

Preliminary Analysis of the Spring 2025 Assessment of the Vertical Model of Writing Instruction at Salem State University

Presented by the Mary G. Walsh Writing Center Team

Amy Jo Minett
Al DeCiccio
Tanya Rodrigue

I) Introduction

In spring 2025, the Mary G. Walsh Writing Center team created and executed a final plan to assess the vertical model of writing instruction, a plan and assessment piloted in the spring of 2024. To accomplish the 2025 assessment, we completed the following tasks:

- We finalized the assessment tool, an adapted and localized version of the LEAP Value Rubric for Written Communication, which we also used for our 2024 pilot assessment (see addendum at end of report);
- We defined criteria for, then identified and anonymized, 120 final appropriate samples (from over 300 collected artifacts) across the three levels of the vertical model (W-I, W-II, W-III);
- We recruited an assessment team of faculty from across the disciplines, a number which was reduced from **eight faculty across seven disciplines (MCO, MSOE, Healthcare Studies, Sociology, English, Accounting and Finance, Biology)** to five faculty across four disciplines (**MSOE, English, Accounting and Finance, and Biology**), given a budget cut and a delay in final budget approval;
- We divided our final team of eight into pairs and strategically distributed artifacts to each pair, ensuring everyone read and scored papers from across several disciplines;
- We collectively assessed 120 artifacts.

To assess these artifacts, we met in two sessions: as a large assessment team for five hours in total; and with an additional three individual pair meetings for roughly three hours. Work also included individual essay scoring (roughly two hours per person), for a total of 10 hours+ of work per person.

At our first meeting, we reviewed (for a third time) and then contextualized and localized the LEAP Value Rubric for Written Communication, making it appropriate for the Salem State writing context (see addendum at end of report). We then practiced scoring essays followed by group discussion of scores in order to calibrate our scores for interrater reliability. Our **agenda for that first meeting** was as follows:

I) Introductions II)

Objectives:

- 1) To effectively prepare assessment team members to score writing artifacts through
 - Discussion and revision of scoring rubric;
 - Consistent application of rubric through practice;
 - Practice scoring/calibration;
 - Collaborative discussion and reflection; ■Ongoing feedback.
- 2) To use final data to begin to revisit and revise Vertical Model as necessary.

III) Process: Rubric Understanding and Use

- 1) Ensure understanding of rubric wording and clarify ambiguities
- 2) Practice Scoring/Practice Calibration
 - a) Score one row at a time;
 - b) Discuss and calibrate scores one row at a time;
 - c) Practice with multiple sample essays.
- 3) Create Inter-Rater Reliability: Raters read and score paper according to each criterion. Raters note or annotate “dings and diamonds” to explain scores.
- 4) Discuss (large group) and calibrate scores for practice essays by striving to come within two points for final scores.

After rater training at our first meeting, each pair worked on their own and met at least twice to calibrate their scores. Essays whose final scores were more than two points apart were placed in a “Reconcile” folder. They were then discussed and scores finalized at final meeting.

The resulting data our work produced is both rich in insights and limited in its portrayal of the Vertical Model of Writing at Salem State. In this report, we share our preliminary analysis of initial findings and their possible implications—along with the limitations which invite us to continue our assessment in subsequent academic years. We write “preliminary” because, given the lateness of the assessment project’s conclusion midMay of 2025, we simply haven’t had adequate time or resources to examine our data indepth in this report. However, we can make several key observations from our analysis which provide a solid base for future analysis and continued assessment.

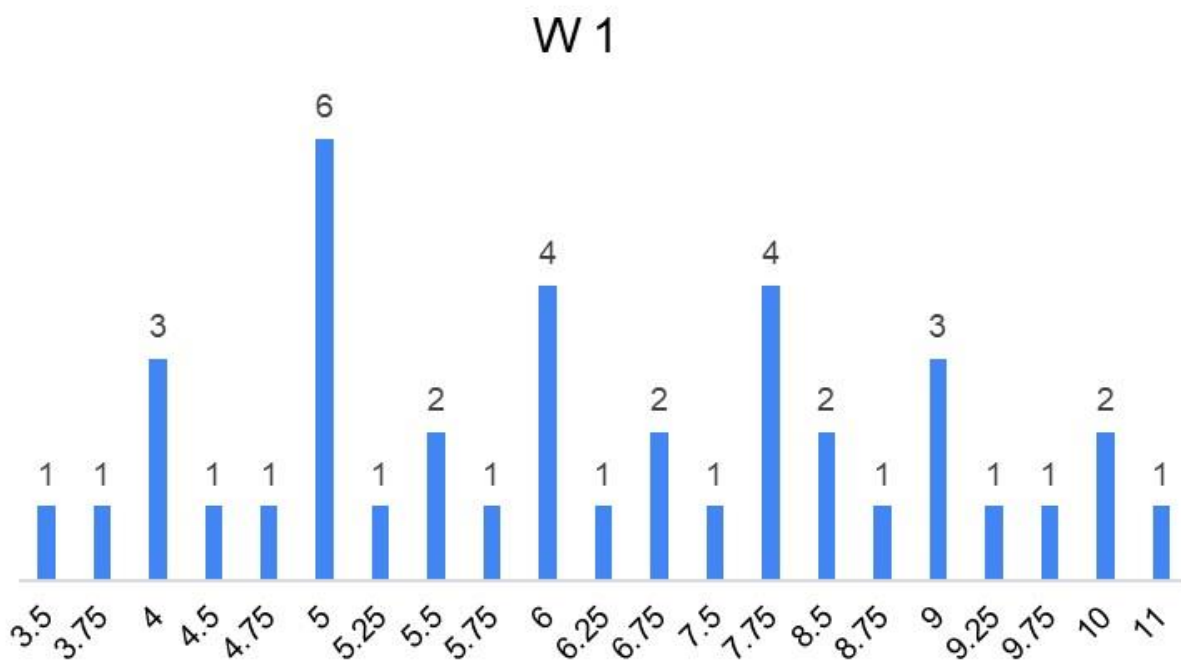
II) Preliminary Findings:

W-I:

In the below bar graph, Figure 1, W-I, essay scores can be seen ascending from 3.5

(the lowest score in this level) to 11, the highest score a W-I essay received. The numbers at the top of each bar indicate how many essays received that score out of a total number of 40.

Figure 1:



Observations: First, assuming that to meet the requirements of the W-1, a student would have to achieve a minimum of “1” for each of the five rubric criterion, **“5” or above for their total score**, we see that seven out of forty student essays—or **17.5%-did not** successfully meet the requirements of the W-I, while **82.5% met or exceeded the minimum requirements**. From this sample (n=40), we can comfortably generalize that *most* artifacts demonstrated that writers met the requirements for the W-1.

Within those criteria, we find the following average scores for rubric criteria in Figure 2.

Figure 2:

W-1	AVERAGE
Purpose and Occasion for Writing	1.557894737
Content Development	1.563157895
Genre and Disciplinary Conventions	1.389473684
Sources and Evidence	1.3
Control of Syntax and Mechanics	1.463157895

While the average differences in all cases are mere 10ths of a point, like Olympic scoring, those 10ths of a point help us discern critical factors to consider in more depth:

- **Sources and evidence** need the most additional curricular and in-class attention in arguably *all* first year writing and subsequent courses, but most significantly, in the W-I course. Specific qualitative comments from rater notes strongly support this observation:

Sources hard to discern;

Evidence (patchwriting?); [as quoted from the University of British Columbia]: [Patchwriting happens when a writer attempts to paraphrase a source by changing few things, but the result is still too similar to the original and “patches” of the original writing remains the same. One could reword every part of another person’s work, but if the ideas and sentences are still in the same order, it is still patchwriting.](#)

Trouble entering and explaining quote;

Works cited not cited, more than any other error - no citation;

Citations of different styles, refs in color, minimal in-text citations;

No sources used in places where needed;

There were no in-text citations, which limited the strength of many of the claims made;

Quotes had no citations associated with them. I did not know what sources were from where in this presentation;

Students are across the board struggling with source use--how to incorporate sources, how to cite sources, how to synthesize and analyze sources.

Recommendations: Offer increased professional development opportunities on integrating source use in all levels of Vertical Writing Model but especially the foundational W-I. Emphasize source use more in curriculum (summarizing, analyzing, synthesizing, citing sources). Create a palette of W-I course syllabi with a range of pedagogical approaches (e.g., the multigenre research course; “Writing Back to the New York Times”) which offer faculty teaching the W-I both **course variety and assignment consistency** in W-I courses).

- The second criterion which suggests a need for greater curricular and classroom attention is “**Genre and Disciplinary Conventions,**” the definition of which can be found in the rubric (see addendum at end of report for detailed explanation). Certainly, part of the problem we encountered with scoring for this criterion was **not having the assignment sheets accompanying the written artifacts** we were scoring: raters therefore struggled to ascertain what those conventions

should be. This is a **limitation** we'll review in more detail in the final section of this report. Importantly, the writers' struggles to make **the genre and discipline** explicit speak also to the intersections between criterion one, "**Purpose and Occasion for Writing**," which includes "**considerations of audience**." Some raters believed assignments should be able to stand on their own as understandable pieces of writing without the actual assignment; others viewed artifacts more generously, believing that well-designed assignment sheets accompanying the artifact could considerably enlighten the writer (and therefore reader) on purpose, occasion, audience, and—in turn—**genre and disciplinary conventions**. As we know both approaches to reading and rating have merit, for now, we acknowledge here a clear need for 1) better-designed assignments which both communicate genre and disciplinary conventions and how those conventions can be made clearer with specific statements of purpose; 2) increased focus on explicit teaching of those conventions and statements of purpose in the W-I and subsequent courses; *and* 3) emphasized work/practice to help students clarify their audience and purpose, which can foster recognition of genre and disciplinary conventions. Here are relevant statements from rater notes that address the issues raters (and writers) encountered—primarily, uncertainty about genre and disciplinary conventions:

Disciplinary conventions did not seem to be in use [assumes reader had knowledge of expected disciplinary conventions]

Not sure genre conventions are met;

No title, subheadings, v. long first paragraph;

The first-person and casual nature of the writing was distracting, but it was unclear if that was fine in the context of the genre and disciplinary convention.

If this was written for a journalism/blogging class or assignment, it probably works. But the hyperlinks and visuals were distracting.

Recommendations: Offer increased professional development opportunities to all faculty which explore and encourage articulation of genre and disciplinary conventions across disciplines—and how those conventions intersect with statements of purpose. Devote increased curricular attention and class time to exploring genre and disciplinary conventions with students—and how those conventions can be clarified with statements of purpose.

- The third criterion we will address here is "**Control of Syntax and Mechanics**." It is both interesting and heartening that these primarily sentence-level issues were scored higher by raters than source use and genre/disciplinary conventions. This fact demonstrates that raters were attending more to [higher-order order concerns](#)

[\(the preceding categories\) than lower-order concerns](#), that is, the communication of meaning and purpose over strict adherence to standard

academic English with standardized, prescriptive grammar. At the same time, rater comments for this criterion are worth extensive further analysis, which will happen in subsequent study:

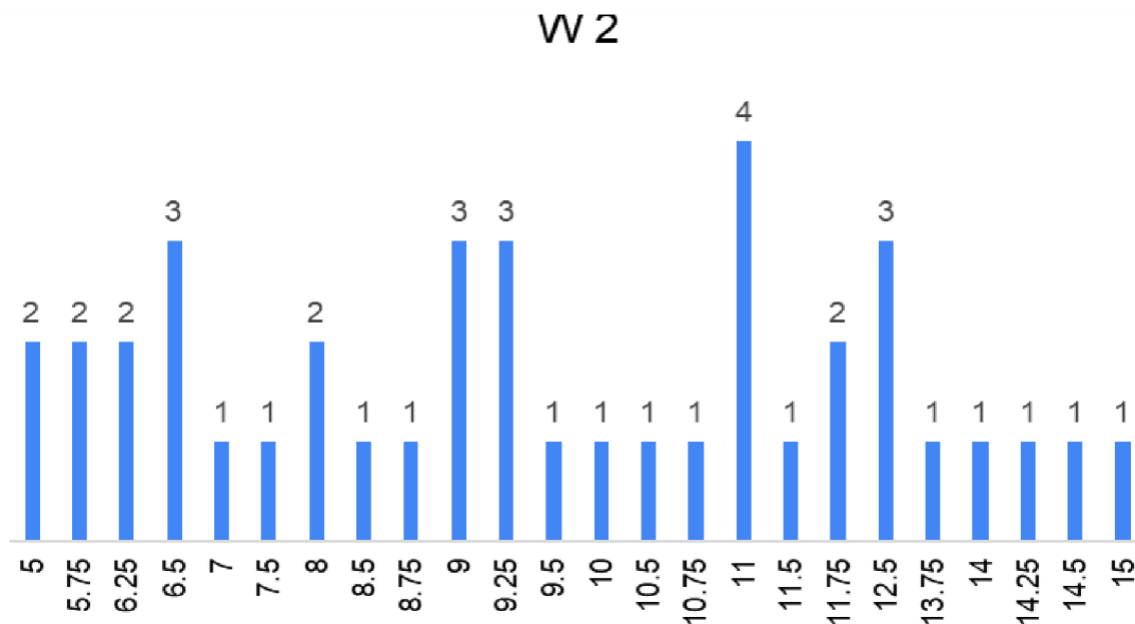
The writing was poor, with many spelling and grammatical errors.
Poor syntax and mechanics; Lapses
in syntax, mechanics;
This paper needs to be edited.

Recommendations: Celebrate and continue to foster in professional development opportunities and weekly Writing Tips understanding of [higher order concerns and lower order concerns](#). Conduct more in-depth study of faculty *and student* perceptions of nonstandard academic language. Provide professional development opportunities exploring *the stages* of additional and academic language development and acquisition; then provide professional development opportunities for teaching [research-based best practices for teaching syntax, grammar, and editing](#). Implement these practices in writing instruction at all levels.

W-II

In the below bar graph, Figure 3, W-II, essay scores can be seen ascending from 5 (the lowest score in this level) to 15, the highest score any essay could receive. The numbers at the top of each bar indicate how many written artifacts received that score out of a total of 40.

Figure 3:



Observations: From this bar graph, Figure 3, we can see, first, that *all* W-2 writing artifacts met or exceeded the minimal score for **meeting requirements of the W-I**, evidence suggesting writers' continued development after the W-I. However, while we haven't *finally* determined a cut-off score indicating that students have met the requirements of the W-2, for this initial report, we have settled upon 10 out of 15 maximum points, or a minimum score of 2 ("Milestone") for each of the five categories of the rubric. With 10 as our current cut score, Figure 3 shows that 22 out of 40 artifacts received a score of less than 10, or 55%: that is, **55% did not meet the minimum cut score, while 45% did**. We also note that **all average scores for discrete rubric criteria are higher than averages for the W-I**.

Within the specific rubric criteria, we find the following average scores in Figure 4.

Figure 4:

	AVERAGE
Purpose and Occasion for Writing	2.08125
Content Development	2.06875
Genre and Disciplinary Conventions	1.6875
Sources and Evidence	1.8625
Control of Syntax and Mechanics	3.893617021

Similar to the breakdown of average scores for the W-1 in Figure 2, here we see the following:

- Again, scores for **"Sources and Evidence"** and **"Genre and Disciplinary Conventions"** are the lowest in the group, though this time **"Genre and Disciplinary Conventions"** scored lower than **"Sources and Evidence"** by .175 of a point. Qualitative comments from raters' notes echo limitations and concerns expressed under the W-I comments.

Under **"Genre,"** we find considerable rater uncertainty over genres students may be attempting:

Headings (blog posts or Annotated Bibliography?). Formal and informal. What should the conventions be? HARDEST PART OF RUBRIC? generous (of blog posts);

Attempts--seems a hodgepodge rather than deliberate attention to conventions;

Not clear purpose except to explain lab-made diamonds? Genre?

I feel like the assignment is: tell me a story about your research...right??

Under “**Source Use**,” rater comments are again similar to those under the W-I:

Does not adequately cite sources in text or even link;

But no Works Cited or references. Just links.

There are limited citations and those used are not done correctly.

Citations need work, but the sources appear good. They could be integrated more effectively.

Direct quotes with no citations. Refers to articles as first article, second article, etc. but unsure which this refers to. Lack of using articles to advance the argument in a linear fashion.

Recommendations: Given the lower rate of students meeting the W-II cutoff scores, intervention with/review of “**Source Use**” seems particularly urgent at the W-II level, regardless of citation style and/or discipline. Moreover, the difference between average scores for “Source Use” at the W-I and W-II levels is only .3875, suggesting increased urgency around and attention to source use in the W-I as well. Again, at this stage, we recommend 1) offering increased professional development opportunities on integrating source use in all levels of Vertical Writing Model but especially the foundational W-I. Likewise, we recommend 2) emphasizing source use more in curriculum and classroom (summarizing, analyzing, synthesizing, citing sources). As for “**Genre and Disciplinary Conventions**,” we recommend again offering increased professional development opportunities to all faculty which explore and encourage articulation of genre and disciplinary conventions across disciplines—and how those conventions intersect with statements of purpose. We further recommend devoting increased curricular attention and class time to exploring genre and disciplinary conventions with students—and how those conventions can be clarified with statements of purpose. As with the W-I, part of the problem we encountered with scoring for this criterion was **not having the assignment sheets accompanying the written artifacts**: raters therefore struggled to ascertain what those conventions should be and noted their relationship to a writer’s clear purpose (or lack thereof), an issue we’ll address further under “Limitations.”

- In the category of “**Syntax and Mechanics**,” the most striking observation may be that writers of W-2 artifacts scored a significant **2.43 percent higher** than writers in the W-I category, testament, we believe, to writers’ continued development through stages of acquiring academic and standard English. While scores were higher, rater perceptions of this category at the W-II level illuminate further:

This reads like a W-1 paper. Some grammatical errors, lack of evidence, and no clear thesis. Needed editing.

Comma splices/sentence boundary;
Minor surface errors;
Some lapses in mechanics and grammar (word usage);
Grammar and mechanics okay (didn't like occasional slang ["screwed"]).

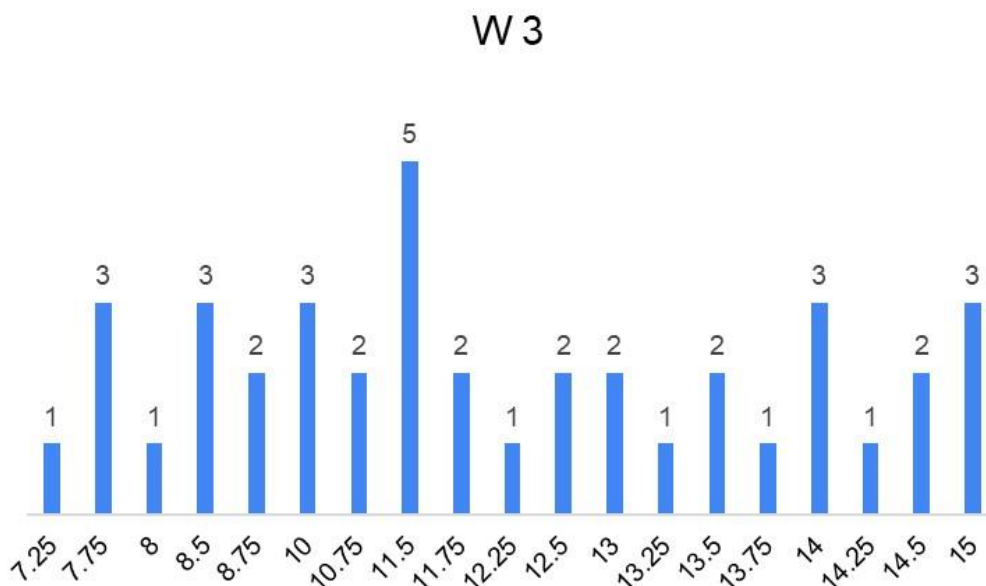
Anecdotally, these and other comments begin to indicate raters scrutinizing tone and word choice more intensely, perhaps, than they did with the W-I artifacts, an observation which invites closer analysis at a later date of the qualitative data.

Recommendations: As with our recommendations for the W-I, we again should celebrate and continue to foster understanding of [higher order concerns and lower order concerns](#) in professional development opportunities and weekly Writing Tips. At the same time, we are eager to conduct more in-depth study of faculty *and student* perceptions of non-standard academic language. We should continue to provide professional development opportunities exploring *the stages* of additional and academic language development and acquisition; then provide professional development opportunities for teaching [research-based best practices for teaching syntax, grammar, and editing](#). We should implement these practices in writing instruction at all levels.

W-III

In Figure 5, W-III, essay scores can be seen ascending from 7.25 (the lowest score in this level) to 15, the highest score possible on the rubric. The numbers at the top of each bar indicate how many written artifacts received that score (out of 40 total).

Figure 5:



Observations: From this bar graph, Figure 5, we can see, first, that while *all* W-III

writing artifacts met or exceeded the minimal score for **meeting requirements of the W-I, 10 W-3 artifacts, or 25%, do not meet the cut score for the W-2**. These numbers represent the greatest gap between levels (W-II and W-III), suggesting the need for greater intervention at the W-II level in all areas (see **recommendations**). As the Mary G. Walsh Writing Center team has loosely designated “13” as the cut score for meeting the W-III requirements (all cut scores need further debate and discussion), it is also of import that **25 out of 40 artifacts do not meet the cut score, or 62.5%**.

Within the specific rubric criteria, we find the following average scores in Figure 6.

Figure 6:

	AVERAGE
Purpose and Occasion for Writing	2.611111111
Content Development	2.404761905
Genre and Disciplinary Conventions	2.301587302
Sources and Evidence	2.120967742
Control of Syntax and Mechanics	2.266129032

From these averages, we observe:

- First, that the **average scores for discrete rubric criteria are higher than averages for the W-II** with the exception of “**Syntax and Mechanics**,” which surprisingly drops by **1.63 points**. Just below that average falls the average for “**Source Use**,” the lowest average of the group, again. In other words, “**Sources and Evidence**” is the category with the lowest average across three levels, and the W-III level scores are just .4334 points higher than the W-II averages. Here are rater comments pertinent to these two categories, beginning with “**Source Use**”:

No sources used in places where needed.

No sources cited in paper, two referenced at end.

Often doesn't cite.

Direct quotes may need more paraphrasing. The conventions of citations are not correct.

- Regarding **“Syntax and Mechanics,”** while the average score is lower than W-II averages, rater comments are somehow more generous, inviting further study of comments and perceptions at a later date:

Doesn't know semicolons :).

Lack of commas to control reading but strong moments (see "paramount")

This essay could use editing.

Some places where editing could be improved.

Recommendations: Given that **25 out of 40 artifacts do not meet the W-III readiness cut score, or 62.5%,** here we see an opportunity to provide vital support and scaffolding that can help students move from the W-II to the W-III fully prepared for the final course in the Vertical Model. The Mary G. Walsh Writing Center team are in discussion over various ways to address this gap. One approach may be to create a **portfolio as “gateway” assessment** between the W-II and W-III levels: students who do not demonstrate readiness for the W-III might then take **a one- or more credit support course (a “Rising Junior” course)** tailored to cultivate writers’ particular needs, from source use, to syntax and mechanics, to genre and disciplinary conventions (averages for which, for the W-III, rose around .6 of a point). Such a course would certainly bolster writers’ readiness for their final, most challenging, W-III writing course. And as before, we imagine offering more targeted professional development opportunities and writing tips that address areas where students are still struggling: workshops on **best practices for teaching source use, syntax and mechanics, and genre and disciplinary conventions.**

III) Limitations of This Report

There are several limitations to this preliminary study which invite further research of our current data and continued assessment of the Vertical Model. Limitations are outlined below.

One of the most critical limitations we faced is the time period in which these student papers were collected. Papers were collected at the end of the AY 23-24 and composed either in that year or within the two years prior. Since then, the use of AI has become more and more prevalent in higher education. If we were to collect papers to score this past academic year or next, we imagine there would be much more evidence of the use of AI in student writing. The role of AI in the teaching of writing and in writing intensive courses at Salem State demands ever greater attention as we move forward in thinking about best pedagogical practices and relevant goals and criteria for courses in the vertical model. Rater comments already question the influence of AI even in these artifacts:

*I feel like a robot wrote this? Vague with no integrated sources
More robot?*

A second limitation is that not all disciplines or schools are represented in this study. Most of the student artifacts are from the College of Arts and Sciences with some samples from the School of Education and the School of Business. By not examining papers in all colleges, we are limited in understanding student work and student needs across the entire institution.

In the same vein, another limitation is the faculty who assessed the writing in this study. The majority of raters on the assessment team were from the school of Arts and Sciences, English being the most represented (more than half of the assessment team). One faculty member was from the School of Education, one from the School of Business, one from Biology. Every discipline has its own expectations and understandings of what constitutes good and effective writing. Therefore, the findings may have been different with the inclusion of more faculty from across more disciplines. We strongly hope to continue this assessment on a larger scale, especially as we find even these very preliminary findings exciting in their instructiveness.

As discussed throughout the report's observations, we also realized as an assessment team that findings would be more informative if a) we had the accompanying assignment sheets with artifacts; and b) we knew we were collecting artifacts from the ends of semesters. An artifact from the beginning of the W-II course should not be expected to meet the same milestones as artifacts collected at semester's end. Thus, as an assessment team working within a limited time frame and with minimal support, we learned a tremendous amount which can richly inform future assessment projects, which we hope we are encouraged and supported to conduct. We have further developed data organization systems which can help us more easily track data and trends, both holistically and through discrete analytic ("primary trait") scoring.

IV) Conclusion

In short and in closing, what we have learned in this process—both about the Vertical Model *and* about assessment—should be a wealth of information we can continue to draw from and leverage. In addition to understanding targeted needs for professional development and curricular/classroom interventions and points of focus for teaching writing, we further generated substantive quantitative and qualitative data we are eager to analyze in much more detail than possible in this preliminary report. We look to the future, eager to continue learning and eager to continue supporting our students the best ways possible.

References

- Hammond, K. (n.d.). Teaching grammar in the 21st century classroom. Center for Professional Education of Teachers, Teachers College, Columbia. Retrieved May 31st, 2025, from <https://cpet.tc.columbia.edu/news-press/teaching-grammar-inthe-21st-century-classroom>
- Higher order concerns (HOCs) and lower order concerns (LOCs) (2025). Purdue Online Writing Lab. Retrieved May 31st, 2025, from https://owl.purdue.edu/owl/general_writing/mechanics/hocs_and_locs.html.
- Plagiarism, patch-writing, and paraphrasing (2023). University of British Columbia. Retrieved May 31st, 2025, from <https://students.ok.ubc.ca/wp-content/uploads/sites/90/2023/06/2AIM-Plagiarism-Patchwriting-ParaphrasingHandout-05182023.pdf>

**Addendum to Report: Rubric Adapted and Localized from LEAP Value Rubric for
Written Communication**

	Capstone 3	Milestones 2	Benchmark 1	Does Not Achieve Benchmark 0
Purpose and Occasion for Writing <i>Includes considerations of audience, purpose, and context for the writing task(s).</i>	Demonstrates a thorough understanding of audience and context and a focused purpose throughout text.	Demonstrates adequate consideration of audience and context and maintains a clear purpose throughout.	Demonstrates some consideration of context, audience, and purpose in places. Purpose may not be consistently clear.	
Content Development <i>Includes use of evidence and data for arguments and claims</i>	Consistently uses appropriate, relevant, and compelling content to demonstrate mastery of the subject and convey writer's understanding of development.	Uses mostly appropriate, relevant, and compelling content to explore ideas within the writing.	Uses some appropriate and relevant content to develop ideas in some parts of the work.	

Genre and Disciplinary Conventions <i>Includes formal and informal rules/conventions expected for writing in particular genres, forms,</i>	Demonstrates detailed attention to and successful execution of a wide range of conventions particular to a specific discipline and	Demonstrates consistent use of important conventions particular to a specific discipline and genre-including organization, content,	Demonstrates occasional use of expected conventions in areas of genre, organization, content, presentation,	
<i>and/or academic fields</i>	genre including organization, content, presentation, formatting, and stylistic choices.	presentation, and stylistic choices.	and stylistic choices.	
Sources and Evidence <i>Includes locating credible sources of evidence, incorporating and citing them appropriately in text(s)</i>	Demonstrates skillful use and citation of high-quality, credible, relevant sources to develop ideas that are appropriate for the discipline and genre of the writing.	Demonstrates consistent use and citation of credible, relevant sources to support ideas that are situated within the discipline and genre of the writing.	Demonstrates an attempt to use and cite credible, relevant sources to support ideas in the writing.	

Control of Syntax and Mechanics <i>Includes following grammatical rules (or breaking them intentionally) depending on the rhetorical situation; Using grammar and syntax intentionally and rhetorically to convey meaning with clarity, consistency, creativity, and fluency</i>	Uses graceful language/ rhetorical grammar that skillfully communicates meaning to readers with clarity, consistency, creativity, and fluency.	Uses language/rhetorical grammar that conveys meaning to readers with clarity, consistency and fluency.	Uses language/grammar that usually conveys clear meaning to readers.	
--	--	---	--	--