WASHBURN UNIVERSITY

Oral Communication Artifact Assessment Report 2024-25

Executive Summary

In Academic Year 2025, **21 Oral Communication artifacts** were scored by two separate raters using a revised version of the AAC&U Oral Communication VALUE rubric. The rubric consists of seven criteria:

- Introduction
- Organization
- Language
- Supporting Material
- Delivery - Voice
- Delivery - Nonverbals
- Conclusion

Each criterion was rated on a five-point scale: Capstone (4), Milestone (3), Milestone (2), Benchmark (1), and Not Present (0); there was also an option to mark an artifact as unscorable.

In the event the average of the scores across all criteria from the 1st and 2nd raters differed by 1.00 or greater, a 3rd rating was generated to normalize the scores. For Academic Year 2025, **six artifacts (28.6%) required a 3rd score** from an additional rater who did not initially rate the artifacts.

Overall, more than three-quarters of the artifacts were scored on average at the **Milestone (3) level or higher (85.7%)**, with all remaining artifacts being scored at the Milestone (2) level (14.3%). Over one-quarter of the artifacts were scored at the Capstone (4) level (28.6%), and **none of the artifacts were scored at or below the Benchmark (1) level**.

When examining overall normalized ratings by criteria on a five-point scale from 4.00 - 0.00, mean ($\bar{x}$) scores ranged from 3.12 - 1.83, and the **overall mean score was 2.54** (standard deviation (σ) = 1.01). Criterion-specific highlights include:

- Highest mean score: Language ($\bar{x}$ = 3.12, σ = 0.80)
- Second highest mean score: Delivery - Voice ($\bar{x}$ = 2.71, σ = 0.86)
- Lowest mean score: Conclusion ($\bar{x}$ = 1.83, σ = 1.17)

When examining the distribution of criteria scores, **Language accounted for one-quarter of criteria scores at the Capstone (4) level (25.0%)**; conversely, Conclusion accounted for over half of the criteria scores at the Benchmark (1) level (60.0%).

When compared to previous assessments of Oral Communication Artifacts, **six of the seven criteria had mean score averages higher in 2024-25** when compared to the otherwise most recent artifact review in 2021-22 (Delivery - Nonverbals was the exception).

Introduction

Oral Communication was assessed during the 2024-25 Academic Year using a modified version of the Association of American Colleges and Universities (AAC&U) Oral Communication Valid Assessment of Learning in Undergraduate Education (VALUE) rubric (see Appendix A). The AAC&U VALUE rubrics were developed by teams of educational professionals and include the most frequently identified criteria of learning for different learning outcomes.

Washburn University (Washburn) uses a modified version (2022) of the AAC&U VALUE rubrics for assessing Oral Communication every three years. Artifacts in video format are collected from students who conducted course presentations and are scored by two or more independent raters using the Outcomes Assessment Projects, formerly Aqua, by the Watermark software platform.

Review Process

Washburn faculty members were invited to attend the calibration training conducted on May 6, 2025. The seven faculty who attended the training were assigned 22 artifacts. The artifacts were assigned to be reviewed by two independent raters on seven criteria: Introduction, Organization, Language, Supporting Material, Delivery - Voice, Delivery - Nonverbals, and Conclusion. These seven criteria were scored on a five-point scale: Capstone (4), Milestones (3), Milestone (2), Benchmark (1), and Not Present (0). Reviewers could also assign the status of unscorable to those artifacts that were not appropriate for the purpose of oral communication assessment. One artifact was designated as unscorable, resulting in 21 artifacts being assigned for review.

Of the seven reviewers who participated in the artifact review process, six scored seven (7) artifacts and one did not score any artifacts. Overall, a total of 21 artifacts were reviewed two times each, for a total of 42 reviews. When the average difference in scores between the 1st and 2nd raters differed by 1.00 or greater, a 3rd rater was utilized to normalize the ratings. After the initial round of scoring, six artifacts (28.6%) did meet the criteria which required a 3rd rating. The process of ascertaining which artifacts required a 3rd reviewer is outlined in the next section.

3rd Rater Review Process

The differences in averaged ratings ranged from 1.71 - 0.14. The greatest percentage of averaged ratings differed by 0.14 (28.6%). The distribution of scores was positively skewed (right-skewed), meaning the mean of 0.64 was greater than the median of 0.57.

The distribution of averaged rating differences and descriptive statistics are shown in Table 1, Figure 1, and Table 2. A total of six artifacts required a 3rd rating, highlighted in blue in Table 1. These artifacts had an averaged rating difference range of 1.71 - 1.00.

Table 1. Rating Differences

Rating Difference	Frequency	Percent
0.14	6	28.6%
0.29	2	9.5%
0.43	2	9.5%
0.57	2	9.5%
0.71	2	9.5%
0.86	1	4.8%
1.00	1	4.8%
1.14	2	9.5%
1.29	1	4.8%
1.43	1	4.8%
1.71	1	4.8%
Total	**21**	**100%**

Figure 1. Percent of Rating Differences
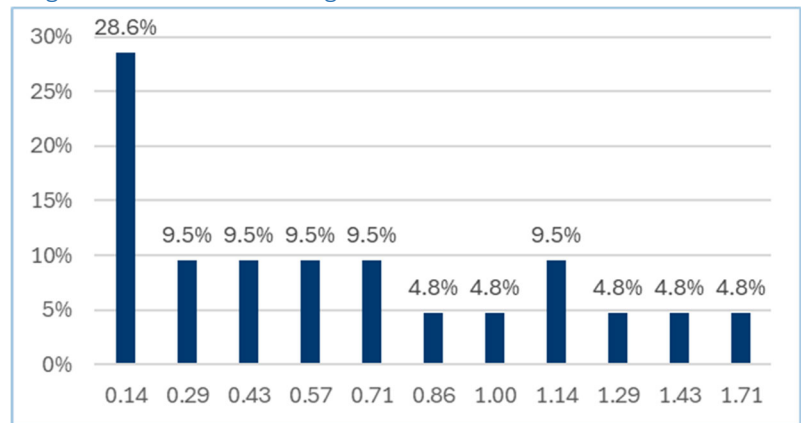
Table 2. Descriptive Statistics for Rating Differences

Mean	Median	Mode	St. Dev
0.64	0.57	0.14	0.49

The six artifacts that required a 3rd rating were scored by a faculty member who did not initially rate those artifacts. The 3rd rater's scores were compared to the first two raters' scores, and the score farthest from the 3rd rater's score was removed from further analysis. See Table 3 for means, standard deviations, and both minimum and maximum differences between averaged ratings for artifacts which either needed or did not need a 3rd rating.

Table 3. Descriptive Statistics for 3rd Rated Artifacts

	Mean	St. Dev	Minimum	Maximum
No 3rd Rating Needed	0.38	0.25	0.14	0.86
3rd Rating Needed	1.29	0.25	1.00	1.71

The mean of averaged rating differences for the artifacts which did not require a 3rd rating was 0.38, compared to 1.29 for the artifacts which did. The standard deviation for both groups of artifacts was the same at 0.25. The minimum difference the artifacts not requiring a 3rd rating was 0.14; the maximum difference for artifacts which did require a 3rd rating was 1.71. The distribution of normalized averaged rating differences and descriptive statistics after incorporating scores from the 3rd rater are shown in Table 4, Figure 2, and Table 5.

Table 4. Normalized Rating Differences

Rating Difference	Frequency	Percent
0.00	1	4.8%
0.14	6	28.6%
0.29	2	9.5%
0.43	4	19.0%
0.57	5	23.8%
0.71	2	9.5%
0.86	1	4.8%
Total	**21**	**100%**

Figure 2. Percent of Normalized Rating Differences
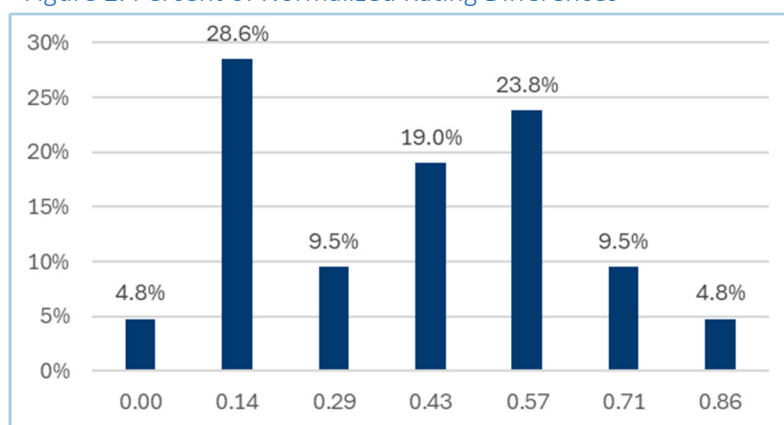
Table 5. Descriptive Stats for Normalized Rating Differences

Mean	Median	Mode	St. Dev
0.39	0.43	0.14	0.24

Results

Differences by Criterion

The AAC&U Oral Communication VALUE rubric defines Oral Communication as "*a prepared, purposeful presentation designed to increase knowledge, to foster understanding, or to promote change in the listeners' attitudes, values, beliefs, or behaviors.*" The rubric contains seven criteria: Introduction, Organization, Language, Supporting Material, Delivery - Voice, Delivery - Nonverbals, and Conclusion. Washburn's Oral Communication artifacts had each of these seven criteria rated on a five-point scale: Capstone (4), Milestones (3), Milestone (2 (Basic)), Benchmark (1), and Not Present (0).

The 21 Oral Communication artifacts were rated by two reviewers on seven dimensions for a total of 147 scores. These scores were reviewed to examine differences in averaged ratings per criterion. Table 6 provides descriptive statistics for the differences in averaged ratings by criterion.

Table 6. Descriptive Statistics for Differences in Ratings by Criterion

	Introduction	Organization	Language	Supporting Material	Delivery - Voice	Delivery - Nonverbals	Conclusion	Total Difference
Mean	0.71	0.57	0.43	0.86	0.48	0.43	0.90	0.63
St. Dev.	0.56	0.60	0.60	0.96	0.51	0.60	0.77	0.68
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max	2.00	2.00	2.00	3.00	1.00	2.00	2.00	3.00

The mean difference for averaged ratings was greatest for Conclusion and Introduction, meaning the ratings, on average, differed by 0.90 and 0.71 with standard deviations of 0.77 and 0.56, both respectively. Conversely, the smallest mean differences were for Language and Delivery - Nonverbals, both differing by 0.43 with a standard deviation of 0.60. The minimum averaged rating was 0.00 (or no difference) across all criteria, whereas the maximum averaged rating was 3.00 for Supporting Materials.
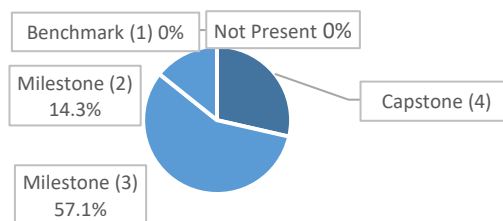
Distribution of Scores

The scores from the two raters were averaged to provide a normalized score for each Oral Communication artifact. Table 7 provides the overall distribution of 147 scores for 21 Oral Communication artifacts. The ranges of average scores were defined as 4.00-3.01 for Capstone (4), 3.00-2.01 for Milestone (3), 2.00-1.01 for Milestone (2), and 1.00-0.01 for Benchmark (1). Those criteria that were not rated were designated as Not Present.

Table 7. Descriptive Data and Statistics for Overall Averaged Ratings

	Capstone (4)	Milestone (3)	Milestone (2)	Benchmark (1)	Not Present	Mean (sd)
	4.00 - 3.01	3.00 - 2.01	2.00 - 1.01	1.00 - 0.01	0.00	
Overall (n = 21)	6 (28.6%)	12 (57.1%)	3 (14.3%)	0 (0%)	0 (0%)	2.54 (1.01)

Most artifacts (57.1%) were scored at Milestone (3), followed by Capstone (4) (28.6%). Three artifacts (14.3%) were rated at Milestone (2), while no artifacts were rated at Benchmark (1) or Not Present (0). See Figure 3 for a visual representation.

Figure 3. Distribution of Scores for Overall Oral Communication Artifacts

Ratings by Criterion

The distribution of average ratings and descriptive statistics for each of the seven criteria are in Table 8. The table shows how the artifacts (n = 21) were rated on each of the seven criteria (total of 147 scores).

Table 8. Descriptive Data and Statistics for Overall Averaged Ratings by Criterion

	Capstone (4)	Milestone (3)	Milestone (2)	Benchmark (1)	Not Present (0)	Mean (sd)
	4.00 - 3.01	3.00 - 2.01	2.00 - 1.01	1.00 - 0.01	0.00	
Introduction	5 (23.8%)	9 (42.9%)	6 (28.6%)	1 (4.8%)	0 (0%)	2.55 (0.99)
Organization	3 (14.3%)	11 (52.4%)	7 (33.3%)	0 (0%)	0 (0%)	2.57 (0.77)
Language	8 (38.1%)	11 (52.4%)	2 (9.5%)	0 (0%)	0 (0%)	3.12 (0.80)
Supporting Material	5 (23.8%)	8 (38.1%)	6 (28.6%)	2 (9.5%)	0 (0%)	2.52 (1.11)
Delivery - Voice	6 (28.6%)	8 (38.1%)	7 (33.3%)	0 (0%)	0 (0%)	2.71 (0.86)
Delivery - Nonverbal	4 (19.0%)	6 (28.6%)	10 (47.6%)	1 (4.8%)	0 (0%)	2.45 (0.92)
Conclusion	1 (4.8%)	5 (23.8%)	9 (42.9%)	6 (28.6%)	0 (0%)	1.83 (1.17)
Overall (n = 147)	32 (21.8%)	58 (39.5%)	47 (32.0%)	10 (6.8%)	0 (0%)	2.54 (1.01)

The averaged ratings of the Introduction, Organization, Supporting Material, and Delivery – Nonverbal scores all fell within the 2.57 - 2.45 range, indicating these criteria scores were in the mid-point of the Milestone (3) range. The averaged ratings of the Delivery - Voice scores came out to 2.71, indicating these criteria scored on the higher end of the Milestone (3) range. The averaged ratings of the Language scores came out to 3.12, making it the only criterion scoring in the Capstone (4) range; conversely, the averaged ratings of the Conclusion scores came out to 1.83, making it the only criterion in the Milestone (2) range.

For the Language criterion, over 90% of the averaged scores were in the Capstone (4) and Milestone (3) range (90.5%), contributing to the highest overall mean score of the criteria; additionally, Language had the second lowest standard deviation of 0.86, indicating its scores were among the most closely clustered around its mean score. Conversely, Conclusion had the highest percentage of ratings at or below the Benchmark (1) range (28.6%). Refer to Figure 4 for a visual representation of the distribution of ratings by each criterion (excluding Not Present). Supporting Material and Delivery - Nonverbals had a normal distribution of ratings. Introduction, Organization, Language, and Delivery - Voice were positively skewed, while Conclusion was negatively skewed.

Figure 4. Distribution of Averaged Scores by Criterion
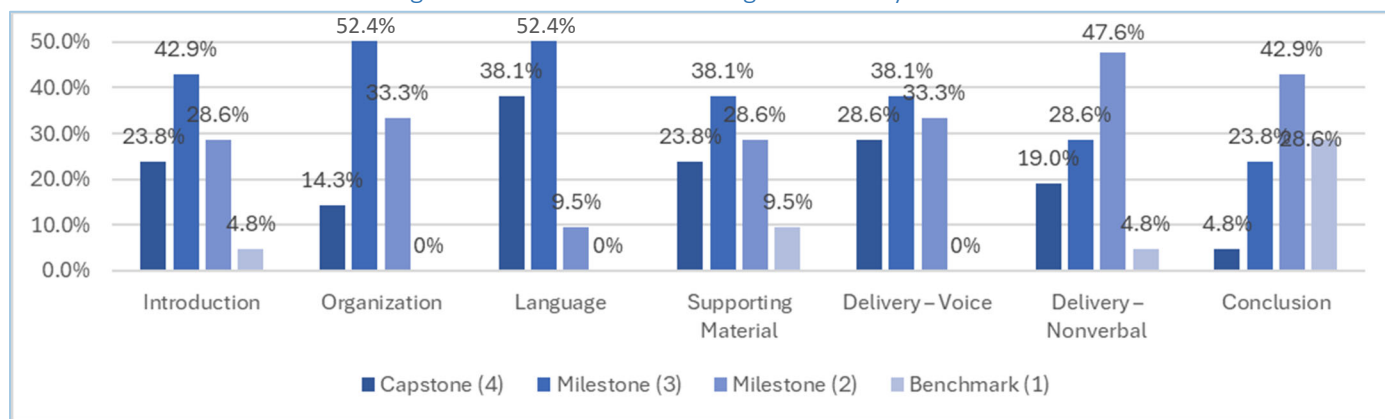
Comparison to Previous Years Results

Washburn piloted review of Oral Communication artifacts in 2014-15 (*n* = 35), from which only 8.6% of the artifacts attained an overall averaged score at or above the Milestone (3) performance level. In 2018-19 (*n* = 199) and 2021-22 (n = 42), 77.4% and 81.0% of the artifacts were at the same level, respectively. For the current year, 2024-25 (*n* = 421), 85.7% of the artifacts attained an overall averaged score at or above Milestone (3); additionally, the mean of overall normalized ratings has continued to increase, from 2.29 in 2018-19, to 2.37 in 2021-22, and 2.54 in 2024-25.

Six of the seven criteria had their highest mean scores for averaged ratings by criterion in 2024-25; Delivery - Nonverbals was the exception, which saw its highest ranking in 2021-22. Delivery - Nonverbals had been showing consistent growth each time Oral Communication was assessed, with the decrease from 2.52 in 2021-22 to 2.45 in 2024-25 being the first ever decrease for its mean score rankings. Otherwise, notable increases in mean scores include Introduction, which increased to 2.55 in 2024-25 from its previous highest ranking of 2.31 in 2021-22, and Conclusion, increasing to 1.83 in 2024-25 from 1.51 in 2018-19. Finally, 2024-25 marks the first instance of a criterion's mean score for averaged ratings being within the Capstone (4) range, with Language's 3.12 being the highest ranking of any criterion throughout Washburn's history of assessing Oral Communication artifacts.

Areas of Consideration and Limitations

Following the 2018-19 Oral Communication artifact rating, reviews were adjusted to not be done in pairs, but rather by having each rater receive a random set of artifacts to rate. Removing repeated pairs of raters for artifacts requiring a 3rd rater and the overall decreased number of artifacts requiring a 3rd rater over time could indicate that assigning raters a random set of artifacts has been normalizing scores over time.

Although 85.7% of Oral Communication artifacts were scored at or above the Milestone (3) level, the mean score of overall averaged ratings was 2.54, indicating that scores were between the Milestone (3) and Milestone (2) levels; additionally, the standard deviation of 1.01 is the highest of the four times Oral Communication artifacts have been assessed, indicating these scores are the least clustered around the mean when compared to previous assessments.

Language was the highest rated criteria with most ratings at or above the Milestone (3) level (90.5%); additionally, there was a minimal amount of Milestone (2) ratings (9.5%) – as well as no ratings at or below the Benchmark (1) level – notable due to all other criteria having more instances of Milestone (2) ratings. This trend was noted during the 2021-22 assessment of Oral Communication artifacts as well, reinforcing the need to ensure the rubric is being thoroughly reviewed and ensuring the categories for Language are distinct enough during future training sessions.

The methodology of determining 3rd ratings and normalizing scores was modified in 2018-2019, which is important to note when comparing Oral Communication assessment results longitudinally. Specifically, the threshold for artifacts requiring a 3rd rater was increased from a difference of more than 1.00 or greater (≤ 1.01) to a difference of 1.00 or greater (≤ 1.00), potentially causing a need for more 3rd raters in 2018-19. Regarding normalizing rater's scores: beginning in 2018-19, the 3rd rater's score was compared to the first two raters' scores, with the score farthest from the 3rd rater's score being removed from further analysis. Prior to this, the methodology was to average the 3rd rater's score with the first two raters' scores, resulting in more varied normalized scores due to the normalized ratings for artifacts requiring a 3rd rater averaging the scores of three raters instead of two. Consequently, the ranges of the scoring levels for artifacts were modified in 2018-19 to accommodate the normalization of the two ratings as opposed to three.

Finally, although efforts were made to draw from a representative profile of Washburn students for generalizability purposes, only Senior level students' oral presentations from seven departments were reviewed; consequently, results cannot be generalized to the entire Senior student population.

ORAL COMMUNICATION VALUE RUBRIC

for more information, please contact value@aacu.org

The VALUE rubrics were developed by teams of faculty experts representing colleges and universities across the United States through a process that examined many existing campus rubrics and related documents for each learning outcome and incorporated additional feedback from faculty. The rubrics articulate fundamental criteria for each learning outcome, with performance descriptors demonstrating progressively more sophisticated levels of attainment. The rubrics are intended for institutional-level use in evaluating and discussing student learning, not for grading. The core expectations articulated in all 15 of the VALUE rubrics can and should be translated into the language of individual campuses, disciplines, and even courses. The utility of the VALUE rubrics is to position learning at all undergraduate levels within a basic framework of expectations such that evidence of learning can by shared nationally through a common dialog and understanding of student success.

The type of oral communication most likely to be included in a collection of student work is an oral presentation and therefore is the focus for the application of this rubric.

Definition

Oral communication is a prepared, purposeful presentation designed to increase knowledge, to foster understanding, or to promote change in the listeners' attitudes, values, beliefs, or behaviors.

Framing Language

Oral communication takes many forms. This rubric is specifically designed to evaluate oral presentations of a single speaker at a time and is best applied to live or video-recorded presentations. For panel presentations or group presentations, it is recommended that each speaker be evaluated separately. This rubric best applies to presentations of sufficient length such that a central message is conveyed, supported by one or more forms of supporting materials and includes a purposeful organization. An oral answer to a single question not designed to be structured into a presentation does not readily apply to this rubric.

Glossary

The definitions that follow were developed to clarify terms and concepts used in this rubric only.

- Orientation/Central message: The main point/thesis/"bottom line"/"take-away" of a presentation. A clear central message is easy to identify; a compelling central message is also vivid and memorable. The central message is established throughout the entire speech, but effective speakers orient the audience to the speech topic and purpose in the introduction. Orientation refers to those parts of the introduction that explain the speaker's interest in the subject and/or rationale for engaging in the project. Orientation may also refer to why the topic is important for the audience. Very complex topics may require additional information in the beginning of the presentation to fully orient a lay audience.
- Delivery techniques: Posture, gestures, eye contact, and use of the voice. Delivery techniques enhance the effectiveness of the presentation when the speaker stands and moves with authority and purpose, looks more often at the audience than at his/her speaking materials/notes, uses the voice expressively, and uses few vocal fillers ("um," "uh," "like," "you know," etc.). Vocal variety refers to intentional changes in pitch, tone, volume, and pace intended to enhance or clarify a speaker's words.
- Language: Vocabulary, terminology, and sentence structure. Language that supports the effectiveness of a presentation is appropriate to the topic and audience, grammatical, clear, and free from bias. Language that enhances the effectiveness of a presentation is also vivid, imaginative, and expressive.
- Organization: The grouping and sequencing of ideas and supporting material in a presentation. An organizational pattern that supports the effectiveness of a presentation typically includes an introduction, one or more identifiable sections in the body of the speech, and a conclusion. An

organizational pattern that enhances the effectiveness of the presentation reflects a purposeful choice among possible alternatives, such as a chronological pattern, a problem-solution pattern, an analysis-of-parts pattern, etc., that makes the content of the presentation easier to follow and more likely to accomplish its purpose.

- Supporting material: Explanations, examples, illustrations, statistics, analogies, quotations from relevant authorities, and other kinds of information or analysis that supports the principal ideas of the presentation. Supporting material is generally credible when it is relevant and derived from reliable and appropriate sources. Supporting material is highly credible when it is also vivid and varied across the types listed above (e.g., a mix of examples, statistics, and references to authorities). Supporting material may also serve the purpose of establishing the speakers credibility. For example, in presenting a creative work such as a dramatic reading of Shakespeare, supporting evidence may not advance the ideas of Shakespeare, but rather serve to establish the speaker as a credible Shakespearean actor.

	Capstone 4 (Advanced)	Milestones 3 (Milestone (3))	Milestones 2 (Basic)	Benchmark 1 (Beginning)	Not Present 0
Introduction	Excellent orientation of audience to topic; clear topic statement; previews main points.	Good orientation of audience to topic; discernible topic statement; previews main points.	Orientation of audience to topic is mundane; topic statement unclear; preview provides little direction for audience.	No orientation of audience to topic; abrupt jump into body of speech.	Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.
Organization	Very well organized; main points clear; effective transitions.	Organizational pattern is evident, main points are apparent; transitions present between main points.	Organizational pattern somewhat evident; main points are present but not distinct; transitions are present but are minimally effective.	Speech did not flow well; speech was not logically organized; transitions present but not well formed.	Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.
Language	Language is exceptionally clear, vivid; free from grammar errors and inappropriate usage.	Language appropriate to the goals of the presentation; free from grammar errors and inappropriate usage.	Language selection adequate; some errors in grammar; language at times misused (e.g., slang, awkward structure).	Grammar and syntax need to be improved as can level of language sophistication.	Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.
Supporting Material	All key points are well supported with a variety of credible material.	Main points were supported using appropriate material.	Points were generally supported; a greater quality of material needed.	Most points were not supported; a greater quantity and quality of material needed.	Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.
Delivery-Voice	Excellent use of vocal variation; intensity and pacing; vocal expression natural and enthusiastic; avoids fillers.	Good vocal variation and pace; appropriate vocal expression; few if any fillers.	Demonstrates some vocal variation; enunciates clearly and speaks audibly; generally avoids fillers.	Sometimes uses a voice too soft or articulation too indistinct for listeners to comfortably hear; often uses fillers.	Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.
Delivery-Nonverbals	Almost no reliance on notes; posture, gestures, facial expression and eye contact well developed, natural, and display high level of poise and confidence. All components above should be polished.	Almost no reliance on notes; posture, gestures and facial expressions are suitable for the speech; speaker appears confident. The speaker faces the audience (camera) with almost no distracting movement toward a visual aid.	Some reliance on notes, but has adequate eye contact; generally avoids distracting mannerisms. Speaker mostly faces audience (camera) with limited nonverbal distractions.	Speaker relies heavily on notes; nonverbal expressions stiff and unnatural. Speaker faces a visual aid or wall instead of the audience (camera); or has nonverbal distractions (hitting hand on desk, swiveling in chair, turning back and forth toward a computer).	Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.
Conclusion	Develops a clear review of main points; ends with strong and memorable closing statement.	Clear review of main points; clear closing statement	Provides some review of main points; closing technique can be strengthened.	Conclusion lacks clarity, trails off, ends in a tone not consistent with the rest of the speech.	Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.