

Quick Guide to Summarizing and Displaying Program Assessment Data from Student Work Evaluated with a Rubric

This quick guide was prepared by the WSU Office of Assessment for Curricular Effectiveness (ACE) and is intended to help WSU programs and faculty consider strategies for summarizing and presenting data collected from the evaluation of student work with a rubric as part of program-level assessment. ACE is also available to collaborate with WSU undergraduate degree programs to analyze and create visual displays of assessment data to engage faculty in discussions of assessment results. Contact us at ace.office@wsu.edu for more information.

Introduction

Program-level assessment data provide a means to look at student performance in order to offer evidence about student learning in the curriculum, provide information about program strengths and weaknesses, and guide decision-making. In this context, a rubric is a scoring tool that identifies component skills and knowledge for the targeted program learning outcomes, with a rating scale that provides information about the level of student performance.

While rubrics come in many forms and no one format is best for every situation, the evaluation of student work with a rubric typically produces **quantitative assessment data**. Analytic, holistic, and single point rubrics include a rating scale to show the degree to which the things you are looking for in an assessment are present. These rubrics may use a variety of rating scales (i.e., “exceeds expectations, meets expectations, below expectations”; “outstanding, very good, adequate, marginally adequate, inadequate”; “almost always, often, sometimes, rarely”). This kind of scaled data from a rubric is considered ordinal (where numbers are assigned to ordered categories). While less commonly used in higher education, a checklist rubric indicates the presence of things you are looking for in an assignment (a checklist rubric for a website development project might include: “each page includes a last updated date”, “contact information is present”, etc.). Yes/no response data from a checklist rubric would be considered nominal/categorical data, while a summed score of all checklist items could be interval/ratio data. For more information about quantitative data, see ACE’s [Quick Guide to Analyzing Quantitative \(Numeric\) Assessment Data](#).

Before You Begin

There is no “one size fits all” approach to analyzing assessment data, but there are some ways to make it more approachable. It’s best to start thinking about your data analysis plan when you are first identifying your assessment questions and determining how you will collect the needed information. It is important to match the analysis strategy to the type of information that you have and the kinds of assessment questions that you are trying to answer. In other words, decisions about how to analyze assessment data are guided by what assessment questions are asked, the needs and goals of the audience/stakeholders, as well as the types of data available and how they were collected.

Before you start to summarize and display the data, you should also be familiar with the basic data collection processes, including how the data were collected, who participated, and any known limitations of the data, as this can help you make an informed decision about what the data can reasonably reveal. Other factors to consider may include: How was the random sampling/size determined? Were well-established, agreed-upon criteria used for assessing the evidence for each outcome? How were raters normed/calibrated? Did multiple raters review each piece of evidence? Has this measure been pilot tested

and refined? As a good practice, a short written description of the data collection processes, number of participants, and a copy of any instrument used (i.e. rubric) should accompany the data analysis file, data summary, and/or final report.

Examples of Questions in Context: Student Work Evaluated with a Rubric

One of the first steps in planning a data summary is to review the purpose of the project and the assessment questions guiding it, and consider the audience for the summary. Questions may be achievement based, i.e., “At what level are students performing when they graduate?” or “Did the students meet the target on a particular learning outcome?” Questions may also reflect change across time or differences across groups, i.e., “Has our program improved over time?” or “Does one group of students learn as well as other students?”

The example questions below may be meaningful for data collected from student papers, performances, or projects evaluated with a rubric. For each of these questions, the following pages contain example scenarios and strategies for summarizing and presenting the data collected using frequency distributions and measures of central tendency (i.e., mean, mode, median).

Comparisons between Outcomes

- At what levels are students performing on the learning outcomes? [*Frequency distribution*]
- How many students are performing at an acceptable level for each learning outcome? [*Grouped frequency distribution*]
- What is the “typical” level of student performance for each learning outcome? [*Measure of central tendency*]

Comparisons between Groups

- At what levels are students from different groups (demographic, learning environment, etc.) performing on the learning outcomes? [*Frequency distribution*]
- How many students from different groups (demographic, learning environment, etc.) are performing at an acceptable level for each learning outcome? [*Grouped frequency distribution*]
- What is the “typical” level of performance for students from different groups (demographic, learning environment, etc.) for each learning outcome? [*Measure of central tendency*]

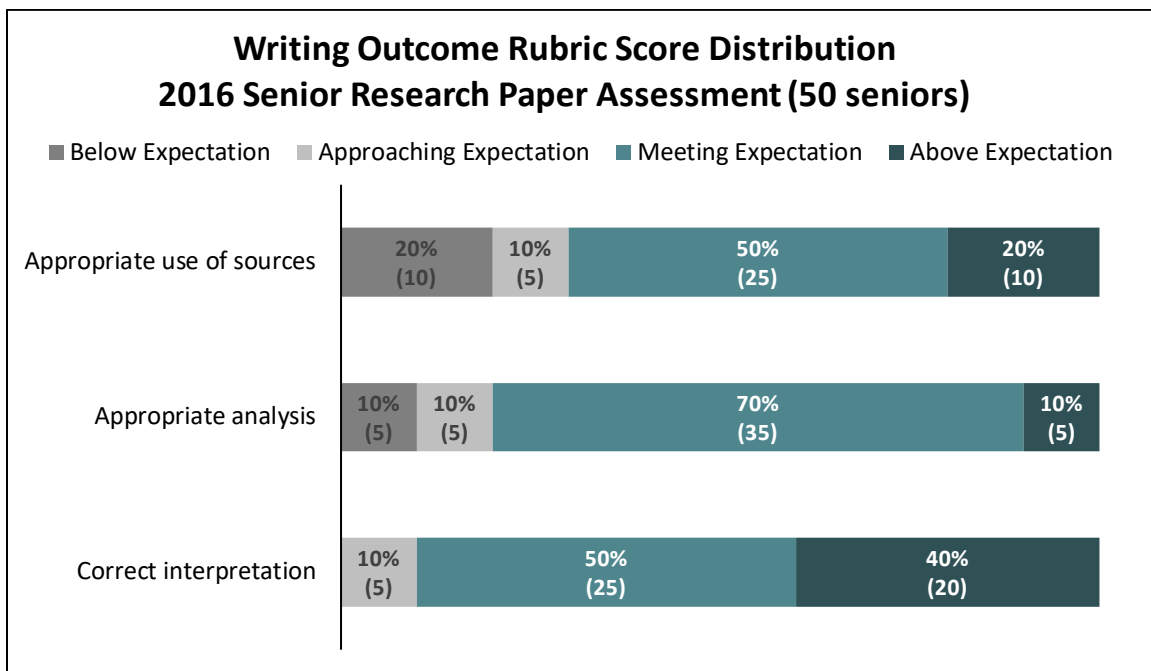
Comparisons over Time

- At what levels are students performing on the learning outcomes over time? [*Frequency distribution*]
- How many students are performing at an acceptable level for each learning outcome over time? [*Grouped frequency distribution*]
- What is the “typical” level of student performance for each learning outcome over time? [*Measure of central tendency*]

At what levels are students performing on the learning outcomes?

Example: A program used a common rubric to evaluate their writing outcome using senior research papers. The program wants to know how many students are below, approaching, meeting, or above the level of expectation for each rubric criterion (i.e. appropriate use of sources, appropriate analysis, and correct interpretation).

- **Frequency distributions** are useful when you want to examine the number/percentage of students across multiple levels of performance



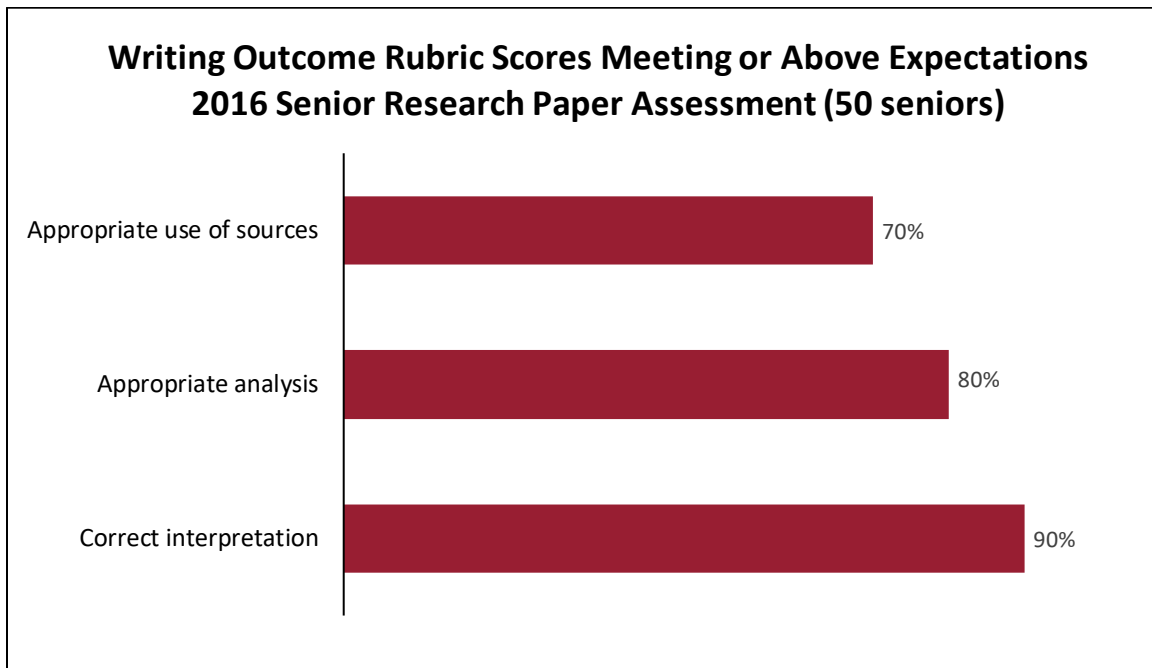
Writing Outcome Rubric Score Distribution 2016 Senior Research Paper Assessment (50 seniors)

Criteria	% (#) of seniors			
	Below Expectation	Approaching Expectation	Meeting Expectation	Above Expectation
Appropriate use of sources	20% (10)	10% (5)	50% (25)	20% (10)
Appropriate analysis	10% (5)	10% (5)	70% (35)	10% (5)
Correct interpretation	0% (0)	10% (5)	50% (25)	40% (20)

How many students are performing at an acceptable level for each learning outcome?

Example: A program used a common rubric to evaluate their writing outcome using senior research papers. The program wants to know how many students are performing at an acceptable level (i.e. meeting or above expectation) for each rubric criterion (appropriate use of sources, appropriate analysis, and correct interpretation).

- **Grouped frequency distributions** are useful when you want to examine the number/percentage of students who are in adjoining levels of performance (i.e., combining adjacent frequency distribution categories)



Writing Outcome Rubric Scores Meeting or Above Expectations 2016 Senior Research Paper Assessment (50 seniors)

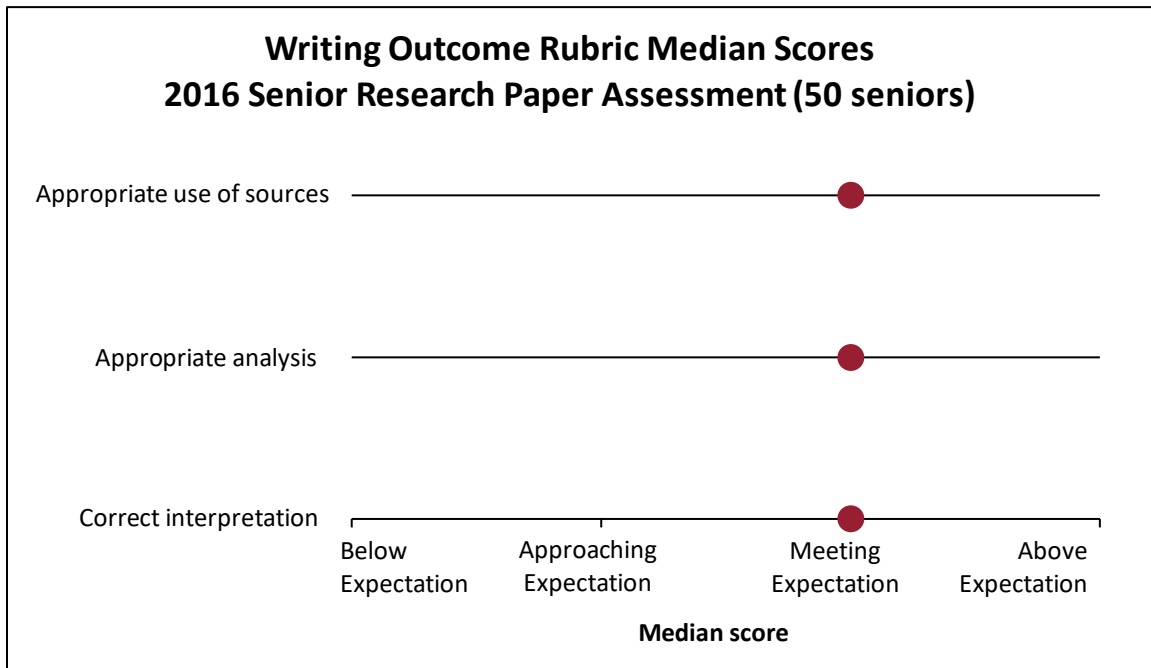
Criteria	% (#) of seniors "meeting or above expectation"
Appropriate use of sources	70% (35)
Appropriate analysis	80% (40)
Correct interpretation	90% (45)

What is the “typical” level of student performance for each learning outcome?

Example: A program used a common rubric to evaluate their writing outcome using senior research papers. The program wants to know the “typical” level of student performance for each rubric criterion (appropriate use of sources, appropriate analysis, and correct interpretation).

- **Measures of central tendency** (i.e., mean, mode, median) are useful when you want to examine the “typical” level of performance

Note: Under certain conditions, some measures of central tendency become more appropriate to use than others. For more information on how the level of measurement can dictate which analysis methods are appropriate, see ACE’s [Quick Guide to Analyzing Quantitative \(Numeric\) Assessment Data](#).



Writing Outcome Rubric Median Scores 2016 Senior Research Paper Assessment (50 seniors)

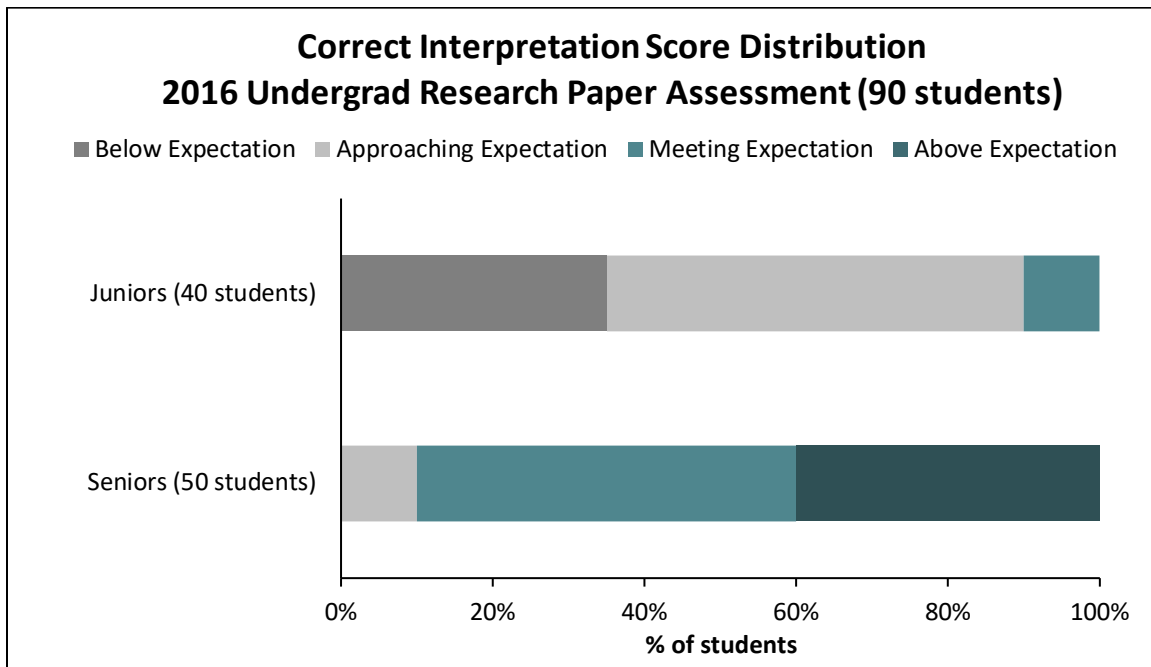
Criteria	Median score
Appropriate use of sources	3
Appropriate analysis	3
Correct interpretation	3

Note: Ordinal scale categories were coded as follows: 1-Below Expectation, 2-Approaching Expectation, 3-Meeting Expectation, 4-Above Expectation

At what levels are students from different groups (demographic, learning environment, etc.) performing on the learning outcomes?

Example: A program used a common rubric to evaluate their writing outcome using research papers. The program wants to know how many juniors and seniors are below, approaching, meeting, or above the level of expectation for a student at the end of the undergraduate experience for the “correct interpretation” rubric criterion.

- **Frequency distributions** are useful when you want to examine the number/percentage of students across multiple levels of performance



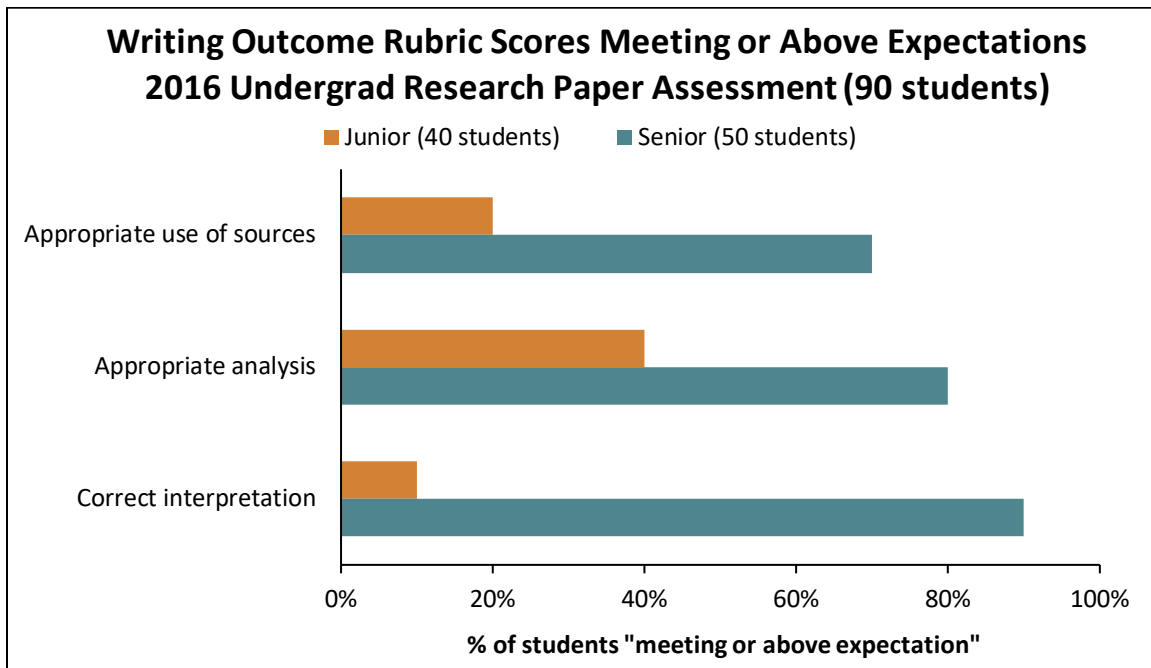
Correct Interpretation Score Distribution 2016 Undergrad Research Paper Assessment (90 students)

Class Level	% (#) of students			
	Below Expectation	Approaching Expectation	Meeting Expectation	Above Expectation
Juniors (40 students)	35% (14)	55% (22)	10% (4)	0% (0)
Seniors (50 students)	0% (0)	10% (5)	50% (25)	40% (20)

How many students from different groups (demographic, learning environment, etc.) are performing at an acceptable level for each learning outcome?

Example: A program used a common rubric to evaluate their writing outcome using research papers. The program wants to know how many juniors and seniors are performing at an acceptable level for a student at the end of the undergraduate experience (i.e., meeting or above expectation) for each rubric criterion (appropriate use of sources, appropriate analysis, and correct interpretation).

- **Grouped frequency distributions** are useful when you want to examine the number/percentage of students who are in adjoining levels of performance (i.e., combining adjacent frequency distribution categories)



Writing Outcome Rubric Scores Meeting or Above Expectations 2016 Undergrad Research Paper Assessment (90 students)

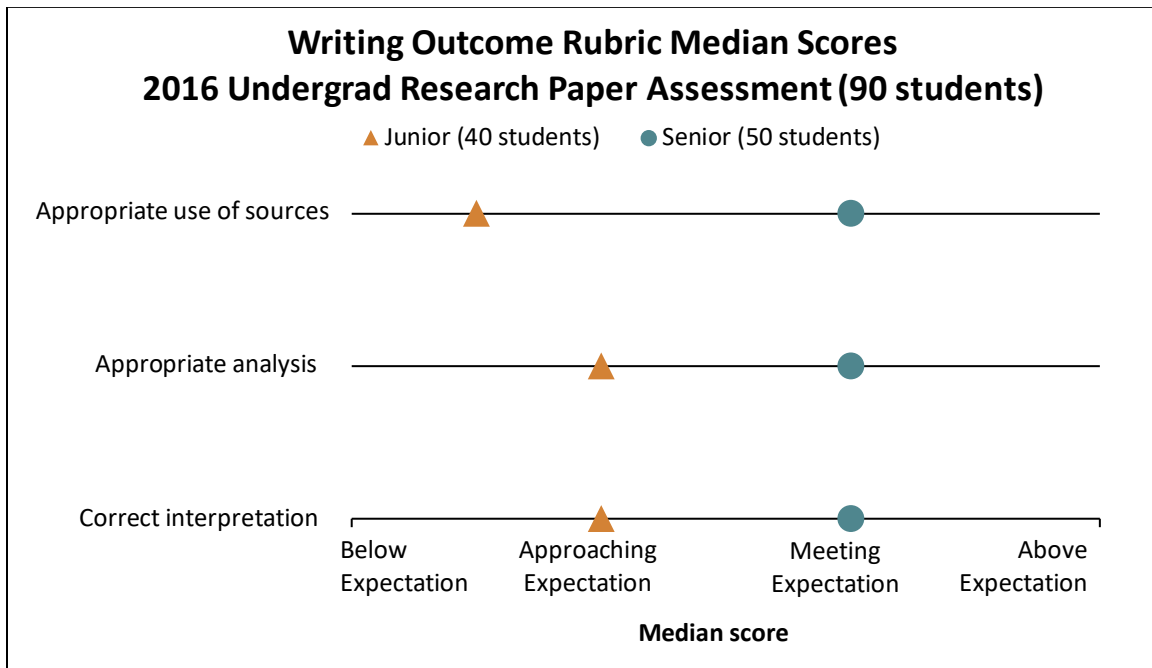
Criteria	% (#) of students "meeting or above expectation"	
	Junior (40 students)	Senior (50 students)
Appropriate use of sources	20% (8)	70% (35)
Appropriate analysis	40% (16)	80% (40)
Correct interpretation	10% (4)	90% (45)

What is the “typical” level of performance for students from different groups (demographic, learning environment, etc.) for each learning outcome?

Example: A program used a common rubric to evaluate their writing outcome using research papers. The program wants to know the “typical” level of student performance for juniors and seniors for each rubric criterion (appropriate use of sources, appropriate analysis, and correct interpretation).

- **Measures of central tendency** (i.e., mean, mode, median) are useful when you want to examine the “typical” level of performance

Note: Under certain conditions, some measures of central tendency become more appropriate to use than others. For more information on how the level of measurement can dictate which analysis methods are appropriate, see ACE’s [Quick Guide to Analyzing Quantitative \(Numeric\) Assessment Data](#).



Writing Outcome Rubric Median Scores 2016 Undergrad Research Paper Assessment (90 students)

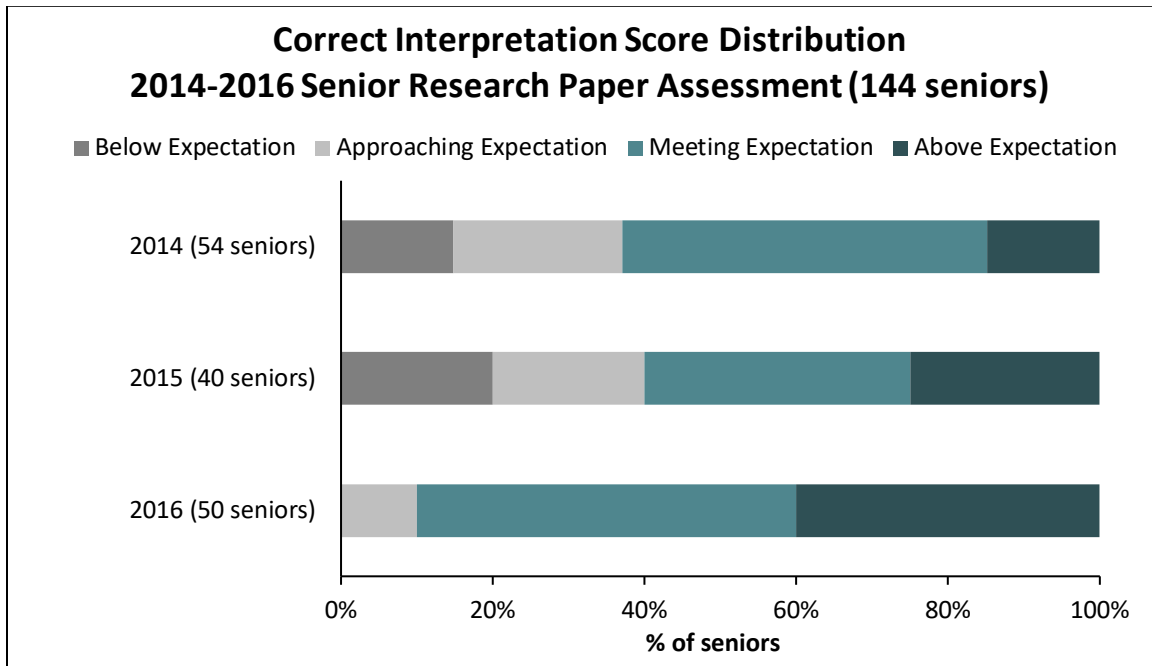
Criteria	Median Score	
	Junior (40 students)	Senior (50 students)
Appropriate use of sources	1.5	3
Appropriate analysis	2	3
Correct interpretation	2	3

Note: Ordinal scale categories were coded as follows: 1-Below Expectation, 2-Approaching Expectation, 3-Meeting Expectation, 4-Above Expectation

At what levels are students performing on the learning outcomes over time?

Example: For the past three years, a program has used a common rubric to evaluate their writing outcome using senior research papers. The program wants to know how many students are below, approaching, meeting, or above the level of expectation for the “correct interpretation” rubric criterion for each year.

- **Frequency distributions** are useful when you want to examine the number/percentage of students across multiple levels of performance



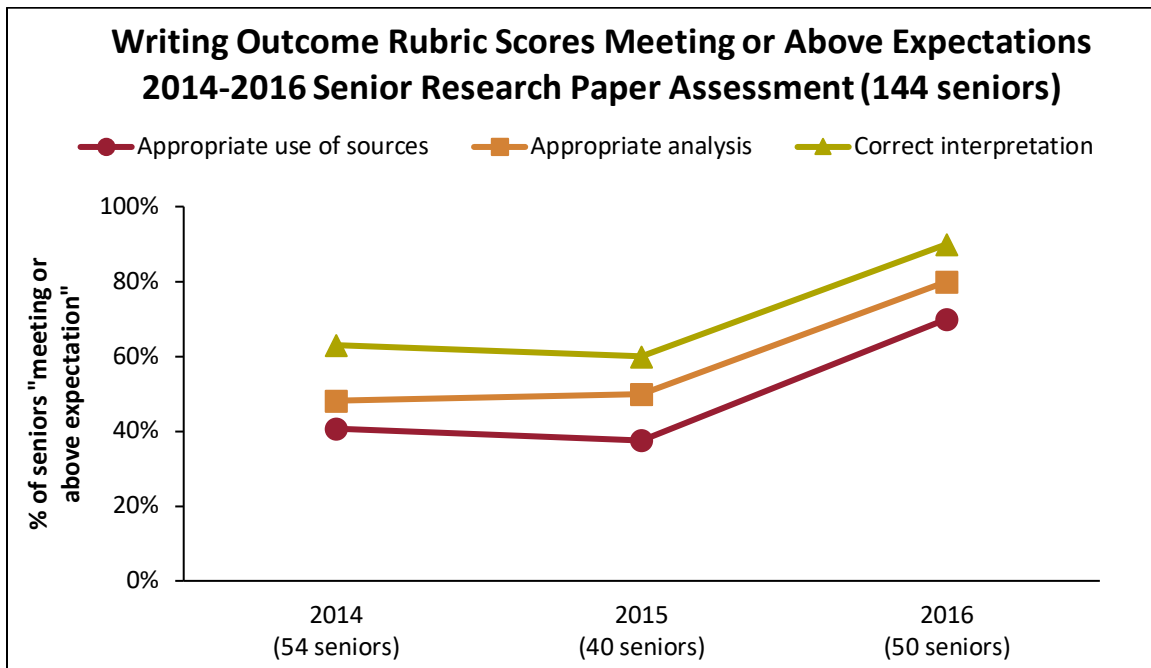
Correct Interpretation Score Distribution 2014-2016 Senior Research Paper Assessment (144 seniors)

Year	% (#) of seniors			
	Below Expectation	Approaching Expectation	Meeting Expectation	Above Expectation
2014 (54 seniors)	15% (8)	22% (12)	48% (26)	15% (8)
2015 (40 seniors)	20% (8)	20% (8)	35% (14)	25% (10)
2016 (50 seniors)	0% (0)	10% (5)	50% (25)	40% (20)

How many students are performing at an acceptable level for each learning outcome over time?

Example: For the past three years, a program has used a common rubric to evaluate their writing outcome using senior research papers. The program wants to know how many students each year are performing at an acceptable level (i.e., meeting or above expectation) for each rubric criterion (appropriate use of sources, appropriate analysis, and correct interpretation).

- **Grouped frequency distributions** are useful when you want to examine the number/percentage of students who are in adjoining levels of performance (i.e., combining adjacent frequency distribution categories)



Writing Outcome Rubric Scores Meeting or Above Expectations 2014-2016 Senior Research Paper Assessment (144 seniors)

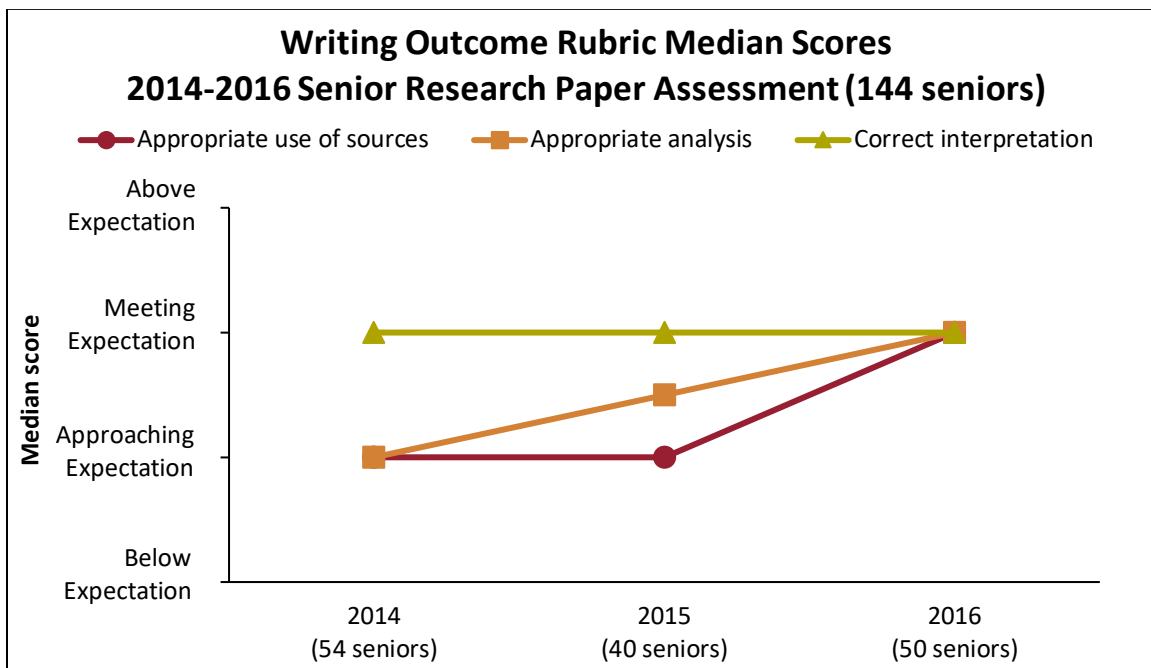
Criteria	% (#) of seniors "meeting or above expectation"		
	2014 (54 seniors)	2015 (40 seniors)	2016 (50 seniors)
Appropriate use of sources	41% (22)	38% (15)	70% (35)
Appropriate analysis	48% (26)	50% (20)	80% (40)
Correct interpretation	63% (34)	60% (24)	90% (45)

What is the “typical” level of student performance for each learning outcome over time?

Example: For the past three years, a program has used a common rubric to evaluate their writing outcome using senior research papers. The program wants to know the “typical” level of student performance each year for each rubric criterion (appropriate use of sources, appropriate analysis, and correct interpretation).

- **Measures of central tendency** (i.e., mean, mode, median) are useful when you want to examine the “typical” level of performance

Note: Under certain conditions, some measures of central tendency become more appropriate to use than others. For more information on how the level of measurement can dictate which analysis methods are appropriate, see ACE’s [Quick Guide to Analyzing Quantitative \(Numeric\) Assessment Data](#).



**Writing Outcome Rubric Median Scores
2014-2016 Senior Research Paper Assessment (144 seniors)**

Criteria	2014 (54 seniors)	2015 (40 seniors)	2016 (50 seniors)
Appropriate use of sources	2	2	3
Appropriate analysis	2	2.5	3
Correct interpretation	3	3	3

Note: Ordinal scale categories were coded as follows: 1-Below Expectation, 2-Approaching Expectation, 3-Meeting Expectation, 4-Above Expectation