

Productive Academic Vocabulary for College Success

Jamie Storyward

School of Education, UNC Chapel Hill

EDUC 795: Learning Analytics

Dr. Matthew Bernacki

April 30, 2024

Problem: A positive correlation between vocabulary knowledge and academic success is well-supported in literature and research. Academic vocabulary deficits in writing are a stumbling block to students beginning college. Can we improve early college success via academic vocabulary? Additionally, students' vocabulary knowledge is frequently assessed using in-class exercises and tests. Ultimately, learners need to transfer this knowledge to their working vocabularies to maintain the gain in the future. Can the ability to track learners' digital writing paired with our increasing reliance on digital communication more accurately judge learners' knowledge of vocabulary?

Goal: Enable students to monitor their development of a productive academic vocabulary to improve college outcomes.

Audience: Middle and high school students receiving vocabulary instruction

Environment: Learners would grant researchers permission to view their social media posts and course assignments.

Learning Theories: The impetus for this project comes from literature on academic vocabulary acquisition and transfer of knowledge to new contexts. It relies on spaced retrieval as an indicator of enduring learning. The presentation of the data to the student is guided by Self-Regulated Learning theory.

Method: We would build a data store and text analysis mechanism. Instructors would select which words from the Productive Academic Vocabulary List (PAVL) they have taught. Students' academic writing and social media posts would be fed into the system. Researchers would use text mining to detect and count students' organic uses of the vocabulary words across sources. By incorporating casual and formal contexts, we could judge both near and far transfer.

Natural language processing would analyze the context to determine whether the term was used correctly.

The learning analytics tool would produce a customized dashboard for each student. Their dashboard would show their personal vocabulary acquisition statistics, including links back to the source material. Use will be broken out between near and far contexts. The dashboard is designed around Self-Regulated Learning theory to empower the individual to control their own experience.

Future directions: With some modifications, this project could determine students' transfer of skills and procedures to other contexts. For example, it could track whether a graduate student incorporates more effective instructional techniques into lesson plans following relevant education.

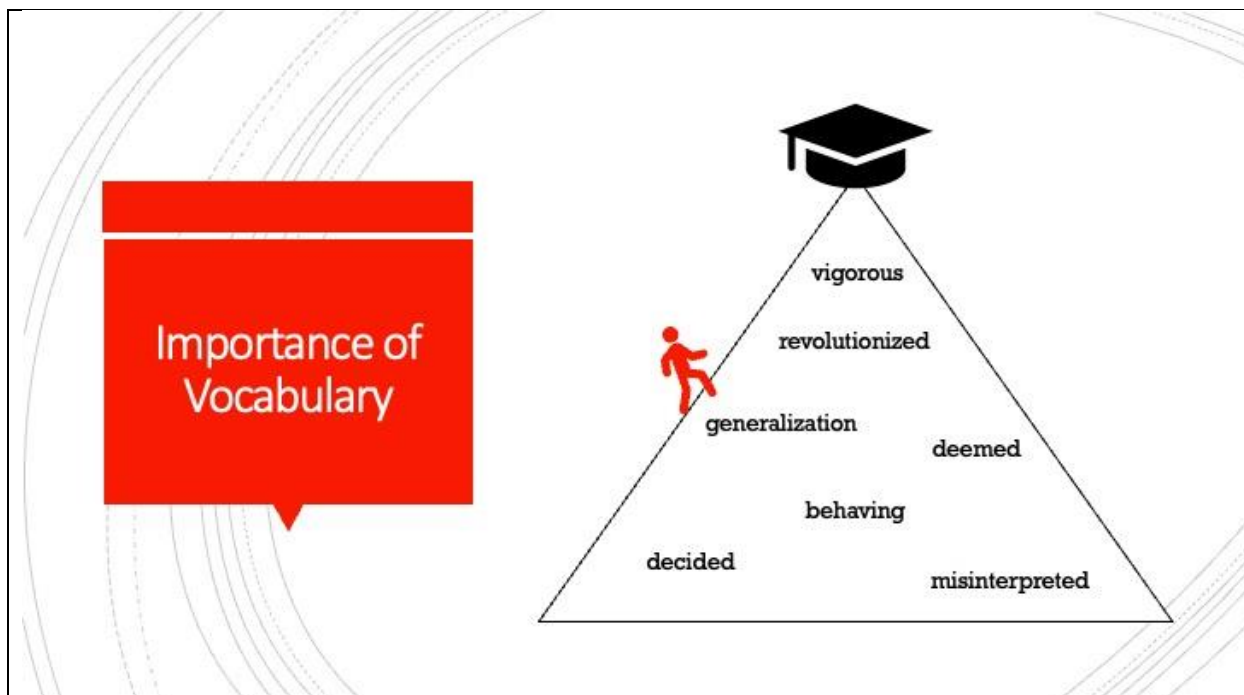


Productive Academic Vocabulary for College Success

Jamie Storyward

EDUC 795 Learning Analytics

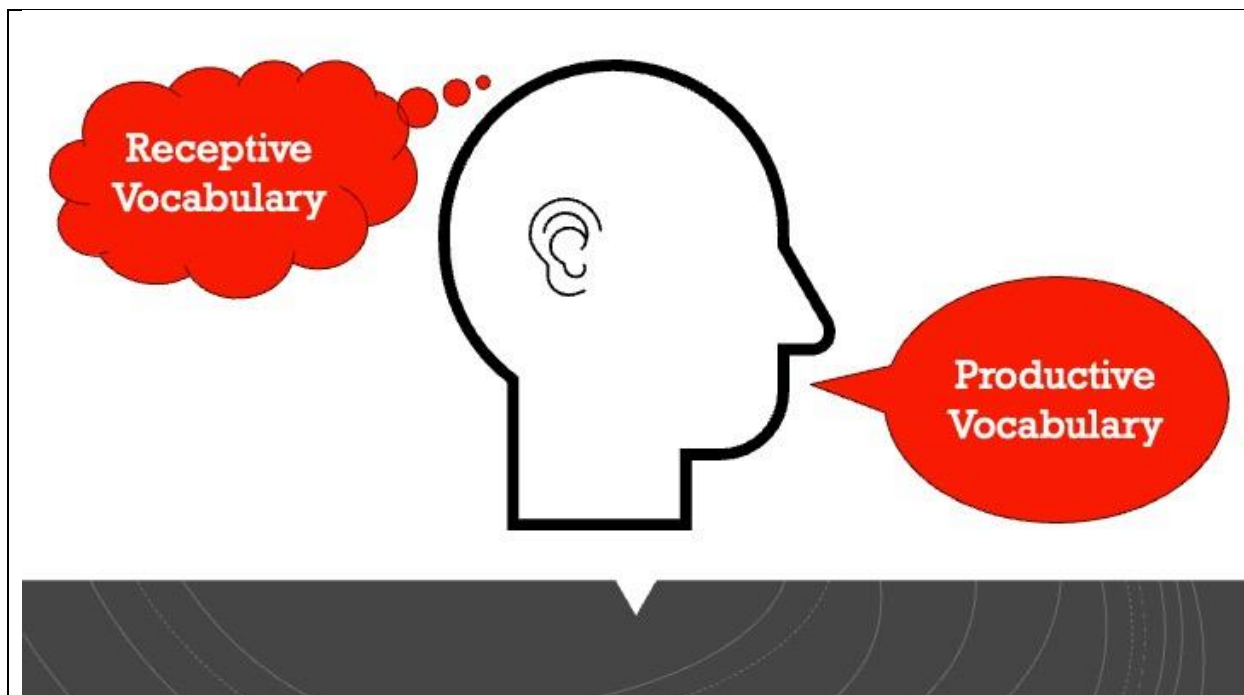
Spring 2024



The link between a student’s vocabulary and success in school is well established.

“There is a growing body of work showing the importance of academic language proficiency for accessing the content of academic texts and academic talk (Bailey & Heritage, [2008](#); Guerrero, [2004](#); Schleppegrell, [2004b](#)); learning to think and learn like a scientist, historian, mathematician, or writer (Honig, [2010](#); Schleppegrell, [2007](#); Shanahan & Shanahan, [2008](#)); and overall academic achievement (Hakuta, Butler, & Witt, [2000](#); Short & Fitzsimmons, [2007](#); Townsend, Filippini, Collins, & Biancarosa, in press).” (Nagy & Townsend, 2012, introduction)

However, knowing this does not mean all students are now entering college with adequate vocabulary. According to one 2007 study of Cantonese speaking students at an English speaking university in Hong Kong, researchers Evans and Green concluded that “inadequate receptive and productive vocabulary in English is the main problem confronting the almost 5,000 students who participated in the survey” (pp. 13–14). There is a need for students to improve their vocabulary to improve their academic outcomes.



As the previous quote brought up, there is a difference between a student's receptive and productive vocabularies.

Receptive vocabulary refers to the body of words someone understands when reading or listening.

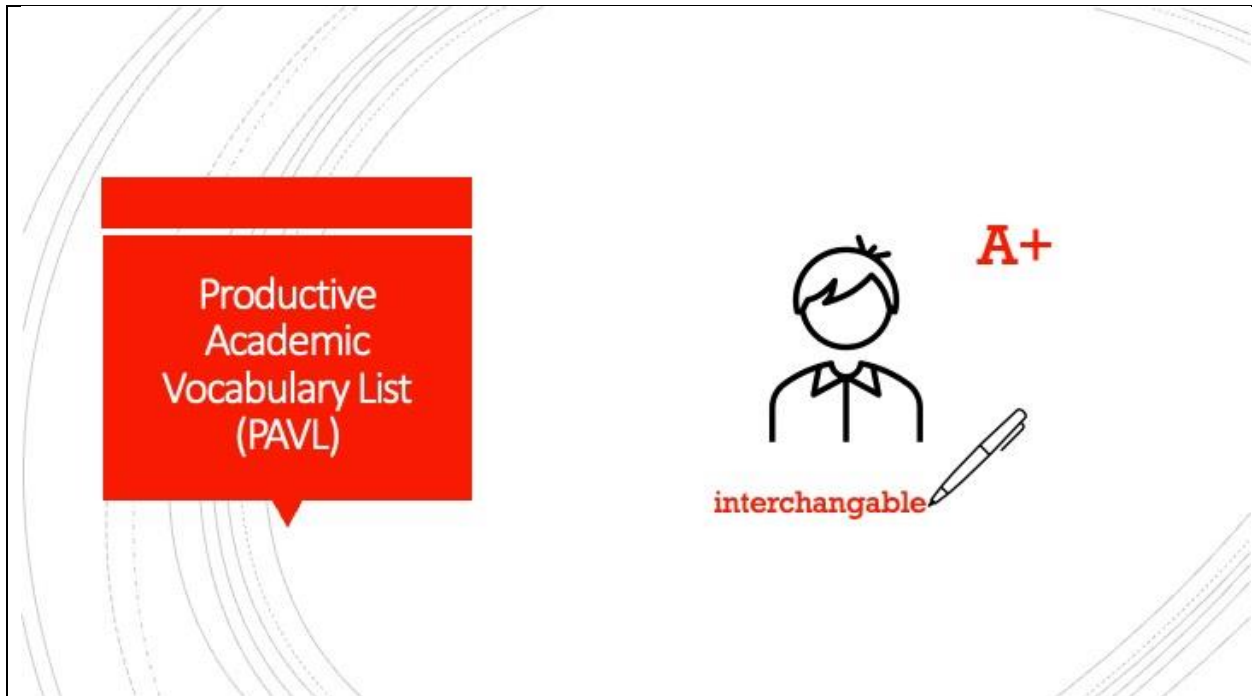
Productive vocabulary refers to the body of words someone uses in writing or speaking.

Receptive vocabulary is typically larger than productive vocabulary, in both primary and secondary languages. This trend is also consistent in both casual and academic contexts.

This project focuses on the importance of expanding student's productive vocabulary with specific academic vocabulary words. For our target words to enter the student's productive vocabulary, they also need to enter or be in their receptive vocabulary.


However, our data collection will track uses of vocabulary words, not times when they were successfully understood.

We focus on productive vocabulary to isolate students' development of college-appropriate writing skills.



In this research, our focus is on improving students' academic success with vocabulary, which means focusing on academic vocabulary. Defining academic vocabulary from general vocabulary, from academic language, and from subject-specific vocabulary is controversial (Nagy and Townsend, 2012).

For this work, we follow the lead of researchers Malmström, Pecorari, and Shaw (2018). They uncovered a list of 591 words college students use in high-scoring academic writing that are not common in general writing. They have identified these as important for students to learn to write with to be successful in college. They are known as the Productive Academic Vocabulary List, or PAVL.



"It is repeated exposures to these words and opportunities to **practice using them in authentic contexts** that allow students to own these words and **use them** with facility in the contexts in which they both garner and support meaning of technical or theoretical ideas."

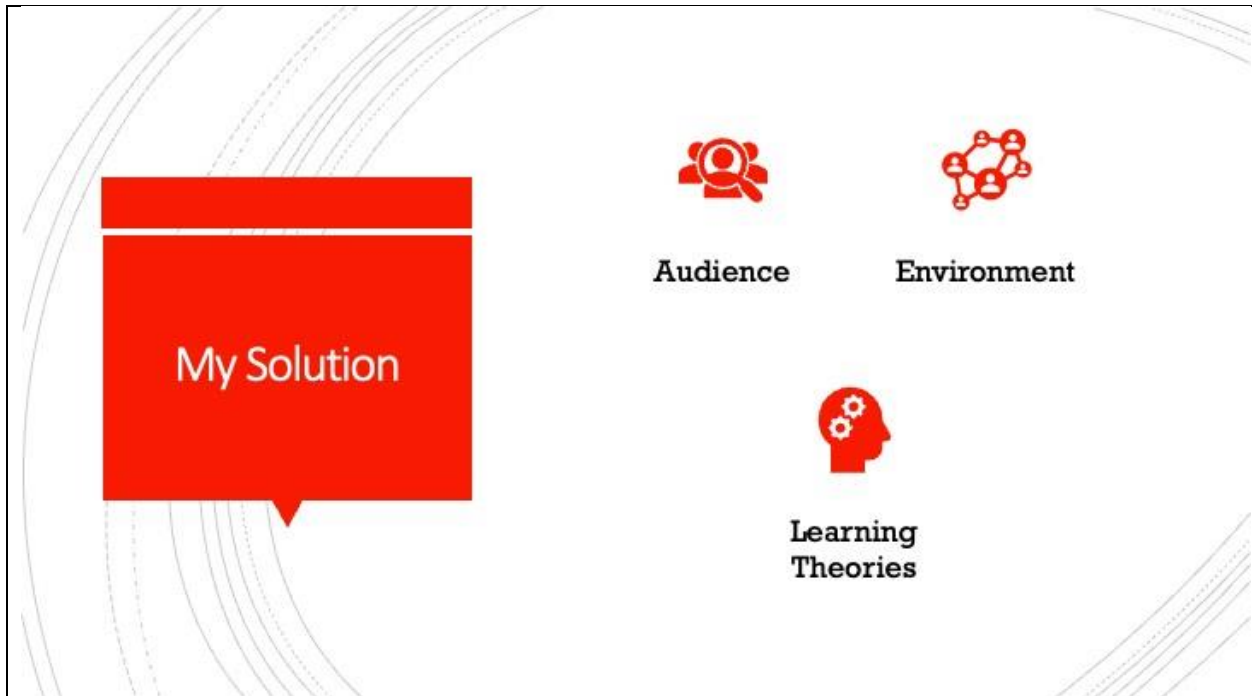
Nagy & Townsend, 2012

For students' writing to succeed, their productive academic vocabulary must grow. By using words in authentic contexts – e.g. outside of assigned vocabulary exercises – whether personal or academic, students will truly understand the words and be able to use them effectively in college writing.



Enable students to monitor
their development of a
productive academic
vocabulary to improve college
outcomes.

My project aims to enable students to monitor their development of a productive academic vocabulary so they may succeed in college.



Audience: In particular, my audience is middle and high school students receiving vocabulary instruction in their English Language Arts classes on Malmström's PAVL.

Environment: For the project, learners would grant researchers permission to view their social media posts and course assignments.

Learning Theories: The impetus for this project comes from literature on academic vocabulary acquisition and transfer of knowledge to new contexts. It relies on spaced retrieval as an indicator of enduring learning. The presentation of the data to the student is guided by Self-Regulated Learning theory.

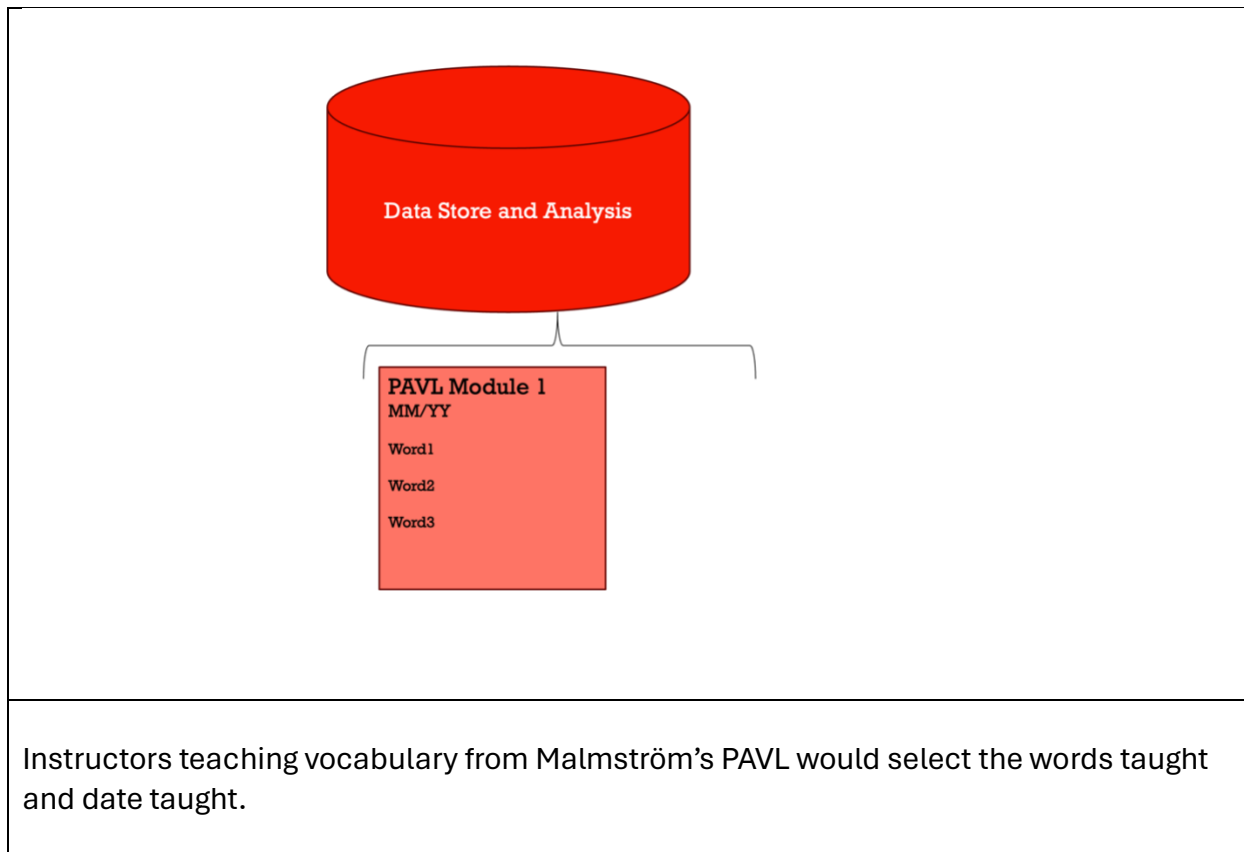
Slide 1 of 7



Here is a diagram of the method.

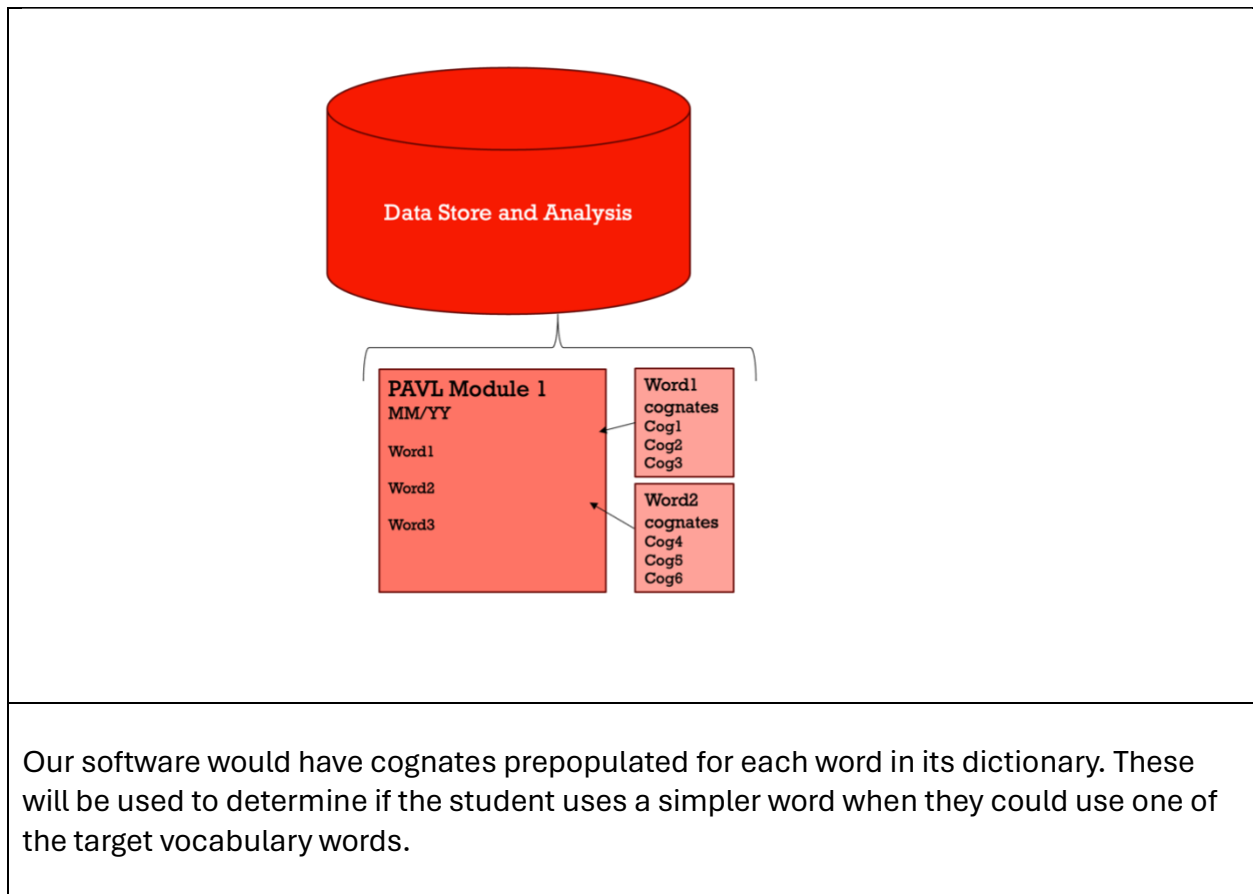
We would start by building our program: a data store and analysis mechanism.

Slide 2 of 7

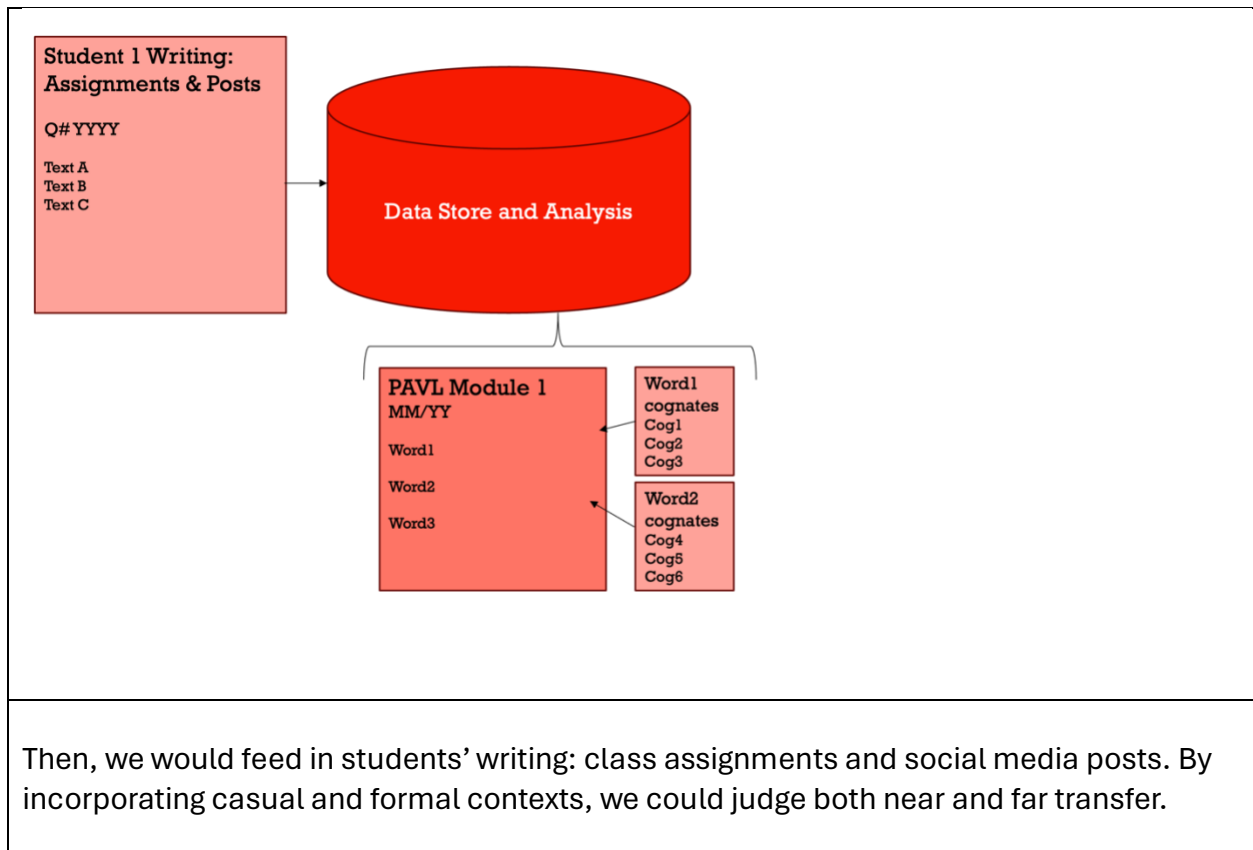


Instructors teaching vocabulary from Malmström's PAVL would select the words taught and date taught.

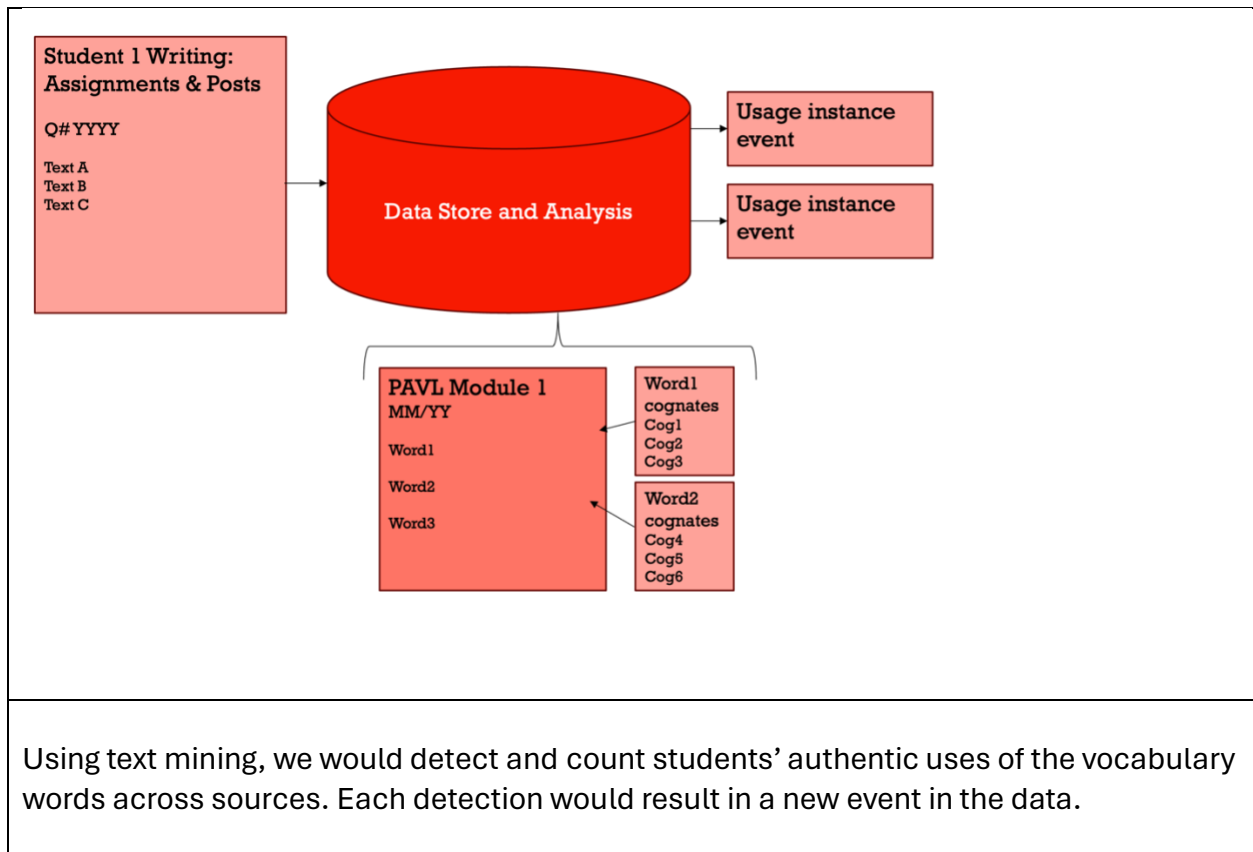
Slide 3 of 7



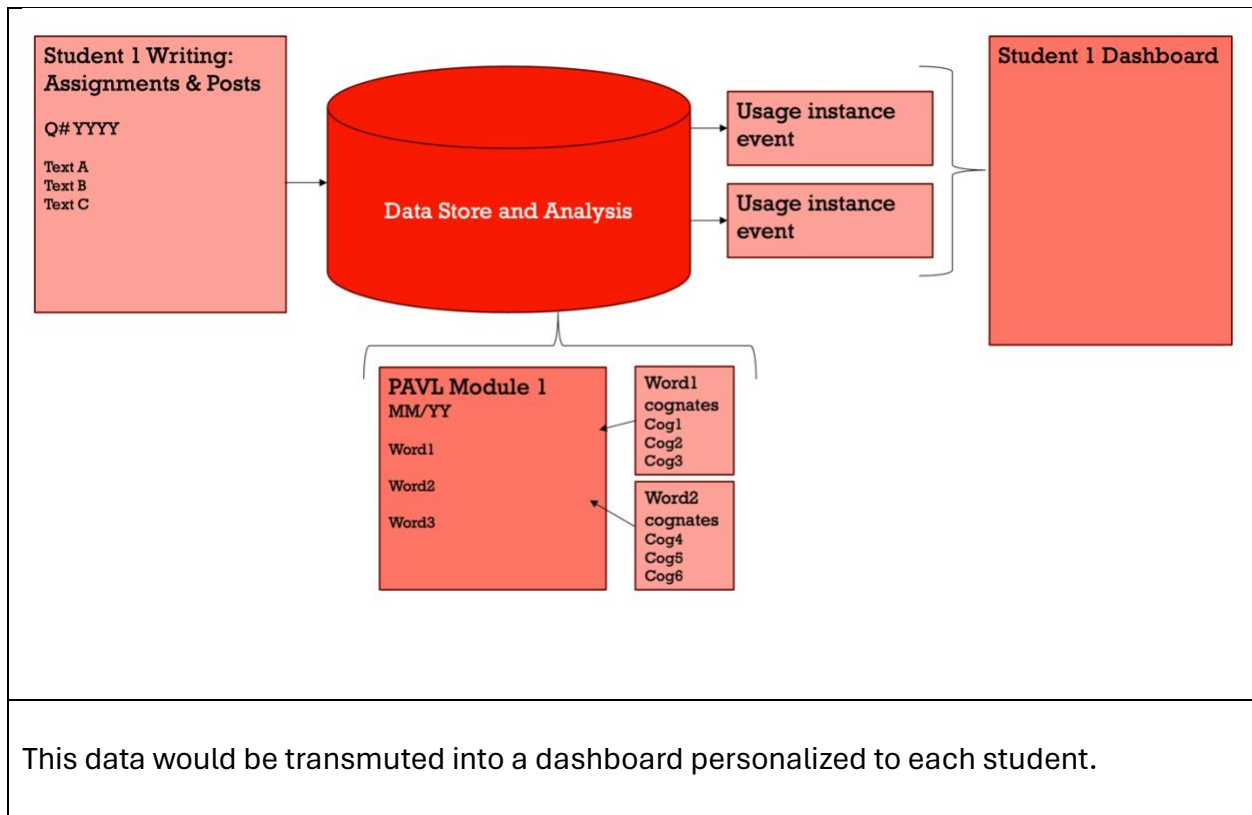
Slide 4 of 7



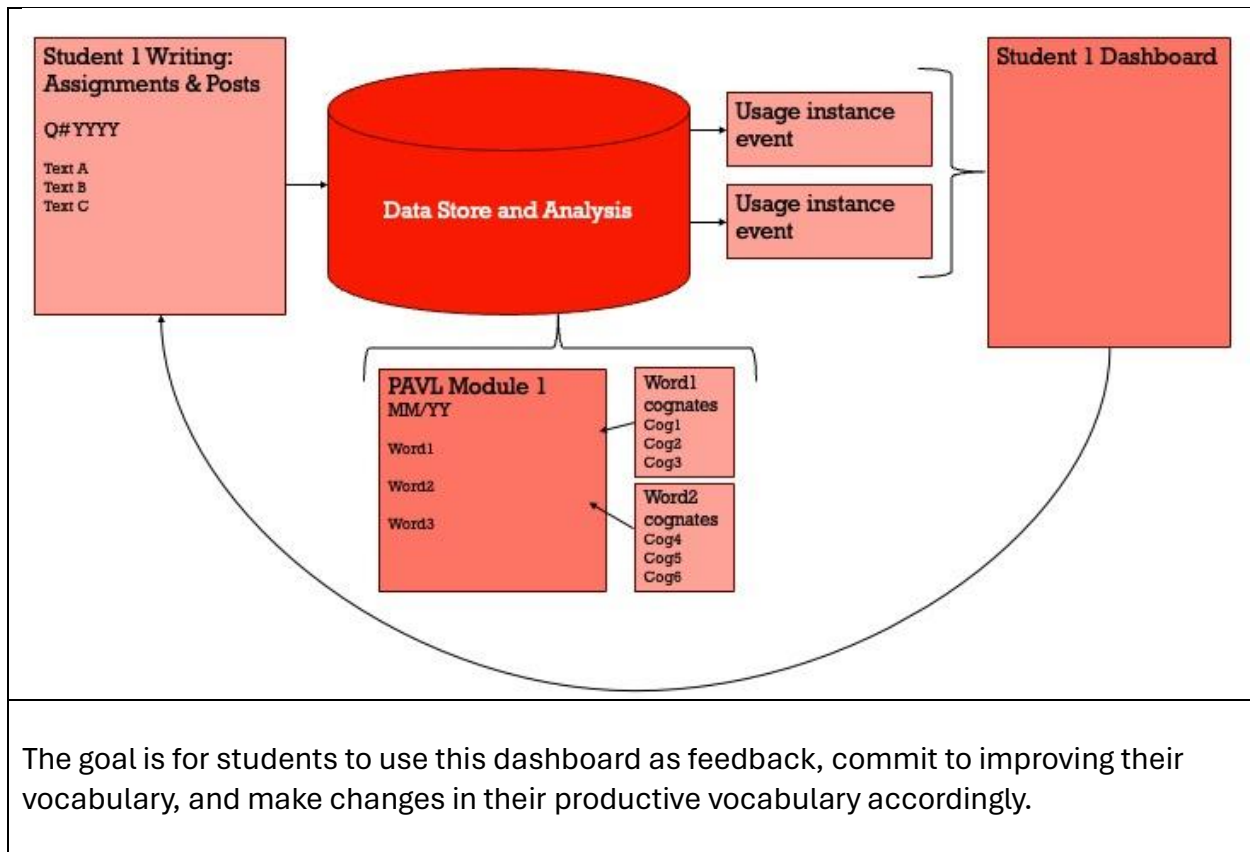
Slide 5 of 7



Slide 6 of 7



Slide 7 of 7





I'll now go over a particular example in more detail. Meet Elena – she is a 9th grade student in 2024, learning PAVL in her English Language Arts class.

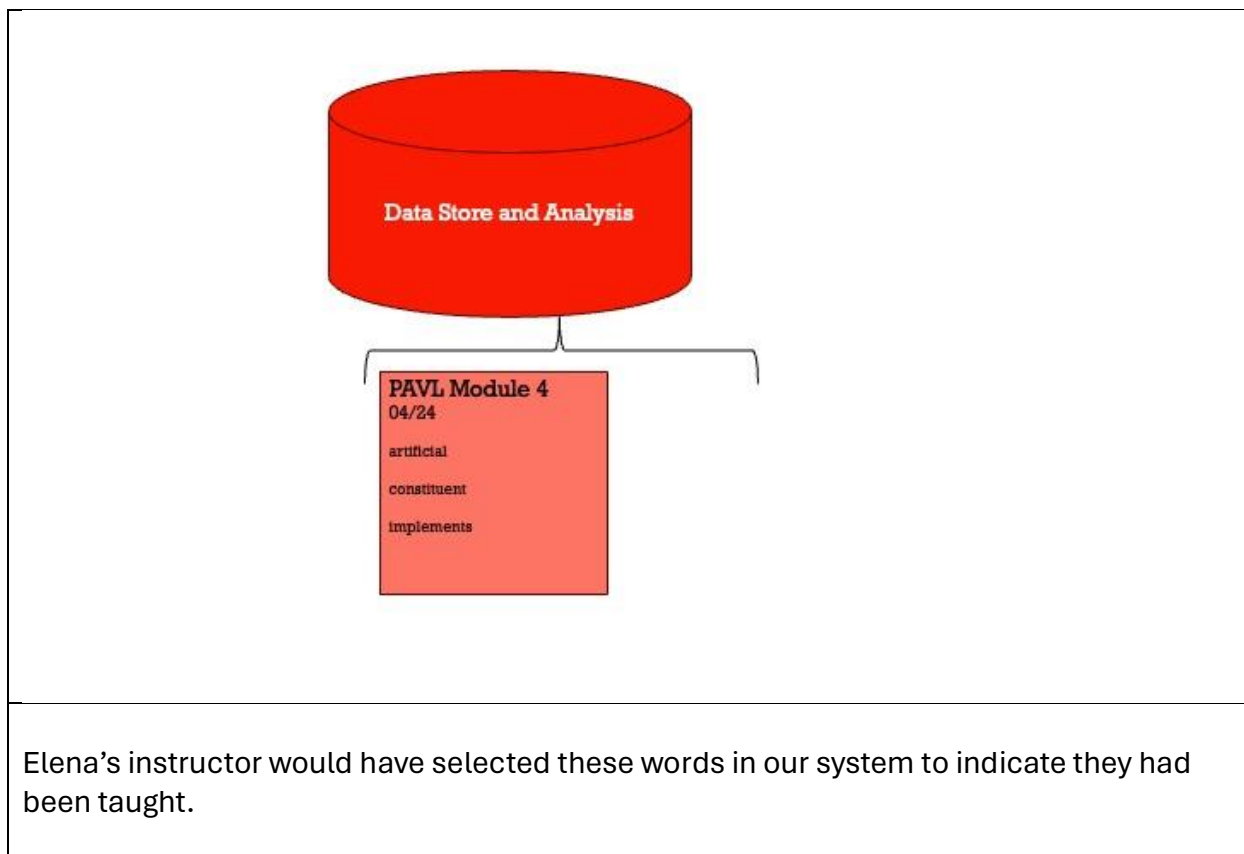
Vocabulary Module 4

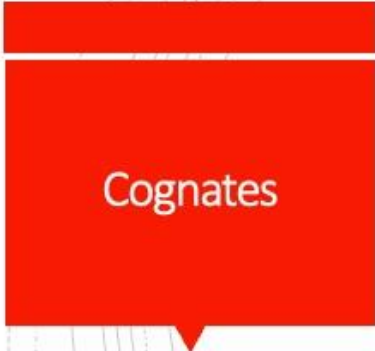
ELA

April 2, 2024

1. artificial
2. constituent
3. deviation
4. generalization
5. implements
6. realistic
7. utilized
8. whether

Here is one set of vocabulary words taught in Elena's English Language Arts class, containing words such as constituent, deviation, implements, and realistic.

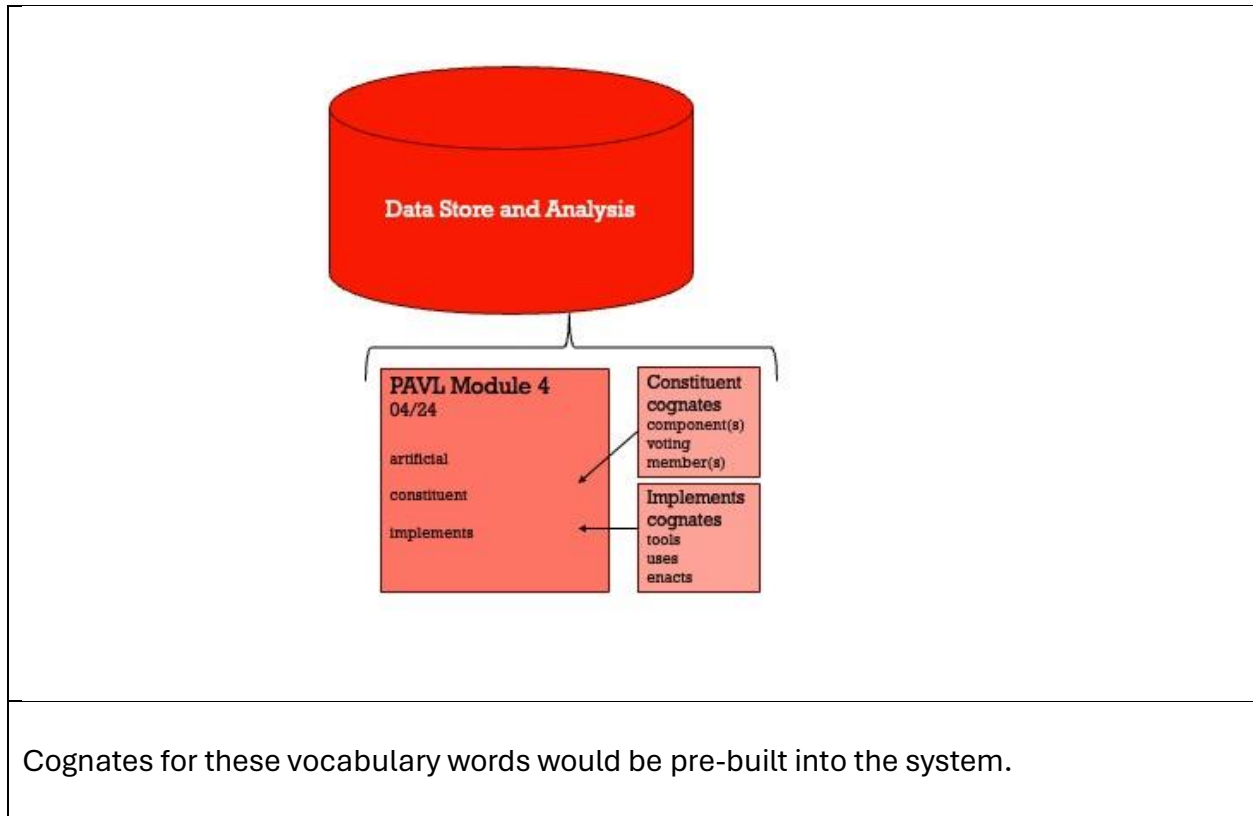




- **Constituent(s)**
 - Component(s)
 - Voting member(s)

- **Implements**
 - Tools
 - Uses
 - Begins using
 - Makes use of
 - Applies
 - Enacts

Here are a few cognates for some of the vocabulary words on Elena's list.



Romeo and Juliet

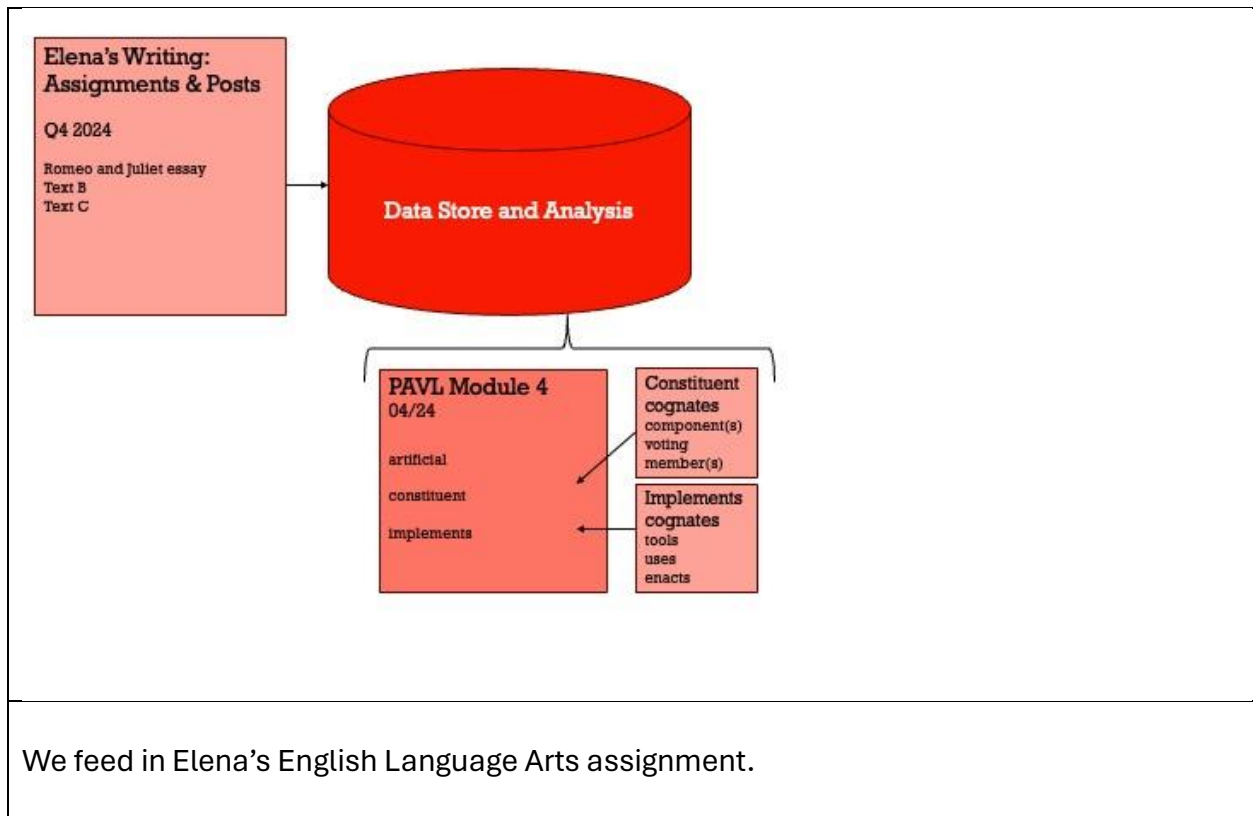
April 7, 2024
ELA
Elena

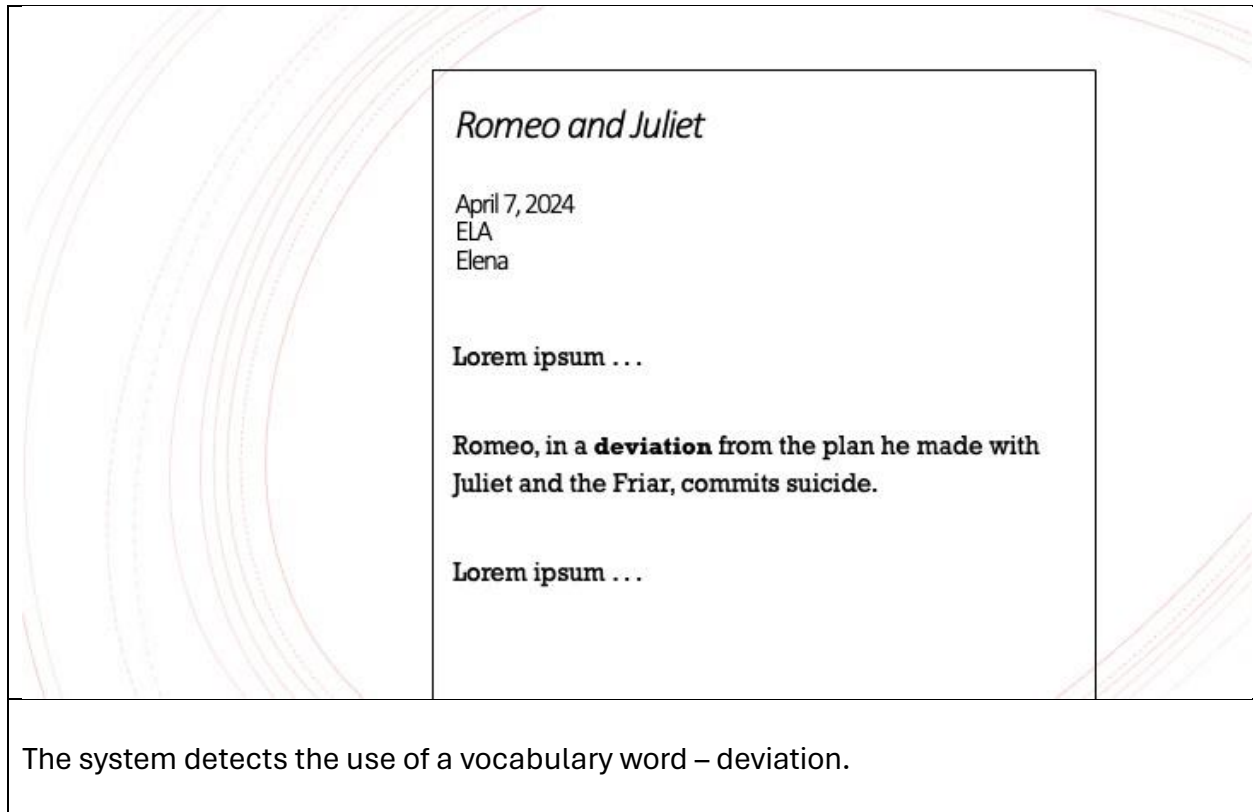
Lorem ipsum . . .

Romeo, in a deviation from the plan he made with Juliet and the Friar, commits suicide.

Lorem ipsum . . .

Here is a sample of Elena's writing, an English Language Arts assignment on *Romeo and Juliet*.





Romeo and Juliet

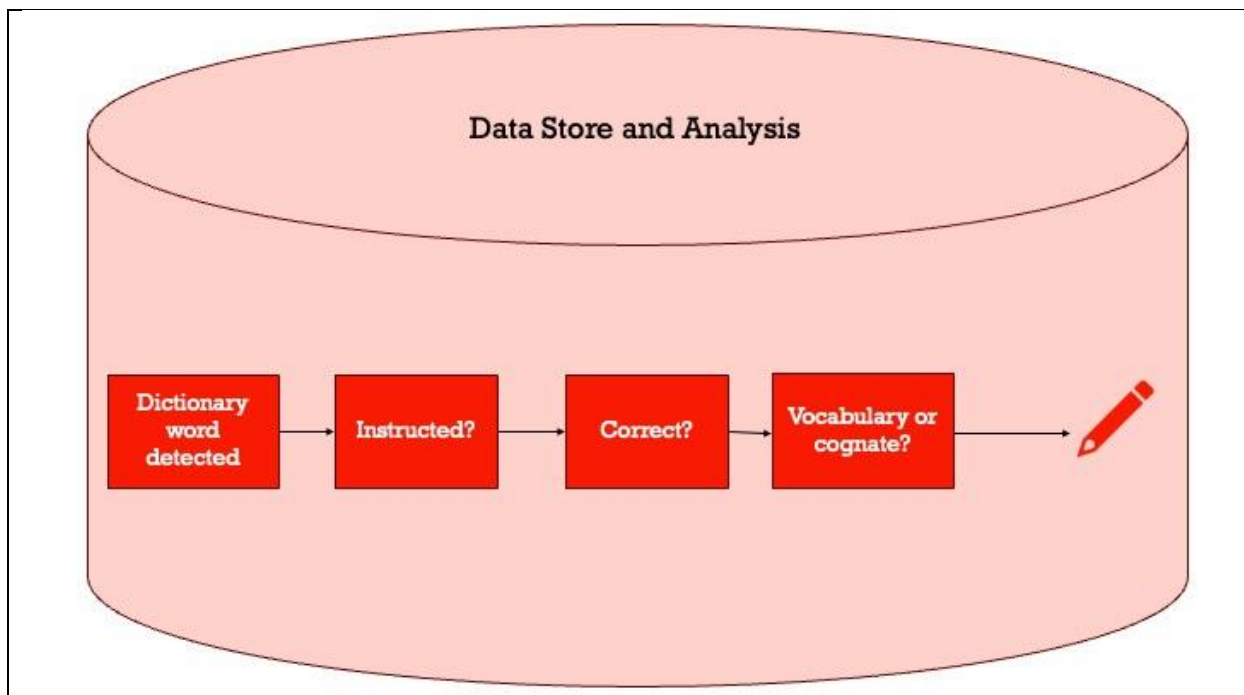
April 7, 2024
ELA
Elena

Lorem ipsum . . .

Romeo, in a *deviation* from the plan he made with Juliet and the Friar, commits suicide.

Lorem ipsum . . .

The system detects the use of a vocabulary word – deviation.



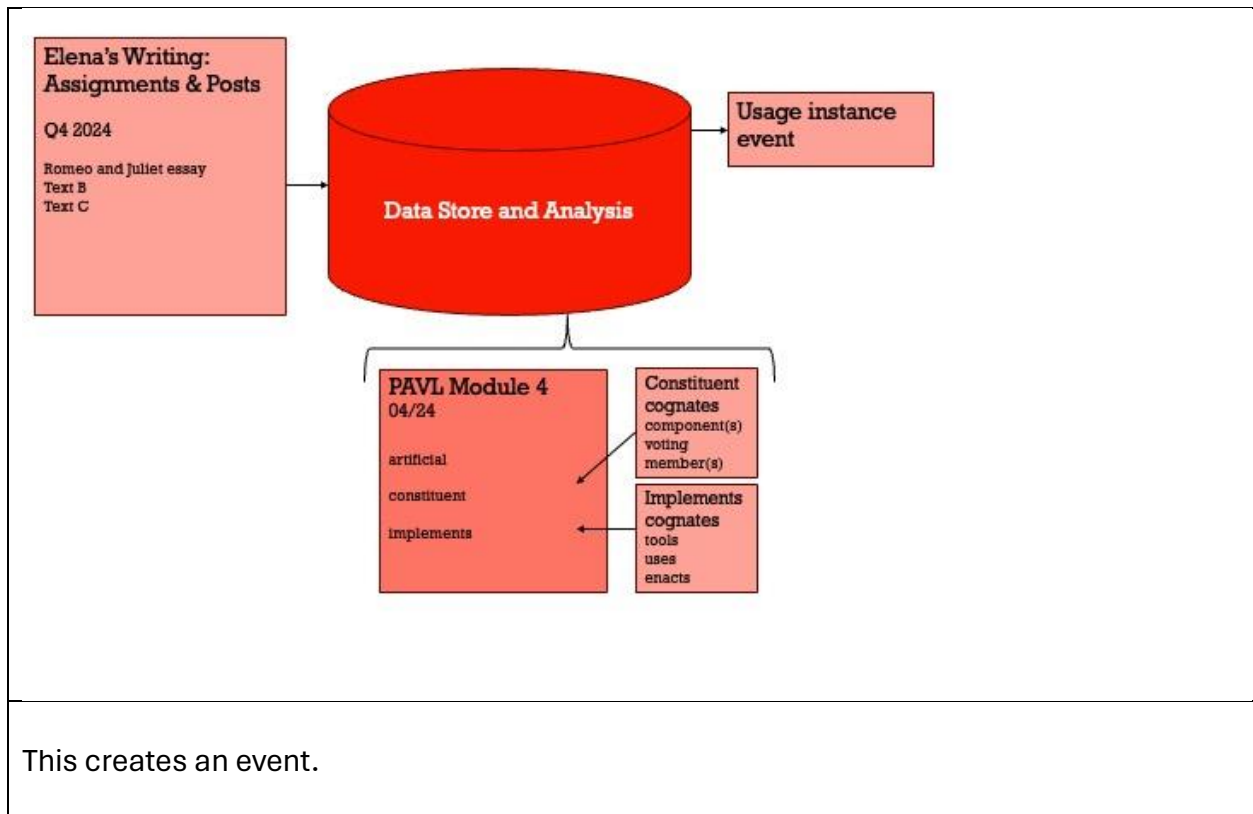
Within the data store, important analysis is occurring based on text mining.

When a detection occurs, it searches the word lists to see if the word has been taught yet, and if so, when. Detections will still be recorded for uninstructed vocabulary words to reflect advanced students' independent progress on the dashboard.

Natural language processing is also used to determine if the word is being used correctly, or if the vocabulary word could be used appropriately, if this is a cognate.

A simple step: the system determines whether the detected item is on the list of vocabulary words or the list of cognates. This determines the type of event.

All of this data is recorded.



User ID	Event ID	Timestamp	Event type	Vocab word ID	Instructed ?	Number of transfers	Source	Correct?
8675	48151642	20240407	001 – Vocabulary word used	2024_4_3 - deviation	Yes	1	Schoolwork	Yes

Here is how the event might look in the data.

- User ID indicates the student’s identity – in this case, Elena.
- The event ID is a unique identifier for this detection.
- Timestamp records the date of the vocabulary use – in this case, the date on the paper.
- The event type records whether this a vocabulary word use or a cognate detection. This is a vocabulary word use.
- Vocab word ID identifies which word – or its cognate – was used.
- Instructed? Determines whether the word has been taught yet. Deviation was a part of Module 4.
- Number of transfers tracks how many times the student has used the word authentically. This is the first use, so the number of transfers is one.
- Source tracks which context the detection occurred in: schoolwork or social media. Schoolwork could always be uploaded or pulled through the API from Canvas, while social media sources are pulled from Mastodon.
- Correct? records the results of the natural language processing. Elena has in fact used “deviation” correctly.

The Civil Rights Act of 1964

April 30, 2024
Social Studies
Elena

Lorem ipsum ...

Some activists thought waiting for rights to be granted voluntarily was not a *realistic* strategy, that those in power would never voluntarily extend rights to Black people.

Lorem ipsum ...

Another class assignment is fed in – one in which another vocabulary word is detected.

User ID	Event ID	Timestamp	Event type	Vocab word ID	Instructed ?	Number of transfers	Source	Correct?
8675	48151642	20240407	001 – Vocabulary word used	2024_4_3 - deviation	Yes	1	Schoolwork	Yes
8675	48151643	20240430	001 – Vocabulary word used	2024_4_6 - realistic	Yes	1	Schoolwork	Yes

Once again, this is analyzed by our system and an event is recorded. Much of the data is the same as for the last detection, but the event ID, timestamp, and identified vocabulary word have all changed.



Laney2007 6 weeks ago

There are alternatives to the metal detectors?? I hope Ridgedale **begins using** the AI-powered weapon detectors! I'm tired of my phone triggering an alert every time I enter the building.

[Xtract One: Alternatives for Weapons Detection](#)

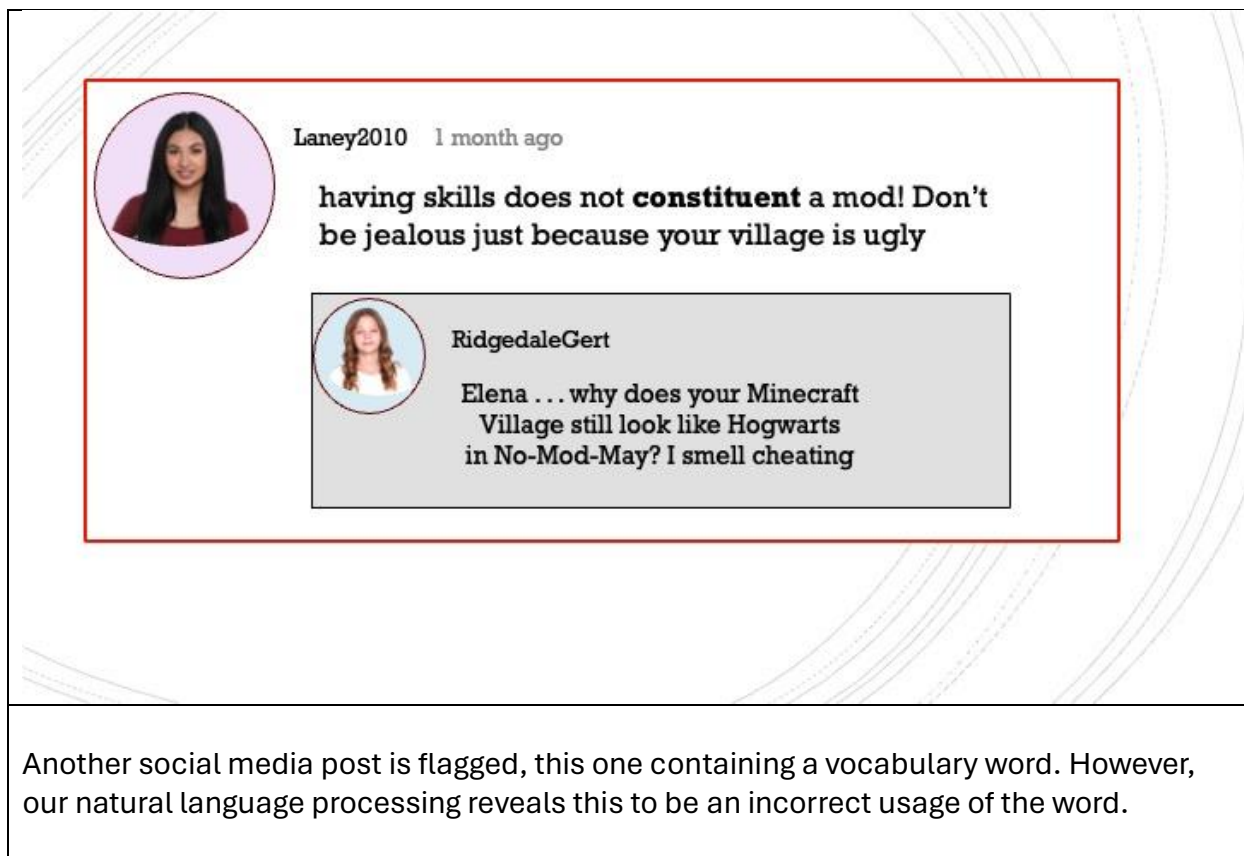
We also feed in Elena's social media posts.

In this one, our system detects a cognate – “begins using” – for “implements.” It uses natural language processing to identify that the word “implements” would fit here, and records this as a different type of event . . .

User ID	Event ID	Timestamp	Event type	Vocab word ID	Instructed ?	Number of transfers	Source	Correct?
8675	48151642	20240407	001 – Vocabulary word used	2024_4_3 - deviation	Yes	1	Schoolwork	Yes
8675	48151643	20240430	001 – Vocabulary word used	2024_4_6 - realistic	Yes	1	Schoolwork	Yes
8675	48151644	20240503	002 – Cognate used	2024_4_5 - implements	Yes	0	Social media	NA

... a cognate used event

Note the number of transfers remains 0. This is tied to the vocabulary word “implements,” but this is an unsuccessful transfer, so that counter is not incremented.



The image shows a screenshot of a social media post and its reply. The post is by user 'Laney2010' and is dated '1 month ago'. The text of the post is: 'having skills does not **constituent** a mod! Don't be jealous just because your village is ugly'. The word 'constituent' is bolded. Below the post is a reply from user 'RidgedaleGert' which says: 'Elena ... why does your Minecraft Village still look like Hogwarts in No-Mod-May? I smell cheating'. The entire screenshot is enclosed in a red border.

Laney2010 1 month ago

having skills does not **constituent** a mod! Don't be jealous just because your village is ugly

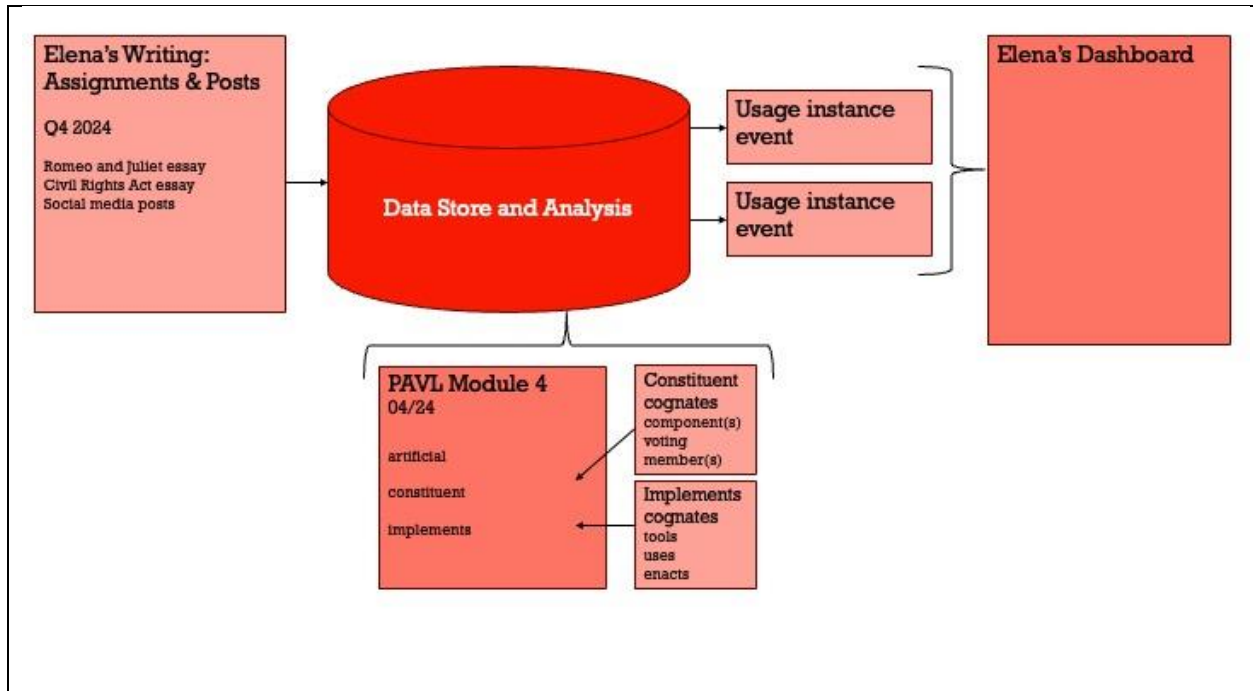
RidgedaleGert

Elena ... why does your Minecraft Village still look like Hogwarts in No-Mod-May? I smell cheating

Another social media post is flagged, this one containing a vocabulary word. However, our natural language processing reveals this to be an incorrect usage of the word.

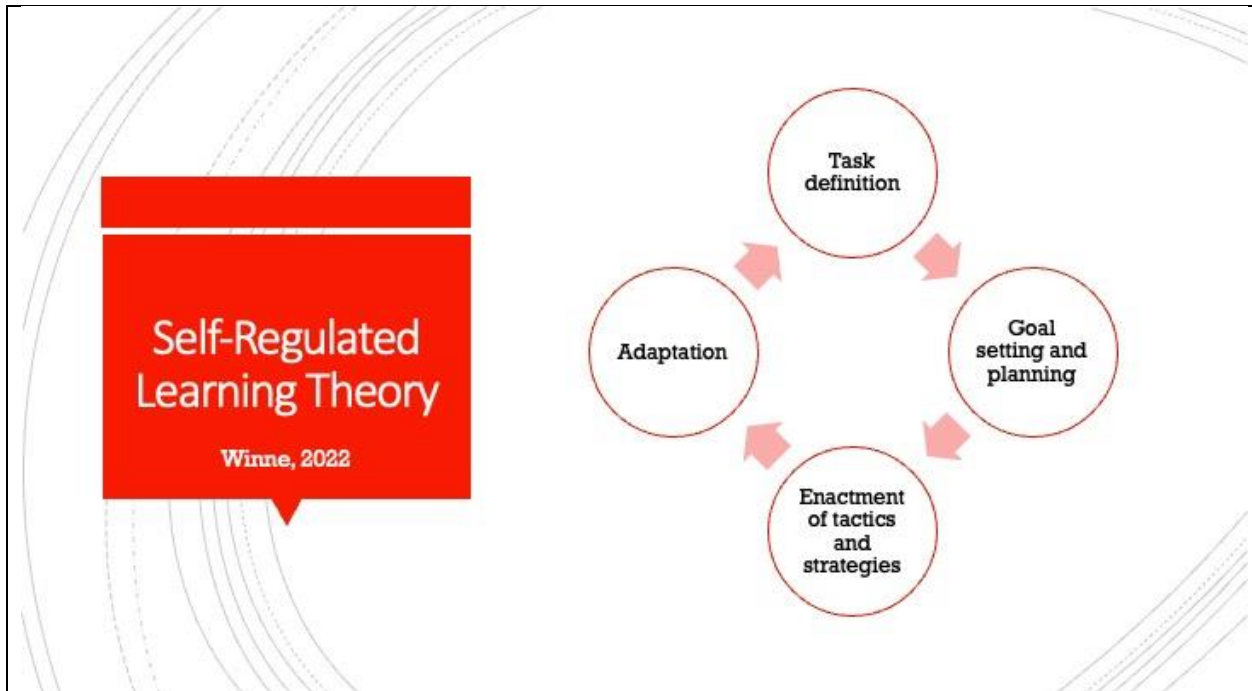
User ID	Event ID	Timestamp	Event type	Vocab word ID	Instructed ?	Number of transfers	Source	Correct?
8675	48151642	20240407	001 – Vocabulary word used	2024_4_3 - deviation	Yes	1	Schoolwork	Yes
8675	48151643	20240430	001 – Vocabulary word used	2024_4_6 - realistic	Yes	1	Schoolwork	Yes
8675	48151644	20240503	002 – Cognate used	2024_4_5 - implements	Yes	0	Social media	NA
8675	48151645	20240510	001 – Vocabulary word used	2024_4_2 - constituent	Yes	0	Social media	No

This also fails to increment the transfer counter.



These events form the data for Elena's Dashboard.

The goal of this dashboard is to provide feedback on and encourage self reflection on her vocabulary use. We will nudge her to use particular words and to use the words correctly. To accomplish this, I organized the dashboard around self-regulated learning theory.



Self-regulated learning theory has four main stages.

- Task Definition - students survey relevant resources and constraints
- Goal setting and planning - make goals and plans to approach them
- enactment of tactics and strategies – carry out plans
- Adaptation – review data and feedback on the enactment. Revise task definition, goals, and plans accordingly.

The process then repeats throughout the learning.

Elena's Dashboard (1 / 3)

Successful college students use these 591 vocabulary words in their academic writing (Malmström, 2018). Make sure you're prepared for a great first year by using these words now!

Category	Count	Percentage
Untaught Words	491	83%
Taught Words	100	17%

What's your reason for studying vocabulary?

I

How will you use these words in class writing? In social media posts?

I

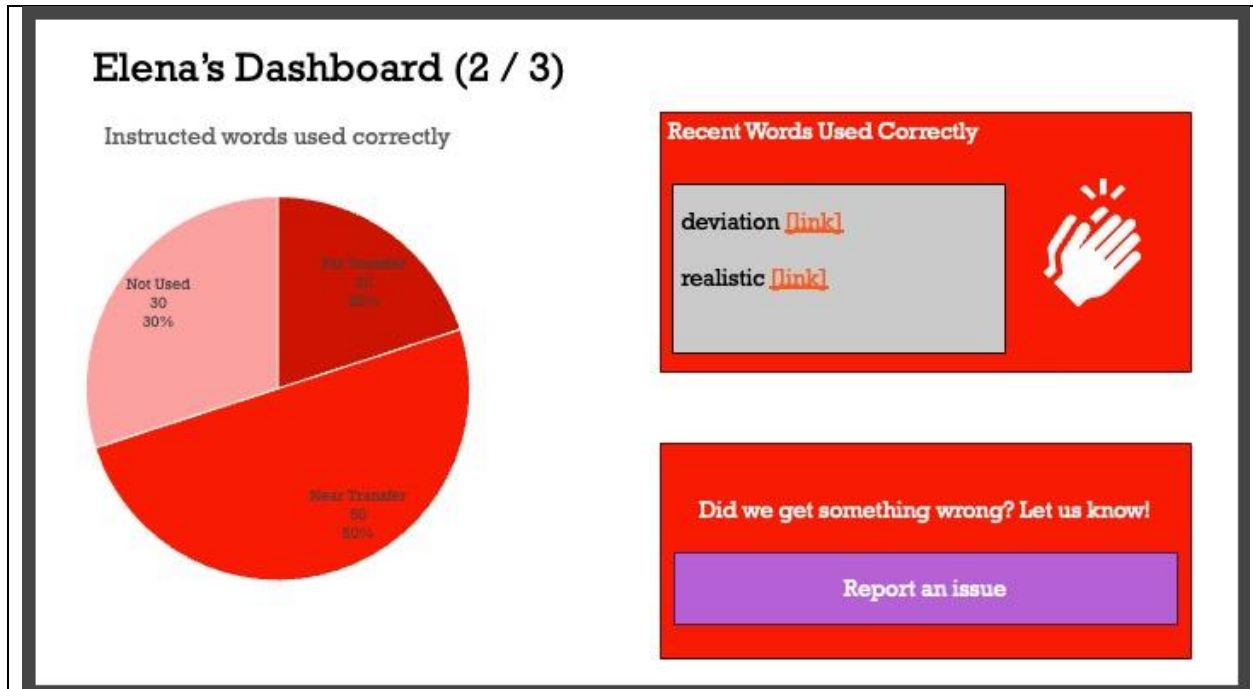
Here is the top of the dashboard, focusing on the first two stages of self-regulated learning: task definition and goal setting.

At the top of the page, we have a motivating statement explaining why this exists and why Elena should care. It is explicit, to be clear for the student.

I have presented a pie chart showing the total number of words to be learned before college – 591 – but also setting the expectation that only 100 of these have been taught so far. Hopefully this communicates to Elena that the expectation is only for her to use those 100 words.

For explicit goal setting, the dashboard asks students:

- What's your reason for practicing academic vocabulary?
- How can you use these vocabulary in your classes? Social media posts?



When Elena scrolls down a little, she will see information about her enactment of tactics and strategies.

- A pie chart shows the number and percent of instructed words she has used correctly. The pale section shows words she has not used. The dark color shows words with far transfer and the lighter color, near.
- A widget displays the words she has used correctly recently – in the last month.

Additionally, this section includes a button that applies to the whole dashboard: a way for the user to report incorrect data. “Allowing users to update their user models when they potentially disagree or find discrepancies in data or results of data analysis” was recommended by Matcha et al. (2020).

Elena's Dashboard (3 / 3)

Opportunity Words

Instead of "begins using" [\[link\]](#),
write "implements"

artificial	utilized
generalization	whether

Words to Review

constituent [\[link\]](#) def:

- (n) a member of a constituency
- (n) an essential part
- (adj) serving to compose a unit

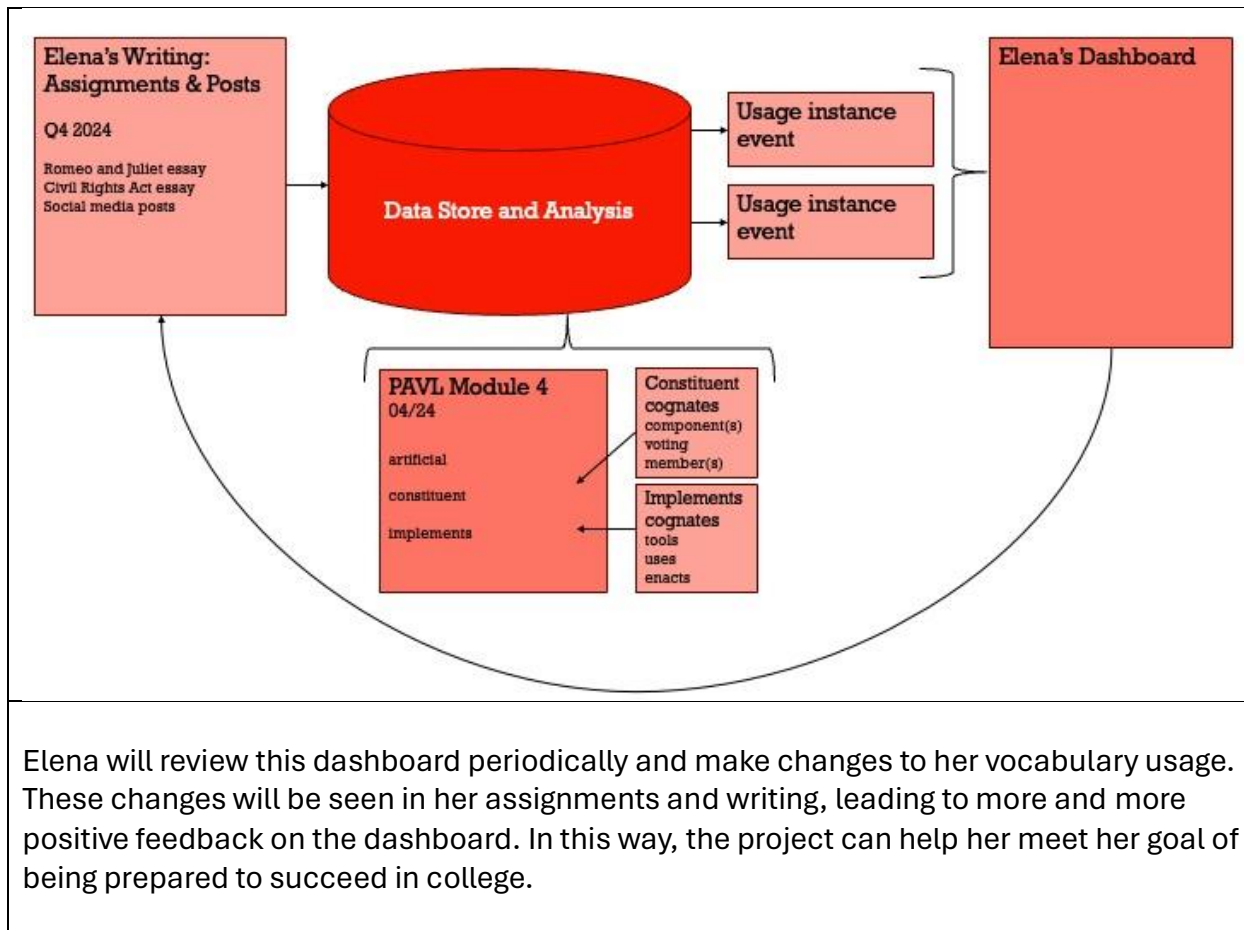
Which of these words will you use this week?

I

At the bottom of the dashboard is information about enactment and for adaptation.

For enactment, we show words that Elena has an opportunity to use: words that Elena used a cognate for recently, as well as learned vocabulary words that Elena has never used authentically in the observed contexts.

For adaptation, she is encouraged to select words from the target lists to use this week. For this to be effective, the program would likely need to send an email reminder about these words mid-week.





Sources

- Hopkins, R.F., Lyle, K.B., & Hieb, J.L. (2016). Spaced retrieval practice increases college students' short- and long-term retention of mathematics knowledge. *Educational Psychology Review*, 28, 853–873. <https://doi.org/10.1007/s10648-015-9349-8>
- Malmström, H., Pecorari, D., & Shaw, P. (2018). Words for what? Contrasting university students' receptive and productive academic vocabulary needs. *English for Specific Purposes*, 50, 28–39. <https://doi.org/10.1016/j.esp.2017.11.002>
- Nagy, W., & Townsend, D. (2012). Words as tools: learning academic vocabulary as language acquisition. *Reading Research Quarterly*, 47(1), 91–108. <https://doi-org.libproxy.lib.unc.edu/10.1002/RRO.011>
- Winne, P. (2022). Learning analytics for self-regulated learning. In *Handbook of Learning Analytics* (pp. 78–85). Society for Learning Analytics Research.

Thank you for reviewing my project!

I appreciate your time and consideration. If you have any questions or feedback, please send them to jamiestoryward@gmail.com. This was a fun project to work on and I would love to work on an implemented project tied to learning analytics.

To learn more about me and my work, visit my website, jamiestoryward.com. Scan the QR code below to visit.

