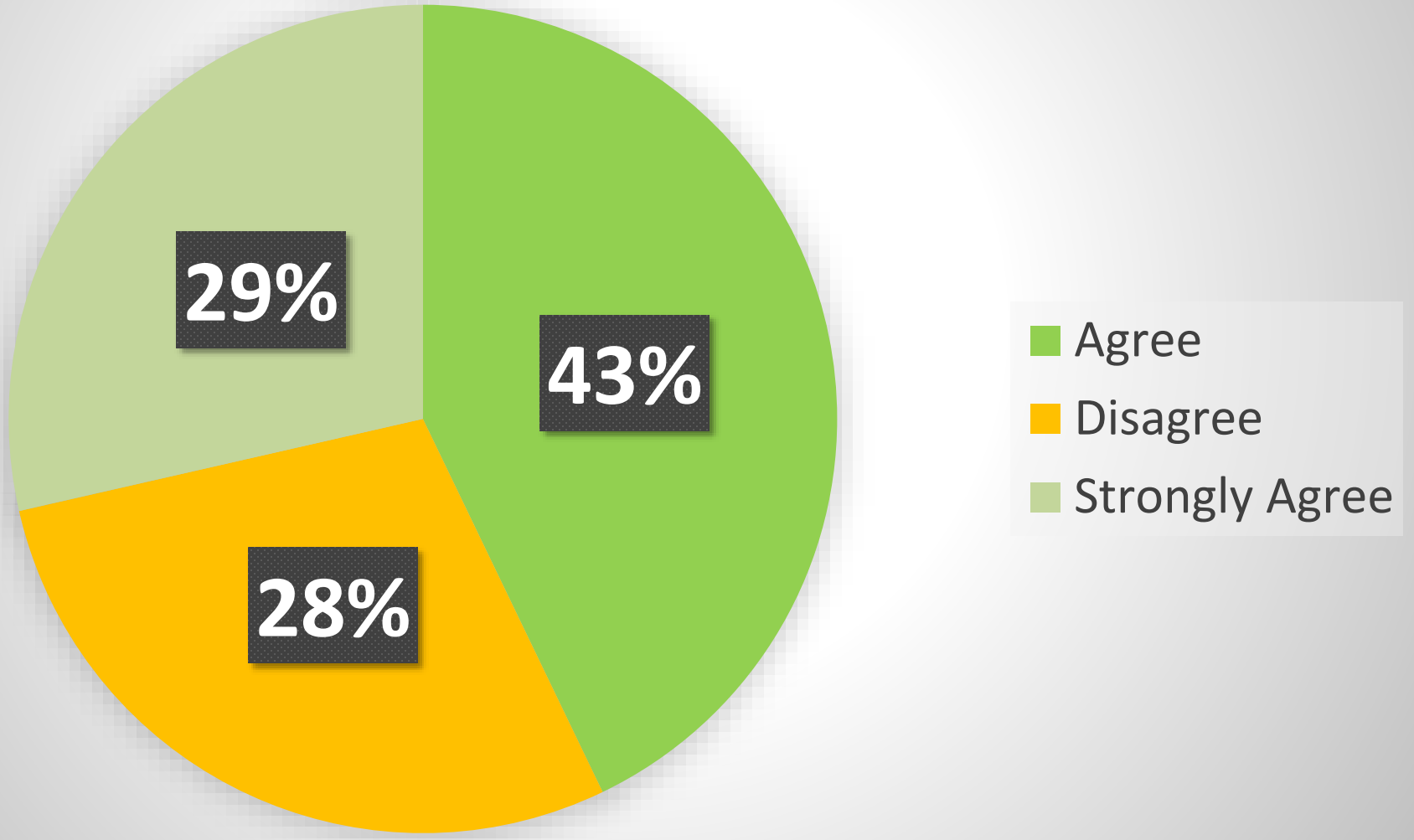
Activities align with state/district standards



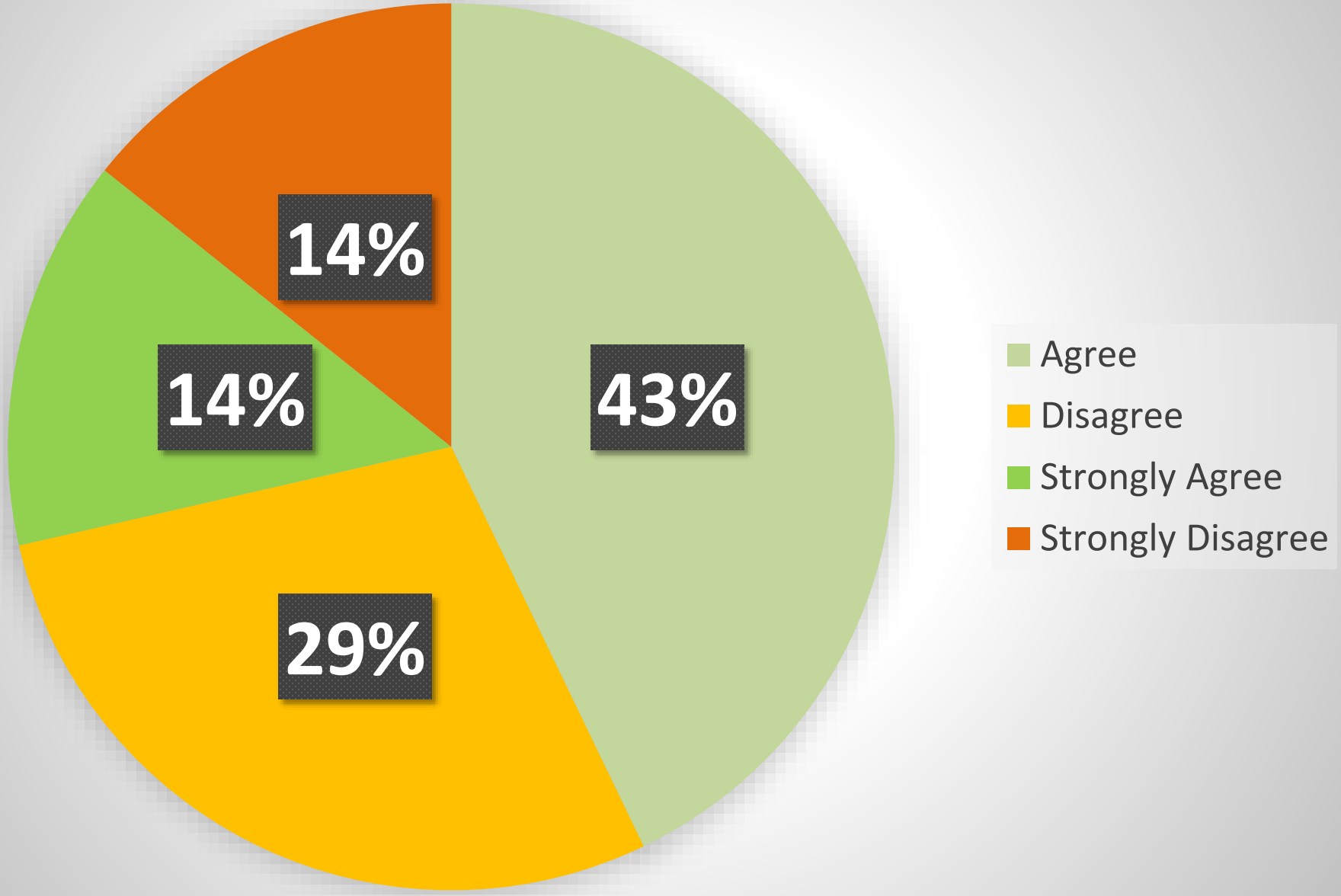
Activities build ELLs' language skills



Activities support students' understanding of text

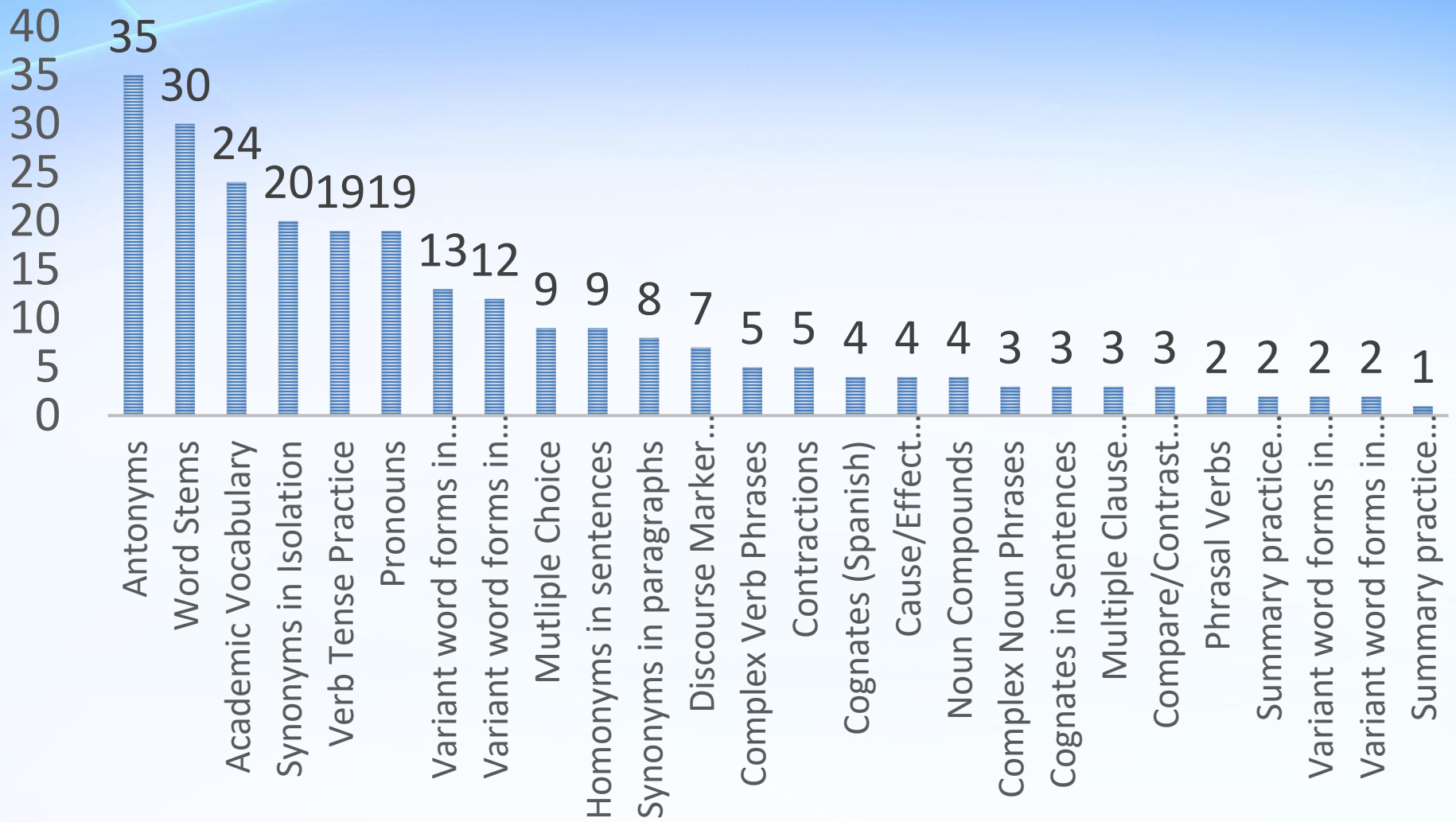


Students easily engage with activities

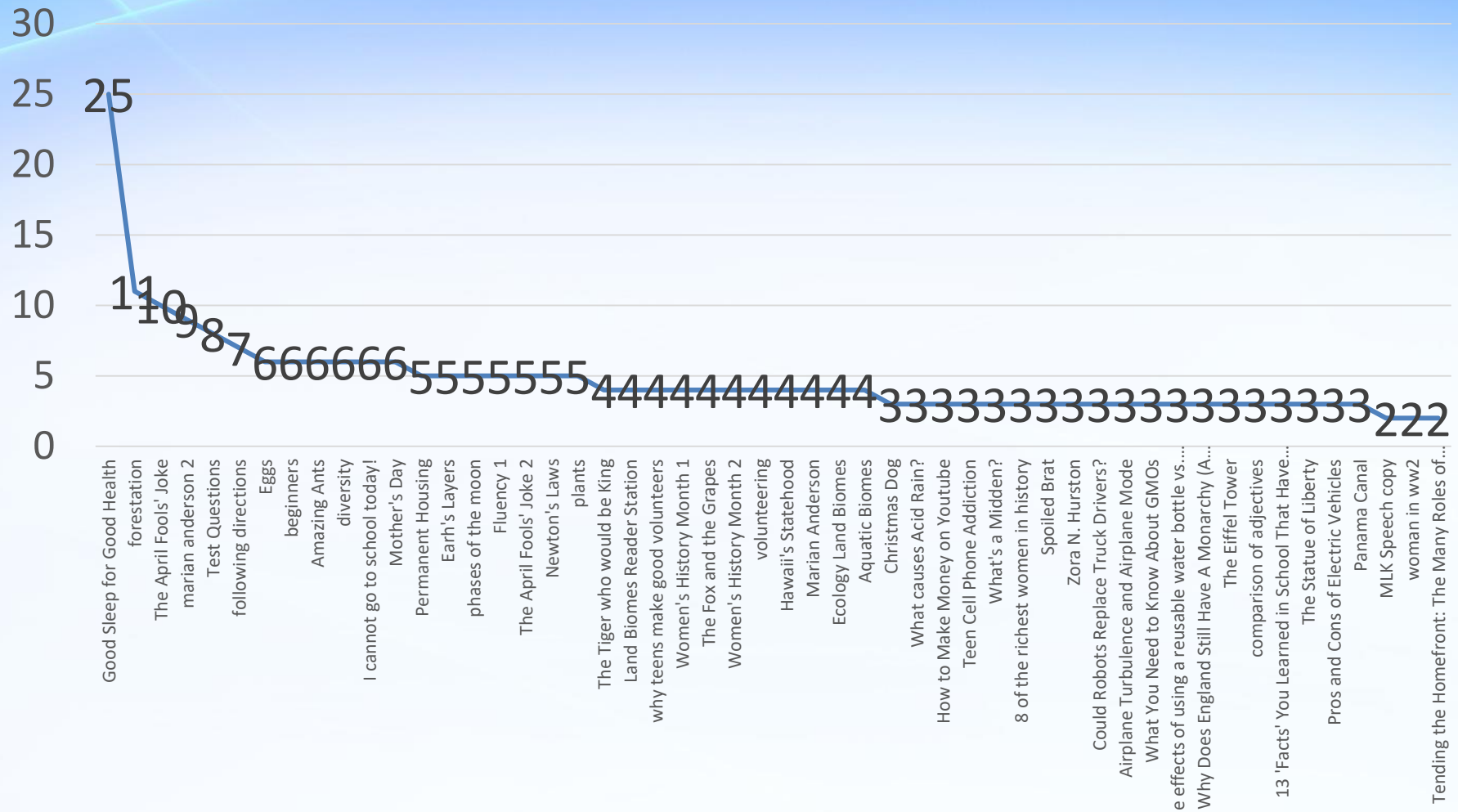


PRACTICE: USAGE LOG DATA

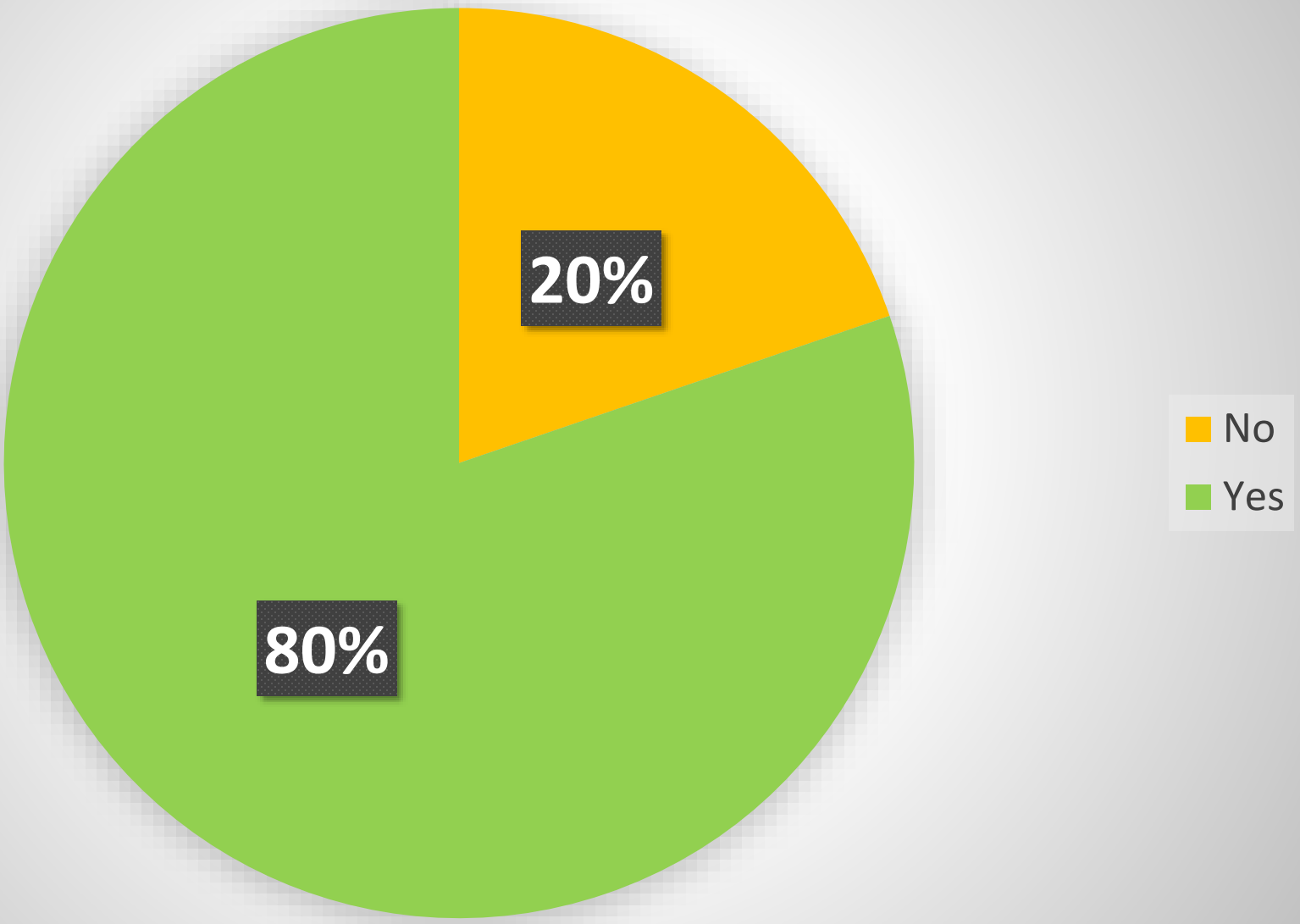
ACTIVITY TYPES



Activities Per Palette (51)



% with Automated Scoring



Summary

Language Muse Activity Palette

- NLP solution to support reading and language skills development
- Freely accessible here: https://languagemuse.org

Survey: What did teachers think?

- Generally, positive experiences
- Teachers percieved appropriateness of activities, standards alignment & student learning

Usage logs: How are teachers using it?

- Productive Palette & assignment creation
- Focus on vocabulary activities, but varied activity creation
- Efficiency: Likely to select automatically-scored activities

Student Outcomes:

TBD in Spring 2018

NLP Solutions for Writing

Some ETS history

E-RATER

E-rater®

NLP methods used to detect 50+ linguistic features

Features aggregated into 10-12 high-level features aligned with the human *holistic* scoring criteria (typically 4 – 6 point scale)

Each feature is represented by a module

- Rule-based: collection of manual rules and/or regular expressions
- Statistical: Statistical models compute feature values

Feature modeling with multiple regression

Linear equation with feature weights yields final score

E-rater Facts & Use Cases

Evaluates *expository, argumentative, & source-based essay writing*

Used for high-stakes assessment since 1999

Scores about 16M submissions yearly

High-stakes assessment

- Test-takers: Supports multiple writing measure types
- Adminstrators: Acceptance decisions

Classroom Instruction: *Criterion*®

- Students: support for writing quality
- Educators: supplemental grading support with scores & feedback
- Available with institutional subscription only

Next-Generation Writing Feedback

U.S. Literacy Challenge in Postsecondary Contexts

- 1.7 of 18 million college students lacking prerequisite skills
- Disjuncture in writing requirements from K12 to postsecondary (Bridgeman & Carlson, 1984; Melzer, 2014; Burstein et al, 2016)

Growing Body of Evidence that AWE feedback helps

- Attali (2004), Shermis et al (2004) showed increased production with *Criterion* use
- Chapelle et al (2015): relationship between correct *Criterion* error feedback & improved revisions
- Cassidy et al (2016) :positive teacher perceptions of feedback utility

Wider accessibilty & construct coverage needed

- *Criterion* accessible only through institutional subscriptions
- Grammarly version is "free" and accessible, but limited to English Conventions

Motivation

27% of U.S. 12th graders scored at or above "proficient" on the NAEP writing assessment (U.S. DoE, IES, & NCES, 2012)

Burstein et al (2016): Two college faculty surveys

- Collect perceptions of student writing competencies
- Inform AWE development
- Findings: Perceive a general lack of writing preparedness

How can an NLP solution provide convenient & relevant support?

The Writing Mentor™ app -- a Google Docs Add-on

http://mentormywriting.org/

Future Work: Usability Study

- Writing Mentor collects usage log data (JSON files)
- Planned Studies:
  - Amazon Mechanical Turk in Fall/Winter 2017
  - Post-secondary & Adult literacy settings
    - Integrate tool use into instruction
    - Collect student-user data
    - Evaluate usability: what do users use?
    - Examine utility: how does writing change?

Exploring Writing Achievement and Its Role in Success at 4-Year Postsecondary Institutions Funded by U.S. Department of Education, IES (Co-PI, Dan McCaffrey, ETS)

What can we learn about student success from writing features?

Significance

- Writing is a challenge, esp. for at-risk students w/o prerequisite writing skills required to persist in U.S. 4-year postsecondary institutions.
- Educators could benefit from a clearer understanding of writing achievement and its role in postsecondary success

Solution

- AWE allows processing and generation of linguistic features for large-scale data sets
- Writing-based linguistic & skill relationships can inform actionable analytics for students, educators, parents and policy-makers

Impact

- AWE features may provide meaningful information about student success predictors
- AWE has potential for educational analytics beyond assessment and instruction

4-year Exploratory Study

- **Year 1: Secondary data analysis with writing assessment data collected from college students to examine writing features & success predictors**
- Year 2: Collect *authentic* student writing data, writing assessment data, & writing attitudes survey data from students enrolled in 4-year institutions to examine writing features, writing attitudes, & success predictors
- Years 3-4: Examine relationships between student writing data and longitudinal success factors (e.g., continued enrollment)

Study Overview

Exploratory, secondary data analysis to examine relationships between responses to an on-demand essay writing task & broader success predictors.

Data

Writing: Essay assessment responses from 929 students from 22 4-year institutions.

Success predictors: Critical thinking assessment scores, SAT/ACT college admissions composite & subject scores & GPA.

Methods

26/200+ AWE features (English conventions, Coherence, Organization, Vocabulary Usage) selected using statistical evaluations, e.g., eliminate highly correlated features.

Regression analyses conducted to predict 6 success indicators. Independent variables:feature+length+human score.

Results

AWE features emerged across sub-constructs as predictors for all 6 indicators: (1) critical thinking assessment score, (2) writing assessment selected response, (3,4) SAT/ACT composite scores, (5,6) SAT/ACT subject area scores, (7) college GPA.

Example: SAT Verbal Score

Variable Construct	Coefficient	Std. Error	R^2	Inc. R^2
grammar	0.11	0.04	0.18	0.01
word usage	0.14	0.04	0.18	0.02
mechanics	0.15	0.04	0.18	0.02
sentence variety	0.29	0.06	0.21	0.04
vocabulary sophistication	0.15	0.04	0.19	0.02
vocabulary complexity	***0.29***	***0.04***	***0.24***	***0.07***
word usage	0.12	0.05	0.18	0.01
argumentation	0.13	0.05	0.18	0.01
personal reflection essays	-0.15	0.04	0.19	0.02
phrasal variety	0.11	0.05	0.17	0.01
derivational morphology	0.13	0.05	0.18	0.01
infllectional morphology	0.22	0.05	0.20	0.04
vocabulary richness	***0.33***	***0.05***	***0.23***	***0.07***
coherence	0.28	0.13	0.17	0.01
sentiment	0.12	0.04	0.18	0.01

Current Study Underway

- *Amidst* Year 2: Collect *authentic* student writing data, writing assessment data, & writing attitudes survey data from students enrolled in 6 4-year institutions to examine writing features, writing attitudes, & success predictors
- Conduct similar analyses, but will have *authentic* student writing data and self-efficacy and beliefs responses!!

NLP & the Literacy Challenge

Accessible NLP solutions & the literacy challenge

- Reading and language skills development: *The Palette*
- Writing instruction: The *Writing Mentor*
- Potential educational analytics beyond assessment and instruction

More NLP-Literacy solutions in more domains

- Peer review systems
- Automated evaluation of speech (e.g., reading apps)
- Multimodal (spoken dialog & video): workforce interviewing tools

Language Muse Team

- **ETS Research Team**: John Sabatini (Co-PI), Dan McCaffrey (Stat Lead), Nitin Madnani (Tech Lead), Kietha Biggers & Kelsey Drier (School Leads)
- **ETS Engineering Support**: Slava Andreyev

Writing Mentor Team

- **ETS Research Team**: D. McCaffrey (Co-PI), Beata Beigman Klebanov, Nitin Madnani, & Guangming Ling.
- **ETS Engineering Team**:10clouds.com

Thanks! & Questions?