2023

Focused on Pedagogy: QR Grading Rubrics for Written Arguments

Ruby Daniels

Kathryn Appenzeller Knowles

Emily Naasz

Amanda Lindner

Focused on Pedagogy: QR Grading Rubrics for Written Arguments

Abstract

Institutional assessments of quantitative literacy/reasoning (QL/QR) have been extensively tested and reported in the literature. While appropriate for measuring student learning at the programmatic or institutional level, such instruments were not designed for classroom grading. After modifying a widely accepted institutional rubric designed to assess QR in written arguments, the current mixed method study tested the reliability of two QR analytic grading rubrics for written arguments and explored students' reactions to the grading tools. Undergraduate students enrolled in a business course (N = 59) participated. A total of 415 QR artifacts from 40 students were assessed; an additional 19 students provided feedback about the grading tools. A new QR writing rubric included three main criteria (numerical evidence, conclusions, and writing), while a second rubric added a fourth criterion for assignments with data visualization. After two coders rated students' QR assignments, data analysis found both new QR rubrics had good reliability. Cohen's kappa found the study's raters had substantial agreement on all rubric criteria (κ = 0.69 to 0.80). Both the QR writing (α = 0.861) and data visualization (α = 0.859) grading rubrics also had good internal consistency. When asked to provide feedback about the new grading tools, 89% of students shared positive comments, reporting the rubrics clarified assignment expectations, improved their performance, and facilitated the writing process. This paper proposes slight modifications to the phrasing of the new rubrics' writing criterion, discusses best practices for use of rubrics in QR classrooms, and recommends future research.

Keywords

rubric, grading, assessment, quantitative literacy/reasoning, pedagogy, teaching

Creative Commons License

Cover Page Footnote

Dr. Ruby A. Daniels is an Instructional Assistant Professor in the Department of Management and Marketing at Texas A&M – San Antonio. With a background in marketing, communication, and research, she teaches various courses in the College of Business.

Dr. Kathryn Appenzeller Knowles is the Director of Quantitative Reasoning and Executive Director of the Center for Academic Innovation at Texas A&M – San Antonio.

Dr. Emily Naasz is a lecturer in political science at Texas A&M-San Antonio where she teaches classes in research methods, international relations, and comparative politics.

Dr. Amanda Lindner is an Assistant Professor in the Department of Curriculum and Instruction at Texas A&M – San Antonio. Her areas of research include determining and providing effective and appropriate literacy instruction for students with diverse backgrounds and needs.

Introduction

After decades of program development and research, there has been an "explosion of quantitative reasoning programs" (Grawe 2011a, 41). About 9 out of 10 member institutions for the Association of American Colleges and Universities (AAC&U) identify QR as an important learning outcome (Roohr et al. 2014). As QR programs continue to grow and mature, educators and researchers share best practices to further the field. As of the fall of 2022, more than a half million articles had been downloaded worldwide from *Numeracy*, the open-access, peer-reviewed journal dedicated to the study of quantitative literacy ("Recent downloads" 2022). An ongoing challenge to implementation of QR programs involves assessment, which depends on the availability of valid and reliable measures of students' quantitative reasoning. While progress has been made in the assessment of QR at the institutional level, far less work investigates assessment best practices at the classroom level. Building on work from prior QR assessment experts (AAC&U 2009a; Grawe et al. 2010), the current study investigates two new pedagogical QR rubrics designed for instructors to use when grading students' written arguments.

Review of Institutional Assessments

While several educational assessments contain questions that require quantitative skills (e.g., CLA+, SAT, ACT, GRE, GMAT, etc.), fewer assessments that explicitly measure quantitative literacy (QL) and/or quantitative reasoning (QR) have been developed and tested. These QL/QR assessments tend to use multiple choice questions, written arguments, or a combination of both. Table 1 includes some instruments used by colleges and universities to assess the quantitative reasoning skills of their students.

Table 1
QR Instruments Used by Colleges and Universities

Assessment	Primary Source	Purpose		Format	
		Institutional	Grading	Written arguments	Multiple choice
Quantitative Reasoning Test (QR-9)	Sundre (2008)	●			●
Quantitative Literacy/Reasoning Assessment (QLRA)	Gaze et al. (2014)	●			●
Quantitative Literacy VALUE rubric	McConnell and Rhodes (2017)	●		●	
Carleton College's QuIRK initiative	Grawe et al. (2010)	●		●	
Quantitative Literacy Assessment Rubric (QLAR)	Boersma et al. (2011)		●	●	

Multiple choice assessments. With the support of a National Science Foundation (NSF) grant, math and science faculty at James Madison University collaborated to determine how to assess college students' QR skills with a multiple-choice assessment. A proprietary instrument called the Quantitative Reasoning Test (QR-9) was developed and evaluated (Sundre 2008). In the test manual for the QR-9, Sundre (2008) emphasized, "test results may be used to inform curriculum and instructional improvements at the program or institution level. However, it is *not appropriate to use test results to make decisions about individual students*" [emphasis added] (3). Reliability analysis of the QR-9 found Cronbach's alpha scores of 0.64 and 0.66. Ranging between 0 to 1, alpha scores above 0.70 are typically considered acceptable (Tavakol and Dennick 2011; Taber 2018; Miller and Lovler 2020).

In 2014 Gaze et al., who also received an NSF grant, created the Quantitative Literacy/Reasoning Assessment (QLRA). The goal of the QLRA project was to "create a non-proprietary, reliable test that would establish a national baseline of QLR abilities across the nation's higher education spectrum" (Gaze et al. 2014, 4). Like the QR-9, the primary purpose of the QLRA was institutional, to evaluate courses and QL/QR curriculum initiatives. The QLRA includes 20 multiple-choice items related to number sense, visual representation, probability/statistics, and reasoning (Gaze et al. 2014). Item phrasing was content validated by a panel of quantitative literacy/reasoning subject matter experts. Each member of the content validation team had more than 10 years' experience teaching and assessing QL/QR in higher education. QLRA data were collected and analyzed from over 3,000 students at 10 institutions. A Cronbach's alpha score for the QLRA indicated good reliability ($\alpha = 0.862$).

Both the QLRA and QR-9 allow for comparisons across different institutions. The multiple-choice format of the QLRA and QR-9 assessments makes them easy to administer and score, especially in courses with multiple sections and/or high enrollment. Like any assessment with a set of correct responses, especially ones that are administered in an unproctored online environment, the validity of instruments may be compromised if students are able to locate a copy of the answers (Lievens and Burke 2011; Cavanagh 2014). In situations where students are graded on a completion basis, the risk of cheating is lower. However, scores on such assessments may not validly measure understanding of QR, especially if students rush through the questions and/or are not interested in the assignment.

Written assessments. First introduced in 2009, the Association of American Colleges and Universities developed the Valid Assessment of Learning in Undergraduate Education (VALUE) rubrics (AAC&U 2022) as an alternative to standardized tests of student learning. Faculty subject matter experts from various institutions developed and validated VALUE rubrics that measure 16 broad student

learning outcomes, such as teamwork, problem solving, ethical reasoning, civic engagement, and written communication. "Utilized by more than 5,600 discrete organizations across 142 countries, the VALUE rubrics have made an essential contribution to the dialogue on the assessment of college learning" (AAC&U 2022, para. 2). The Quantitative Literacy VALUE rubric (AAC&U 2009a) uses a four-category scale to measure six dimensions of QL: interpretation, representation, calculation, application/analysis, assumptions, and communication.

Researchers suggest the VALUE rubrics have both face and content validity because they were developed after extensive deliberations by faculty members across the United States (Rhodes and Finley 2013). While sharing that several universities claim to have high inter-rater reliability when using the VALUE rubrics, Rhodes and Finley (2013) did not report a specific reliability for the QL VALUE rubric. In 2017, McConnell and Rhodes reported inter-rater reliability tests of the QL VALUE rubric ranged from 0.52 to 0.62. When the QL VALUE rubric was tested with STEM students, Gray et al. (2017) found intraclass correlation coefficients (ICCs) ranging from 0.47 to 0.60 demonstrated the rubric's reliability.

While these results provide some insight into the reliability of the QL VALUE rubric, additional questions remain. Because McConnell and Rhodes (2017) did not report Cohen's kappa statistics, which adjust an inter-rater reliability percentage to account for matches occurring due to random chance, their findings may be inflated. Similarly, the ICCs reported by Gray et al. (2017) are considered poor to moderate by other researchers who interpret ICCs below 0.50 as poor and ones between 0.50 to 0.75 as having moderate reliability (Koo and Li 2016; Bobak et al. 2018).

Researchers at Carleton College developed the Quantitative Inquiry, Reasoning, and Knowledge (QuIRK) initiative with support from an NSF grant (Grawe et al. 2010). Rather than measuring students' QR skills with multiple-choice tests, Grawe (2011a) suggests a direct assessment of written arguments tends to be a more valid measure of whether learners can apply their quantitative skills. "Students' ability to construct arguments with quantitative evidence and the habit of mind to seek out that evidence in the first place might be better assessed in essays or oral talks" (50). Over four years, QuIRK researchers developed rubrics to assess students' ability to use quantitative reasoning in written arguments. Their rubrics assessed sophomore students' writing portfolios when QR was either centrally or peripherally relevant.

Wide acceptance in the QR literature (e.g., Grawe 2011a, 2011b; Craver 2014; Elrod 2014; Hubert and Lewis 2014; Jastram et al. 2014; Colombini and Hum 2017) provides strong evidence of the QuIRK rubrics' content validity. When evaluating the reliability of the rubric for papers when QR is centrally relevant, Grawe et al. (2010) found multiple raters had exact agreement for 66.7% of papers when coding with a 4-category rubric. Unlike studies of the QL VALUE rubric that only reported

general inter-rater reliabilities (Gray et al. 2017; McConnell and Rhodes 2017), research about the QuIRK rubric shared Cohen's kappa (κ) statistics to indicate "the degree to which the observed agreement exceeds the expected agreement, relative to the agreement not explained by chance" (10). Grawe et al.'s (2010) analysis found the QuIRK rubric for papers with QR that is centrally relevant (see Table 2) had a Cohen's kappa score of 0.532, which indicates moderate inter-rater reliability (Sun 2011).

Table 2
Institutional QuIRK Rubric When QR is Centrally Relevant

1	2	3	4
Use of numerical evidence is so poor that either it is impossible to evaluate the argument with the information presented or the argument is clearly fallacious. Perhaps key aspects of data collection methods are missing, or critical aspects of data source credibility are left unexplored. The argument may exhibit glaring misinterpretation (for instance, deep confusion of correlation and causation). Numbers may be presented but are not woven into the argument.	**The use of numerical evidence is sufficient to allow the reader to follow the argument. But there may be times when information is missing or misused.** Perhaps the use of numerical evidence itself is uneven. Or the data are presented effectively, but a lack of discussion of source credibility or methods makes a full evaluation of the argument impossible. Misinterpretations such as the confusion of correlation and causation may appear, but not in a way that fundamentally undermines the entire argument.	**The use of numerical evidence is good throughout the argument.** Only occasionally (and never in a manner that substantially undermines the credibility of the argument) does the paper fail to explore source credibility or explain methods when needed. While there may be small, nuanced errors in the interpretation, the use of numerical evidence is generally sound. **However, the paper may not explore all possible aspects of that evidence.**	**The use of numerical evidence is consistently of the highest quality.** When appropriate, source credibility is fully explored, and methods are completely explained. Interpretation of the numerical evidence is complete, considering all available information. There are no errors such as confusion of correlation and causation. **This paper would be an excellent choice as an example of effective central QR to be shared with students and faculty.**

Note: (Grawe et al. 2010) CC BY-NC

Rubrics designed to assess the use of QL/QR in written arguments (AAC&U 2009b; Grawe et al. 2010) are more likely to measure how effectively students apply QR skills in real-world settings. Unlike standardized tests that often approach quantitative concepts abstractly or without context, QR requires the ability to address topics that may be sloppy and unstructured (Wiggins 2003; Grawe 2011a). The QL VALUE and QuIRK rubrics also are well suited for assessing students' use of QR in an interdisciplinary environment. The purpose of QR Across the Curriculum initiatives (Numeracy Infusion Course for Higher Education n.d.; Rosen et al. 2003; Steele and Kiliç-Bahi 2008; Elrod 2014; Science Education Resource Center 2018; Ober et al. 2019) is to encourage the integration of quantitative literacy/reasoning in various courses (e.g., psychology, marketing, biology, communication, history, management, journalism, public speaking, etc.). Unlike standard multiple-choice assessments that may be misaligned with a specific discipline, argument-focused rubrics provide both guidance and flexibility for faculty who may not have explicit training in QR.

Limitations. While measuring QR in written arguments with existing rubrics offers benefits over multiple choice assessments, such instruments were not designed for teachers to grade individual student assignments. Rhodes and Finley (2013) emphasize the VALUE rubrics "were developed as 'meta-rubrics' to be used at the *institutional or programmatic levels* in order to assess student learning overall and over time, not for specific assignments" [emphasis added] (6). Similarly, the QuIRK instrument was "designed to be applied to samples of student writing to assess QR at an institutional level. In particular, the *rubric is not designed to evaluate individual students*" [emphasis added] (Grawe et al. 2010, 2). Instead, the purpose of the QuIRK rubric is to "discern effects of institution-level programs and curricular reforms" (Grawe et al. 2010, 3).

To overcome this limitation, Boersma et al. (2011) adapt the phrasing of the institutional QL VALUE rubric to create the Quantitative Literacy Assessment Rubric (QLAR). The QLAR was originally designed for grading students' written responses to questions in a casebook for a Quantitative Reasoning in the Contemporary World (QRCW) course. While the QLAR retained the six criteria from the QL VALUE rubric, Boersma et al. (2011) modified the descriptions of the QL VALUE rubric's performance levels because they had problems producing reliable results with the rubric's original phrasing. After mapping a series of casebook questions to the six QL competencies in the QL VALUE rubric, two readers compared their QLAR evaluations of answers to various QRCW casebook questions. In two separate tests, the coders had strong inter-rater agreement of 88% and 97%. Boersma et al.'s (2011) study highlighted the importance of testing (and, if needed, modifying) institutional QR rubrics to develop tools that function well in a classroom environment.

Holistic and Analytic Rubrics

Table 3 compares the purpose of the two main types of rubrics: holistic and analytic. Holistic rubrics are often used by educational administrators to assess institutional or programmatic performance (Brookhart 2018). After evaluating various student artifacts, aggregated scores are tallied and shared to give educational leaders a general measure of progress so they can make curricular changes, as needed. The QuIRK assessment is an example of a holistic rubric. Its findings are intended to be used outside the purview of individual students and instructors (Grawe et al. 2010). In contrast, analytic rubrics are often used to grade individual students at the classroom level (Brookhart 2018). These pedagogically focused tools promote student learning by allowing instructors to (1) evaluate each student's performance and (2) provide specific feedback about individual strengths and opportunities for improvement.

Table 3
Comparison of the General Characteristics of Holistic and Analytic Rubrics

General Characteristics	Rubrics	
	Holistic	Analytic
Audience	Administrators	Students
Purpose	Institutional assessment	Classroom grading
Assessment	General	Specific
Criterion	Single	Multiple
Scores	Overall evaluation	Sum of criteria scores

Configured in a grid with rows and columns, rubrics have two main components: criteria and performance level descriptions. Criteria identify what elements are required for an assignment (Brookhart 2018). Performance level descriptions indicate what criteria look like at various levels of quality. Rubrics may assign points, percentages, or letter grades to each performance level (Jensen 1995).

While holistic and analytic rubrics both include performance levels, they structure criteria differently. Holistic rubrics evaluate all criteria at the same time (Brookhart 2018). For instance, the QuIRK rubric has one row that holistically evaluates use of QR. An early version of the QuIRK rubric used an analytic approach that examined three elements of QR quality: implementation, interpretation, and communication. However, because Grawe et al. (2010) found "readers had a difficult time distinguishing between these intertwined concepts" (20), the structure was changed to a holistic rubric. In contrast, analytic rubrics assess criteria separately (Brookhart 2018). While the QL VALUE rubric (AAC&U 2009a) was designed for institutional assessment, its rubric structure is analytic because it contains six distinct criteria. After conducting a systematic literature review of studies that examined the use of rubrics in higher education, Brookhart (2018) found 57% of rubrics were analytic.

In addition to structuring criteria differently, rubrics may use distinct levels of measurement for performance. Holistic assessments score QR artifacts categorically. For instance, a holistic tool might designate a letter grade (A to F) to each column of the rubric. The QuIRK assessment uses such an approach, with categorical measurement that ranges from 1 to 4 (Grawe et al. 2010). In contrast, analytic rubrics score each criterion independently, then sum the row totals to compute an interval-level score. For instance, an analytic rubric with 5 performance levels (1 = lowest rating, 5 = highest rating) and 5 criteria would have scores that range from 5 to 25. Similarly, specific percentages can be assigned to each performance level. Instructors may also allocate different weights to specific criterion in an analytic rubric. For example, Criteria A and B might be worth 40%

each, with Criteria C weighted at 20%. Such weighting is not possible in holistic rubrics.

Holistic assessment tools tend to be quicker, less cognitively demanding, and appropriate when student feedback is not needed (Brookhart 2018). While they work well for institutional assessment, holistic instruments, such as the QuIRK rubric, were not designed to help students learn QR. In contrast, pedagogically focused grading rubrics identify what is expected of student work and describe what learning looks like (Andrade 2000; Arter and McTighe 2001; Arter and Chappuis 2006; Bell et al. 2013; Brookhart 2013, 2018; Nordrum et al. 2013; Panadero and Jonsson 2013). Analytic rubrics also support students by aligning learning with grading (Brookhart 2018). In a classroom setting, an analytic grading rubric is more useful than a holistic one because analytic rubrics provide students with more detailed feedback (Arter and McTighe 2001; Arter and Chappuis 2006; Brookhart 2013, 2018; Brookhart and Nitko 2019). Other benefits of analytic grading rubrics include increased transparency, reduced anxiety, and improved student self-efficacy (Panadero and Jonsson 2013).

Gap in the Literature

Despite the tremendous growth of QR programs, few studies have tested the reliability of grading rubrics to assess QR in written assignments at the classroom level. Studies of QR instruments (see Table 1) primarily focus on programmatic assessment at the institutional level. While recognizing the importance of institutional and program assessments, Sundre and Thelk (2010) indicate QR measurement tools are also "sorely needed" (11) by collegiate instructors. Bressoud (2009) similarly describes various classroom "trials and tribulations" (2) when teaching a collaborative QR session for students enrolled in various introductory courses.

Given the complex nature of QR, Grawe (2011a) suggests "Multiple instruments are almost surely necessary" (50). Research is needed to determine whether institutional rubrics can be reliably modified for classroom use. In their analysis of the VALUE rubrics, Rhodes and Finley (2013) emphasize such revisions need to be made carefully "to reflect the course content and assignments being examined, while still preserving the dimensions of learning in the original rubric" (7). Grawe et al. (2010) similarly acknowledge that future researchers and educators might revise the QuIRK holistic institutional rubric to align with different objectives or student populations.

Despite widespread use of rubrics in college classrooms (Brookhart 2018), only one study has tested the reliability of a pedagogically focused QR rubric. While Boersma et al.'s (2011) QLAR research demonstrated it is possible to successfully adapt an institutional rubric to grade students' written arguments, their study was conducted more than a decade ago and was designed to assess responses

to casebook questions in a specific QR course. To date, the QLAR's reliability findings have not been replicated and no studies have tested use of the QLAR in non-QRCW courses. Most importantly, students in the study "were not familiar with the QLAR rubric" (Boersma et al. 2011, 13). As a result, students were unaware of how their work scored on the rubric. Not sharing the QLAR also forfeited the opportunity to collect student insight about the usefulness of the grading rubric.

Purpose and Research Questions

Building upon the QuIRK institutional assessment, a mixed method study quantitatively tested two new QR grading rubrics, then qualitatively explored students' responses to the grading tools. The research project was part of a larger initiative to assess the effectiveness of a curricular change in an undergraduate business course (Daniels et al. 2022). The purpose of the study was to investigate methodological and pedagogical implications for using new grading rubrics when assessing written QR assignments. Modifications to the holistic QuIRK institutional assessment resulted in the development of two analytic grading rubrics. A new QR writing rubric with three criteria (numerical evidence, conclusions, and writing) was designed to grade written assignments when QR is centrally relevant. A second rubric, which included the same criteria as the QR writing rubric, added a fourth criterion to assess written assignments with data visualization.

When measuring student performance, assessment tools should be both valid and reliable. Validity indicates whether a data collection instrument measures what it intended to measure, while reliability reflects the assessment's ability to measure consistently (Miller and Lovler 2020). The new QR grading rubrics had face validity because much of the phrasing was identical to the QuIRK institutional rubric. However, the reliability of the new grading rubrics was unknown. To provide methodological insight, the study's quantitative research question investigated: *How reliably do the new grading rubrics measure QR?*

Inter-rater agreement between two coders is commonly computed when evaluating open-ended data on a categorical scale (Miller and Lovler 2020). While it is one form of reliability, the percentage of inter-rater agreement does not account for matches that may occur due to random chance (Grawe et al. 2010; Sun 2011). In contrast, Cohen's kappa adjusts for identical evaluations that may occur randomly. Kappa statistics range from 0 to 1, with higher scores indicating greater coder agreement (Sun 2011; McHugh 2012). Researchers report kappa benchmarks indicate when raters have poor (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), to almost perfect (0.81–1.00) agreement (Sun 2011; McHugh 2012). Based on these benchmarks, the study hypothesized the Cohen's

kappa scores for both the QR writing and data visualization grading rubrics would indicate at least moderate coder agreement, with kappa scores equal to or greater than 0.41.

Cronbach's alpha is another widely used measure of reliability that evaluates the internal consistency of a set of items (Miller and Lovler 2020). Ranging from 0 to 1, alpha scores above 0.70 are typically considered acceptable, with 0.80 considered good, and 0.90 or more indicating high reliability (Tavakol and Dennick 2011; Taber 2018; Miller and Lovler 2020). In alignment with these benchmarks, the study hypothesized Cronbach's alpha scores for both the new QR grading rubrics would be equal to or greater than 0.70.

The current study was the first time the new QR grading rubrics were used. As a result, the study's researchers were also interested in students' reactions to the new grading tools. To provide qualitative insight, the study's second research question explored: *How did students respond to the new QR grading rubrics?*

Methodology
Sample

Participants were senior undergraduate students ($N = 59$) enrolled in a 16-week business course at a Hispanic-Serving Institution (HSI) in Texas. Nearly 8 out of 10 of the university's students are Hispanic, with most representing first-generation college students. Because a significant number of the university's students have exceptional financial need, about 70% are eligible for Pell grants. The course's hybrid format included one 75-minute on-campus lecture each week, with the remaining activities (readings, videos, quizzes, and exams) completed asynchronously online. The course required each student to complete 11 QR assessments about 4 teaching interventions. In Fall 2021 and Spring 2022, a total of 415 student artifacts were collected and analyzed from 40 students, 75 of the assignments included data visualization. Intervention topics focused on the interpretation of bell curves, income statements, price elasticity statistics, and chart data. In Fall 2022, open-ended data were collected from 19 students about their reactions to the rubrics.

Data Collection

In accordance with an approved IRB protocol, all students enrolled in the course during the Fall 2021 and Spring 2022 semesters were invited to participate in the study. Before any data collection, participants signed a standard consent form for their data to be included. Students' QR assessments were collected online in the university's learning management system. Prior to any data analysis, each assessment was deidentified to protect student confidentiality.

Instrumentation

As illustrated in Table 4, data were collected at the start of the course (pre-test), immediately after the teaching interventions (case studies), and at the end of the semester (post-test). After data collection, each artifact was assessed with (1) the holistic QuIRK and (2) one of the new analytic grading rubrics.

Table 4
Overview of the Study's Design and QR Interventions

| Phase | Week | QR Intervention Topics | | | |
		Salary	Income statements	Price elasticity	Data visualization
Pre-test	1	●	●	●	●
Case study	2	●			
Case study	4		●		
Case study	6			●	
Case study	10				●
Post-test	16		●	●	●

Note: The dots represent the 11 QR assessments students were required to complete during the course.

Format. In the first week of class, students completed a pre-course assessment that required them to write responses to four open-ended questions, which aligned with the topic of each teaching intervention (see Appendix A). After participating in an on-campus lecture about each topic, students were assigned a case study that required the use of quantitative reasoning. At the end of the course, students took an online final exam that included open-ended questions about three of the teaching interventions (see Table 4).

Table 5 compares general characteristics of the pre-test, case study, and post-test instrumentation. All the QR assessments were submitted online and used open-ended prompts that required students to provide written responses. The pre-test assessments represented low-stake assignments (graded for completion only), while the case study and post-test assessments were high-stake assignments (evaluated on a 100% performance scale). Unlike the case study instructions, which required a response approximately three to four pages long, the pre-test and post-test assessments did not specify a length for the assignment. Students had one week to prepare and submit their work for the pre-test as well as each case study. A maximum of two hours was provided to complete the post-test final exam, which also included 50 multiple choice questions about other course concepts.

Table 5
Comparison of Pre-Test, Case Study, and Post-Test Characteristics

Assessment Characteristics	Pre-test	Case studies	Post-test
Submission	Online	Online	Online
Question format	Open-ended	Open-ended	Open-ended
Grading scale	Completion only	Up to 100%	Up to 100%
Length	Not specified	3 to 4 pages	Not specified
Preparation time	1 week	1 week	2 hours

Rubrics. The new QR writing rubric uses a 5-category scale to assess three criteria: numerical evidence, conclusions, and writing (see Table 6). The numerical evidence criterion includes the same phrasing as the QuIRK rubric. It also has a performance level for written assignments that contain no numerical evidence. While grounded in the original language of the QuIRK rubric, the phrasing of the conclusion criterion's performance levels was slightly modified. For instance, the beginning of each performance level was changed to clearly indicate the rater should evaluate the student's conclusions (e.g., "presented a sufficient conclusion," "presented a good conclusion," "presented a logical and well-reasoned conclusion"). Because the descriptor "poor" in the QuIRK rubric's first performance level might be perceived by some students as discouraging, the word was changed to "significantly underdeveloped." To allow for more nuanced evaluations, the second performance level's phrasing was changed from "makes a full evaluation of the argument *impossible*" [emphasis added] to "made a full evaluation of the argument *difficult*." Like the numerical evidence criterion, a new performance level was added for artifacts that did not draw any conclusions. Because QR requires students to communicate quantitative information in clear, cogent arguments (Rutz and Grawe 2010; Wolfe 2010), a writing criterion was also included to evaluate the clarity, fluency, and mechanics of a student's writing.

A second grading rubric, which added a fourth data visualization criterion to the QR writing rubric, was developed for written assignments that require students to create one or more charts, graphs, infographics, etc. The phrasing of the visualization criteria was adapted from the representation criterion of the AAC&U's (2009a) QL VALUE rubric (see Table 7).

Table 6
Current Study's QR Writing Rubric

Criteria	1	2	3	4	5
Numerical Evidence Integrated appropriate numerical evidence in the narrative explanations	No use of numerical evidence	The use of numerical evidence is so poor that either it is impossible to evaluate the argument with the information presented or the argument is clearly fallacious.	The use of numerical evidence is sufficient to allow the reader to follow the argument. But there may be times when information is missing or misused.	The use of numerical evidence is good throughout the argument.	The use of numerical evidence is consistently of the highest quality.
Conclusion Presented an appropriate, well-reasoned conclusion based on numerical evidence	Did not draw any conclusions	Presented a significantly underdeveloped conclusion with minimum supporting arguments; numbers are presented but not woven into argument	Presented a sufficient conclusion but lacked discussion of the numerical evidence that made a full evaluation of the argument difficult	Presented a good conclusion that aligned with the numerical evidence; may not have explored all possible aspects of the numerical evidence	Presented a logical and well-reasoned conclusion; interpretation of the numerical evidence is complete and considers all available information
Writing Demonstrated college-level writing (clarity, fluency, and writing mechanics)	Major writing errors distract significantly from ideas in the paper	The writing is rushed and/or several ideas are confusing.	The writing rambles and/or is wordy.	The writing expressed most ideas clearly.	Consistently expressed ideas clearly and fluently throughout the paper.

Note: Adapted from Grawe et al. (2010) CC BY-NC

Table 7
Current Study's Data Visualization QR Grading Rubric

Criteria	1	2	3	4	5
Numerical Evidence Integrated appropriate numerical evidence in the narrative explanations	No use of numerical evidence	The use of numerical evidence is so poor that either it is impossible to evaluate the argument with the information presented or the argument is clearly fallacious.	The use of numerical evidence is sufficient to allow the reader to follow the argument. But there may be times when information is missing or misused.	The use of numerical evidence is good throughout the argument.	The use of numerical evidence is consistently of the highest quality.
Conclusion Presented an appropriate, well-reasoned conclusion based on numerical evidence	Did not draw any conclusions	Presented a significantly underdeveloped conclusion with minimum supporting arguments; numbers are presented but not woven into argument	Presented a sufficient conclusion but lacked discussion of the numerical evidence that made a full evaluation of the argument difficult	Presented a good conclusion that aligned with the numerical evidence; may not have explored all possible aspects of the numerical evidence	Presented a logical and well-reasoned conclusion; interpretation of the numerical evidence is complete and considers all available information
Writing Demonstrated college-level writing (clarity, fluency, and writing mechanics)	Major writing errors distract significantly from ideas in the paper	The writing is rushed and/or several ideas are confusing.	The writing rambles and/or is wordy.	The writing expressed most ideas clearly.	Consistently expressed ideas clearly and fluently throughout the paper
Visual Representation Converted numerical information into a visual format (e.g., chart, graph, infographic, etc.)	Did not visually represent any numerical information; only included words or a table	Converted numerical information but the resulting visual representation is inaccurate	Converted numerical information accurately but the resulting visual representation adds little to no understanding	Clearly and accurately converted numerical information into a visual representation that furthers some understanding	Skillfully converted numerical information into appropriately labeled visual representation that furthers deeper understanding

Note: Adapted from Grawe et al. (2010) CC BY-NC and AAC&U (2009a) CC BY-NC-SA 4.0

Coder training. Two doctoral-trained raters attended a synchronous online session to learn how to code the study's data. Both coders previously participated in a yearlong quantitative reasoning faculty fellowship program and had experience teaching QR. The raters reviewed the QuIRK, QR writing, and data visualization rubrics and discussed samples of student artifacts. As shown in Appendix B, samples of each performance level were annotated to provide raters with additional guidance. As shown in Table 6, the conclusion criterion of the rubric required students to "present an appropriate, well-reasoned conclusion based on numerical evidence." This required students to (1) draw at least one conclusion and (2) provide well-reasoned arguments with numerical evidence to support their conclusions. To clarify each performance level, the training sheet for the conclusion criterion highlighted examples of conclusions in blue, while arguments were underlined (see the Important Information boxes at the bottom of each page in Appendix B).

In the second phase of the training, coders practiced using the QuIRK, QR writing, and data visualization rubrics to independently evaluate 10 student artifacts. The coders assessed each artifact twice. Papers that did not include data visualization were assessed with the QuIRK and new QR writing rubric, while artifacts with charts, graphs, or infographics were assessed with the QuIRK and new data visualization rubric. After evaluating the practice artifacts, the raters compared and discussed their scores to clarify their understanding of the rubrics. Once both raters indicated they understood the assessment process, they independently evaluated approximately 25% of the student artifacts ($n = 110$). In situations when different evaluations emerged, a revised assessment protocol was implemented that asked raters to negotiate their differences (Grawe et al. 2010). Then, they collaboratively decided whether to retain their initial coding or adjust for agreement.

Open-ended student responses. After completing a case study that was evaluated by the new grading tool, the instructor asked students to share open-ended written thoughts about the new QR writing rubric during one of the on-campus lectures. They also provided written feedback about a video the instructor created that contained tips for using the QR writing rubric. The video was about 10 minutes long and showed the instructor talking about the QR rubric. During the video, the teacher reminded students where to find the rubric and reviewed each criterion of the grading tool. The video also showed screenshots of the assignment description to help students understand the connection between the case study and rubric requirements. Throughout the video, the instructor encouraged students to recall the in-class practice exercises that used the same QR process required for the case study. Finally, because a writing criterion was included in the rubric, the teacher

recommended students use the writing resources provided in the course's online materials when preparing their assignments.

Two coders evaluated the students' qualitative thoughts about the new QR writing rubric and video with tips. After reading each response, the raters globally categorized each student's comments as positive (favorable, helpful, appreciative), negative (critical, indifferent, apathetic, disapproving), or neutral (impartial). They also content analyzed comments to identify themes in the open-ended data, as needed, to provide additional insight.

Results
Preliminary Analysis

During the study, 40 students in Fall 2021 and Spring 2022 were asked to complete 11 QR assessments. A total of 415 QR artifacts were collected and 75 of these assignments included a data visualization requirement. Gender was equally distributed (50% male, 50% female), with the same number of students participating in both semesters. In Fall 2022, 19 additional students, who participated in the same interventions, shared written thoughts about the new QR grading rubrics.

RQ1: Reliability of the QR Grading Rubrics

Inter-rater reliability. In a preliminary analysis, two raters coded about 25% of the study's QR artifacts. Inter-rater agreement on the numerical evidence, conclusion, and writing criteria of the QR writing rubric was based on ratings of 105 student artifacts, while the inter-rater reliability of the data visualization criterion was assessed on 19 assignments. The numerical evidence criterion had the highest overall percentage of exact inter-rater reliability at 84.76%, followed by the conclusion (83.81%), writing (81.90%), and data visualization (78.95%) criteria. Cohen's kappa analysis indicated the study's raters had substantial agreement on all four rubric criteria, with kappa scores ranging from 0.69 to 0.80 (see Table 8).[1] Because there was substantial agreement between the two coders' evaluations when using the new grading rubrics, one rater assessed the remainder of the study's QR artifacts.

Internal consistency. Cronbach's alpha measured the internal consistency of the QR rubrics' criterion as a second indicator of reliability. The QR writing rubric had a Cronbach's alpha of 0.861, while the data visualization rubric had an alpha of

---

[1] A similar process found the overall percentage of exact inter-rater agreement when using the QuIRK rubric was 83.6%. Cohen's kappa analysis similarly indicated the two raters had substantial agreement when coding with the QuIRK rubric ($\kappa = 0.78$).

Table 8
Reliability of the QR Writing and Data Visualization Rubric Criterion

QR Rubric Criterion	N	Percentage of inter-rater agreement	κ
Numerical evidence	105	84.76%	0.80
Conclusions	105	83.81%	0.78
Writing	105	81.90%	0.75
Data visualization	19	78.95%	0.69

0.859. These results indicated good reliability (Tavakol and Dennick 2011; Taber 2018; Miller and Lovler 2020).[2]

Consistent with the Cronbach's alpha findings, post hoc analysis found highly significant correlations between the study's four rubric criteria. The numerical evidence criterion of the new analytic QR rubrics was highly correlated with the conclusion ($r_s(413) = 0.82$, $p = 0.00$), writing ($r_s(413) = 0.60$, $p = 0.00$), and data visualization ($r_s(73) = 0.53$, $p = 0.00$) criterion. The conclusion criterion was also highly correlated with writing ($r_s(413) = 0.59$, $p = 0.00$) and data visualization ($r_s(73) = 0.54$, $p = 0.00$). Writing and data visualization were also significantly correlated ($r_s(73) = 0.53$, $p = 0.00$).[3]

RQ2: Student Responses to the QR Grading Rubrics

Students ($n = 19$) who attended one of the on-campus lectures in Fall 2022 provided written responses about the QR writing rubric as well as a video with tips about the grading tool (see Daniels 2022). After reviewing the open-ended data, two coders categorized each student's remarks as positive, negative, or neutral. The raters also content analyzed the information to identify themes within each category. Analysis

---

[2] QuIRK scores were significantly correlated with summed scores for the QR writing ($r_s(413) = 0.87$, $p = 0.00$) and data visualization ($r_s(73) = 0.84$, $p = 0.00$) rubrics. Ratings on the QuIRK rubric were also highly correlated with item scores for numerical evidence ($r_s(413) = 0.84$, $p = 0.00$), conclusions ($r_s(413) = 0.80$, $p = 0.00$), writing ($r_s(413) = 0.64$, $p = 0.00$), and data visualization ($r_s(73) = 0.58$, $p = 0.00$). Highly significant correlations were expected because portions of the new grading rubrics were identical or very similar to the phrasing of the QuIRK rubric.

[3] A post hoc factor analysis of the new QR rubric criteria detected only one factor. Factor loadings for the new QR writing rubric were 0.826 or higher, while the visualization rubric had factor loadings of 0.724 or more. With an eigenvalue of 2.369, the QR writing rubric explained about 79% of the variance in the factor. The data visualization rubric had an eigenvalue of 2.878 and explained about 72%. All the communalities exceeded 0.40, with the lowest communalities associated with the writing (0.525) and visualization (0.638) criteria.

of the coders' evaluations of students' open-ended comments about the QR rubric and video found the percentage of inter-rater reliability was 97.54%. Cohen's kappa similarly indicated the two coders had almost perfect agreement ($\kappa = 0.94$).

New QR writing rubric. Most students (89%, $n = 17$) shared positive comments about the new grading rubric. Three themes were identified in the positive responses (see Fig. 1). Several students (58%, $n = 11$) described how the rubric clarified expectations about the assignment. One student shared, "The QR rubric kept me on track and reminded me what needs to be in my case study." Another observed, "The rubric acted like a checklist for me. I could tell what was missing." Students (21%, $n = 4$) also described how the QR rubric helped them improve their performance on the assignment. A student explained, the QR "rubric provided a breakdown of how to achieve a good score." Another indicated, "I like how it shows the percentage each slot is worth. So, if you do a mediocre job or the bare minimum, your score reflects your effort." A few students (11%, $n = 2$) also described how the rubric facilitated their writing process. One student noted, "When resources are provided that actually make sense and break down what is asked of the student, it's easier to start plugging away at the assignment." Another said the grading rubric "was helpful when I was in the middle of writing and just needed to confirm something."
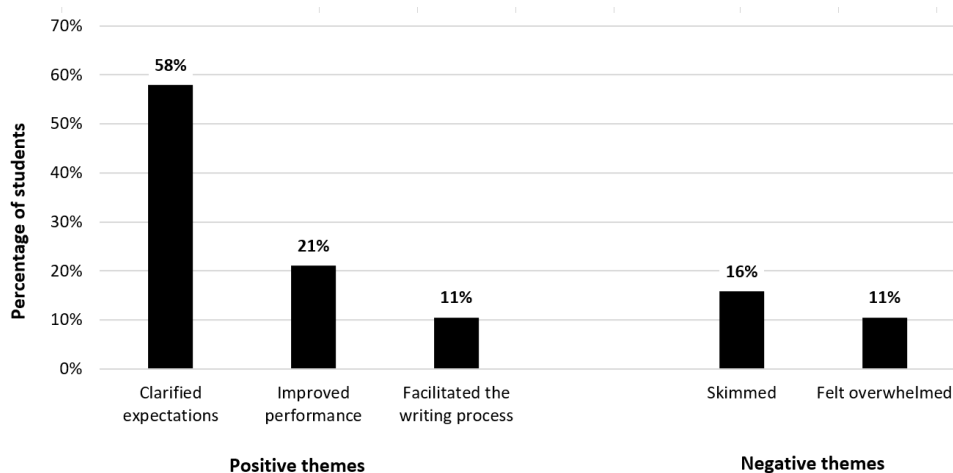

Figure 1. Themes in student comments about the QR writing rubric.

While their overall evaluations were positive, some students (16%, $n = 3$) said they did not thoroughly review the QR rubric. One student wrote, "I skimmed over it and did not pay attention." Another shared, "I didn't read the whole rubric."

No students shared neutral comments about the QR rubric, but a few (11%, $n = 2$) shared negative descriptions. One student explained it was "overwhelming." Another expressed similar concerns. "If I am honest, the rubric has a lot of wording.

I am a visual learner, so when I see all those words, my brain gets lost. [It's] like a sea of words."

QR rubric video. No negative comments were shared in students' open-ended descriptions about the video with tips about the grading rubric. One student (5%) did not express an attitude about the video, instead indicating she preferred working with the rubric. Nearly all the students (95%, *n* = 18) wrote positive comments about the QR rubric video.

Five themes were detected in the content analysis of comments about the video (see Fig. 2). About a third the students (37%, *n* = 7) described how the video clarified what was expected for the assignment. An equal number of students (37%, *n* = 7) said the video simplified the rubric information. These students described how the video broke down the requirements, so the assignment was more understandable. Approximately a quarter of the students (26%, *n* = 5) indicated the general writing resources mentioned in the rubric video (e.g., how to use the "read aloud" function in word processing software) were helpful.
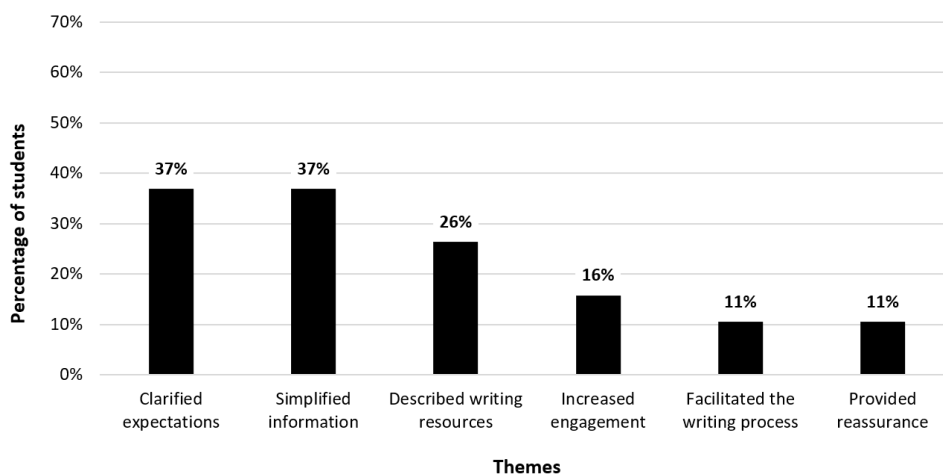

Figure 2. Themes in student comments about the video with tips about the QR writing rubric.

Other students (16%, *n* = 3) shared that their engagement with the QR rubric increased after watching the video. Because the information was visual and conversational, one student said the video "kept my attention. I felt like I was in class, and you were going over it." Another student admitted, "I didn't look at the rubric until watching the video. Once I did, I understood the rubric's importance." A few students (11%, *n* = 2) described how the video facilitated their writing process. For instance, one student said she had trouble starting her case study until she watched the video. Then she "was able to break the assignment down line by line." Another student said the video was "super helpful" because she could pause

it and replay sections as she wrote her paper. A similar percentage (11%, *n* = 2) explained how the video provided emotional reassurance about the assignment. After watching the video, one student said, "I realized I was overthinking the case study questions and was afraid of being incorrect." Another shared that watching the video allowed him "to breathe a bit easier because the instructions and breakdown simplified it immensely."

Discussion and Implications

The study's findings suggest institutional QR rubrics can be successfully modified to create valid and reliable analytic grading rubrics. The new QR writing rubric, which was based on the holistic QuIRK assessment, is intended to grade students' general written arguments. The second new QR rubric includes a data visualization criterion for assignments containing graphs, charts, infographics, etc. The study's findings offer the QR community two new pedagogically focused grading rubrics to help align teaching, grading, learning, and institutional assessment.

QR Rubric Reliability

Because they integrate phrasing from the QuIRK and AAC&U QL VALUE rubrics, the new QR grading tools have good face validity. Analysis of both inter-rater reliability and internal consistency indicate the new rubrics are also reliable. When student artifacts were coded by two raters, the exact percentage of inter-rater reliability as well as Cohen's kappa analysis found substantial agreement between the two coders' evaluations. The internal consistency of the QR rubrics' four criteria also indicated good reliability.

While the current study's findings support the reliability of the new grading rubrics, some variance in ratings occurred during the coding process. This finding was expected given the nuanced nature of QR and the subjective nature of coders' evaluations. To calibrate their ratings during training, coders were provided examples of each criteria's performance levels (see Appendix B). After evaluating a sample of artifacts, raters used a process recommended by Grawe et al. (2010) that had coders review discrepancies then collaboratively decide whether to retain their initial ratings or revise for agreement. This process helped coders reach a consensus on how to differentiate the performance levels for each grading criteria.

Classroom Best Practices

Unlike the holistic QuIRK rubric that combined descriptions of intertwined elements of QR, the new analytic grading rubrics separated QR into four criteria: numerical evidence, conclusions, writing, and data visualization. In addition to providing students with greater clarity about the assignment requirements, the new format encouraged students to reflect on each criterion before they started writing.

Students' responses to the new analytic QR grading rubrics were overwhelmingly positive, with 89% describing how the tools helped them. Students shared how the rubric broke down the larger assignment into manageable pieces, clarified their understanding of the requirements, increased their scores, jumpstarted the writing process, and helped them "breathe a bit easier" when preparing their assignments. To help students leverage the new rubrics when writing arguments, instructors should consider using several pedagogical best practices in their QR classrooms.

Grade transparently. Even when reliable and valid, rubrics are unlikely to promote learning if students are unaware of the criteria the instructor uses to evaluate their work. For instance, Boersma et al.'s (2011) analysis of the Quantitative Literacy Assessment Rubric found the grading tool had strong inter-rater reliability when scoring students' responses to casebook questions. However, students in the Quantitative Reasoning in the Contemporary World course were not provided a copy of the rubric that evaluated their work (Boersma et al. 2011). When uncertain about how their grades are determined, students' assignments represent a "shot in the dark" that may or may not align with the instructor's expectations.

In the current study, copies of the new QR grading rubrics were provided with each assignment in the course's online learning platform. Openly sharing the grading rubrics with students helped them understand how the instructor would evaluate their QR arguments. In the current study, 58% of students who shared comments about the new QR rubrics described how the grading tools clarified their expectations about the requirements. This finding is consistent with pedagogical literature that suggests grading transparency reduces students' perceptions that an instructor's grades are random or unfair (Panadero and Jonsson 2013).

Discuss the rubrics. While sharing the new QR analytic grading rubrics with students is an important first step in the learning process, the current study's findings suggest merely providing a copy of the rubric is not sufficient. Some students reported they did not read the rubric or only skimmed the information. Others felt confused or overwhelmed by what one student described as "a sea of words." QR instructors should not assume students intuitively understand the rubric's criteria and performance levels.

To encourage students to use the new QR grading rubrics when preparing their assignments, the instructor in the current study actively discussed the analytic rubrics before and after each case study. Prior to the first teaching intervention, the instructor presented the new QR writing rubric during an on-campus class. Students were encouraged to ask questions about the expectations for numerical evidence, conclusions, and writing. In the future, teachers may also help students understand what QR looks like at different performance levels for each criterion by sharing examples like the ones in Appendix B.

Provide iterative rubric feedback. In the second week of the course, students were required to write a short paper that analyzed bell curve data about salaries for recent college graduates. After students submitted their first case study online (but before the next on-campus class), the instructor evaluated student assignments with the new analytic QR writing rubric. At the beginning of the next live class, the teacher shared a column chart that illustrated a letter grade distribution for the first case study. The instructor then asked students to use what they learned about bell curve data to interpret the grade distribution chart. During the discussion, the teacher asked questions about central tendency (mean vs. median) and confidence intervals to provide students with an opportunity to apply their bell curve QR skills.

After discussing the aggregated class results, the instructor shared examples of deidentified work from individual students who integrated numerical evidence and presented well-reasoned conclusions. The debriefing about the first case study ended with a discussion about lessons learned from the assignment as well as opportunities for improvement. For instance, after the bell curve teaching intervention, students' use of numerical evidence was better than the pre-course assessment. However, students' conclusions often were underdeveloped and/or asserted. During the class debriefing, examples were shown of well-reasoned conclusions that integrated numerical evidence (the second criterion of the QR writing rubric). Students then reflected about how the conclusions in their first case study might be improved. After the on-campus debriefing about the case study, the instructor posted individual feedback for each student in the university's online learning platform. The relatively small class size allowed the instructor to provide numerical scores, rubric feedback, as well as open-ended comments in each student's assignment. The teacher used the same iterative feedback process after each case study.

Develop engaging rubric resources. Allocating sufficient time to discuss the analytic QR grading rubrics presented the instructor with significant time management issues. The hybrid format of the course, which included only one 75-minute lecture each week, made it difficult to fully address QR concepts as well as allow time to discuss the grading rubric. To overcome this challenge, the instructor developed a 10-minute video with tips about how to use the grading tool. Unlike the rubric alone, which was perceived by some students to be "a sea of words," the video was described by study participants as more conversational, visual, and reassuring. Responses to the video with tips about the QR rubric was overwhelmingly positive, with 95% of students sharing favorable remarks.

Open-ended comments indicated study participants' preferences for rubric resources varied. Some students said they relied primarily on the assignment description when writing their case studies, while others preferred using the rubric and/or video with rubric tips when planning their assignments. Because students'

learning styles differ, instructors should consider providing various resources to keep students engaged with the analytic QR grading rubrics.

Encourage student self-assessment. In addition to helping them understand how their assignments would be graded, students said they used the new QR rubrics to self-assess the quality of their work proactively. Some students reported using the rubric like a checklist to ensure they met all the assignment requirements. Others said they liked being able to rewind and replay the instructor's video explanations to determine if their case study met the requirements.

Differentiating performance levels also helped students self-assess if their work was the "bare minimum" or "mediocre." In this regard, students used the analytic QR grading rubrics as reflective self-assessment tools that allowed them to take ownership of the learning process. The rubrics also permitted students to assess the development of their QR skills over time. Because the QR writing rubric assessed multiple case studies during the course, students were able to compare their criterion scores and check for improvement in their numerical evidence, conclusions, and writing.

Curricular Best Practices

In the current study, two hybrid sections of an undergraduate business course were taught in Fall 2021 and Spring 2022, with about 25 students in each class. Use of the new QR grading rubrics suggested curricular best practices related to learning outcomes, assignment format, and faculty training.

Align learning outcomes. From a teaching perspective, the analytic QR grading rubrics help instructors develop student learning outcomes that support institutional assessment. In the current study, the grading rubrics reliably measured QR in written arguments about a variety of topics. Regardless of the context, each teaching intervention supported three to four core student learning objectives:

- Integrate numerical evidence into written arguments.
- Develop conclusions based on numerical evidence.
- Demonstrate college-level writing.
- Create visual representations of numerical information.

While identifying key elements of QR, the grading rubrics are sufficiently broad to accommodate diverse contexts. For instance, the teacher of an undergraduate special education course could use the grading rubric criteria to write learning outcomes such as: integrate numerical evidence into parent-teacher conferences, develop conclusions about a child's performance based on numerical evidence, and demonstrate college-level writing. Similarly, the instructor of an undergraduate political science course might include outcomes such as: integrate numerical evidence about voting patterns in a congressional election, develop

conclusions about voting based on numerical evidence, and demonstrate college-level writing. Regardless of the context, the criteria in the analytic grading rubrics provide structure as well as flexibility to instructors as they write learning outcomes and develop teaching interventions.

Combine diverse formats. The new analytic QR grading rubrics also provide instructors with a diverse tool that can be used when assessing QR assignments in various formats. In the current study, the pre-course assessment asked students to answer a series of short answer questions. The case studies required a full written paper with references, while the open-ended post-test questions were situated in a timed online final exam. In the current study, each criterion of the grading rubric was equally weighted when evaluating case studies and final exam questions. This approach communicated the importance of effectively combining numerical evidence, conclusions, writing, and data visualization. Because most learning platforms allow teachers to build rubrics into their courses, students' written arguments can be assessed with the study's new analytic QR rubrics regardless of the assignment format.

For instance, the pre-course QR questions in the current study used completion scoring. In addition to receiving full points for submitting the assignment, students also received rubric feedback about their use of numerical evidence, conclusions, and writing. In this low-stakes environment, students were able to identify opportunities for improvement without losing any points. This process communicated the importance of the QR grading criteria and performance levels while also establishing a benchmark for comparison with future assignments. Later in the course, case studies and open-ended final exam questions earned points based on each criterion of the analytic grading rubric.

The use of multiple and diverse assignment formats is important as educators encourage students to develop a QR "robust habit of mind anchored in data" (Steen 2004, 4). The literature consistently indicates there is no "magic bullet" to teach QR. Instead, students need to repeatedly practice their quantitative reasoning skills in various contexts. The current study demonstrated the new analytic grading rubrics can be successfully used to assess QR in low-stake short answer assignments, in-depth case studies, as well as timed open-ended exam questions. An additional benefit to using analytic grading rubrics, like the ones tested in the current study, is these assessments can be used to track the performance of individual students over time.

In situations where a course has multiple sections or high enrollment, the conciseness of the new QR grading rubrics may provide instructors with an efficient way to assess students' written arguments. With a few clicks, an instructor or teaching assistant can generate a score while also providing high-level feedback allowing students to recognize opportunities for improvement. The instructor can

then debrief students during the next on-campus or synchronous online class and/or post asynchronous announcements that summarize observations and suggestions.

Train QR faculty. From an institutional perspective, use of valid and reliable analytic grading rubrics may help colleges and universities improve their overall QR performance. Before they can teach students how to demonstrate QR, instructors must first understand how their college or university will assess students' quantitative reasoning. The interdisciplinary nature of QR often presents institutional assessment challenges because instructors from various disciplines use diverse approaches when teaching and grading student assignments. To coordinate instructors' effort, institutional leaders can encourage teachers to develop learning outcomes that align with the college or university's assessment strategy. Providing instructors with sample learning outcomes and valid/reliable analytic QR grading rubrics may also clarify and reinforce criteria that will be measured in the institution's assessment.

Such a process was employed at Texas A&M University–San Antonio prior to the development and testing of the current study's new QR grading rubrics. The authors collaborated during the university's QR Curriculum Faculty Fellowship program. During the nine-month program, faculty fellows studied the institution's student learning outcomes and reviewed examples of QR assessments. Then, with the guidance of the university's Director of Quantitative Reasoning, instructors developed QR-related learning objectives, designed teaching interventions, and created an assessment plan for one of their courses. The intent of the QR program was to align teaching, grading, learning, and assessment. While institution-wide assessment has not yet been conducted, Texas A&M–San Antonio QR faculty reported the curriculum fellowship process helped them develop interventions to promote student learning while also supporting the university's Quality Enhancement Plan.

Limitations and Future Research

The results of the current study are subject to limitations. Data were collected in an undergraduate business course taught by one instructor at a university in south Texas. The class used a hybrid format, with a 75-minute on-campus lecture once a week and the remaining assignments online. While the QuIRK institutional rubric has been used and tested several times, the current study was the first evaluation of the new QR writing and data visualization analytic grading rubrics. Findings may differ by instructor, course, student enrollment, class format, coders, and/or institution. The first section of the course was taught in Fall 2021, when the delta variant of COVID-19 influenced some students' ability to attend the on-campus sessions. To accommodate those who were unable to participate in the live

instruction, students received a copy of the instructor's lecture slides as well as any class handouts. Future researchers should continue to test the reliability of the analytic QR grading rubrics in different courses that use various formats (fully on-campus, online synchronous, online asynchronous) and represent diverse student populations. The analytic grading rubrics' use in high enrollment QR classes with written assignments should also be explored.

While the new analytic QR rubrics demonstrated good reliability, evaluating assignments with data visualization presented some challenges. Coding discrepancies requiring calibration tended to involve assignments with a data visualization component. Results of some post hoc analysis related to data visualization merit additional investigation. Correlations between the new rubrics' four criteria found data visualization had the lowest correlations with the numerical evidence, conclusions, and writing criteria. Similarly, post hoc factor analysis found the data visualization criterion had the lowest factor loading. These findings may be due to the sample size analyzed for the two new analytic grading rubrics. The QR writing rubric, which measured numerical evidence, conclusions, and writing, assessed all the student artifacts ($N = 415$). However, because only 75 of the QR assignments included a visualization component, the sample size that examined the data visualization rubric was smaller. Future studies should consider testing the data visualization grading rubric on a larger sample of QR artifacts that include charts, graphs, infographics, etc.

The writing criterion of the grading rubrics also presented some challenges. During the training session, the study's coders agreed effective writing requires more than proper mechanics (e.g., spelling, punctuation, grammar, capitalization, etc.). However, raters also believed persistent writing errors tended to distract from the effectiveness of a student's quantitative reasoning. This observation was consistent with Grawe et al.'s (2010) raters who reported written communication and QR are "intertwined" (20). Rather than exclusively assessing written communication based on the frequency of writing errors, the grading rubrics focused on clarity and fluency (see Table 6). When coding, raters occasionally reported difficulty differentiating between the mid-point performance levels on the writing criterion. In the future, researchers should consider modifying the performance levels of the current study's writing criterion to align with the Written Communication VALUE rubric (AAC&U 2009b). When the current study's grading rubrics were originally developed, the authors focused on modifying existing QL/QR rubrics and did not consider integrating information from the AAC&U written communication rubric. The control of syntax and mechanics criterion in the Written Communication VALUE rubric, which has been validated by writing subject matter experts, could be adapted for use when evaluating written arguments with QR. To improve validity when measuring student writing, the authors recommend instructors and future researchers use the updated version of

Table 9
Revised QR Writing Rubric with New Writing Performance Levels

Criteria	1	2	3	4	5
Numerical Evidence Integrated appropriate numerical evidence in the narrative explanations	No use of numerical evidence	The use of numerical evidence is so poor that either it is impossible to evaluate the argument with the information presented or the argument is clearly fallacious.	The use of numerical evidence is sufficient to allow the reader to follow the argument. But there may be times when information is missing or misused.	The use of numerical evidence is good throughout the argument.	The use of numerical evidence is consistently of the highest quality.
Conclusion Presented an appropriate, well-reasoned conclusion based on numerical evidence	Did not draw any conclusions	Presented a significantly underdeveloped conclusion with minimum supporting arguments; numbers are presented but not woven into argument	Presented a sufficient conclusion but lacked discussion of the numerical evidence that made a full evaluation of the argument difficult	Presented a good conclusion that aligned with the numerical evidence; may not have explored all possible aspects of the numerical evidence	Presented a logical and well-reasoned conclusion; interpretation of the numerical evidence is complete and considers all available information
Writing Demonstrated control of syntax and mechanics	Used language that significantly impedes meaning because of errors in usage	Used language that sometimes impedes meaning because of errors in usage	Used language that generally conveys meaning to readers with clarity, although writing may include some errors	Used straightforward language that generally conveys meaning to readers; the language has few errors	Used graceful language that skillfully communicates meaning to readers with clarity and fluency and is virtually error-free

Note: Adapted from Grawe et al. (2010) CC BY-NC and AAC&U (2009b) CC BY-NC-SA 4.0

Table 10
Revised Data Visualization QR Grading Rubric with New Writing Performance Levels

Criteria	1	2	3	4	5
Numerical Evidence Integrated appropriate numerical evidence in the narrative explanations	No use of numerical evidence	The use of numerical evidence is so poor that either it is impossible to evaluate the argument with the information presented or the argument is clearly fallacious.	The use of numerical evidence is sufficient to allow the reader to follow the argument. But there may be times when information is missing or misused.	The use of numerical evidence is good throughout the argument.	The use of numerical evidence is consistently of the highest quality.
Conclusion Presented an appropriate, well-reasoned conclusion based on numerical evidence	Did not draw any conclusions	Presented a significantly underdeveloped conclusion with minimum supporting arguments; numbers are presented but not woven into argument	Presented a sufficient conclusion but lacked discussion of the numerical evidence that made a full evaluation of the argument difficult	Presented a good conclusion that aligned with the numerical evidence; may not have explored all possible aspects of the numerical evidence	Presented a logical and well-reasoned conclusion; interpretation of the numerical evidence is complete and considers all available information
Writing Demonstrated control of syntax and mechanics	Used language that significantly impedes meaning because of errors in usage	Used language that sometimes impedes meaning because of errors in usage	Used language that generally conveys meaning to readers with clarity, although writing may include some errors	Used straightforward language that generally conveys meaning to readers; the language has few errors	Used graceful language that skillfully communicates meaning to readers with clarity and fluency and is virtually error-free
Visual Representation Converted numerical information into a visual format (e.g., chart, graph, infographic, etc.)	Did not visually represent any numerical information; only included words or a table	Converted numerical information but the resulting visual representation is inaccurate	Converted numerical information accurately but the resulting visual representation adds little to no understanding	Clearly and accurately converted numerical information into a visual representation that furthers some understanding	Skillfully converted numerical information into appropriately labeled visual representation that furthers deeper understanding

Note: Adapted from Grawe et al. (2010) CC BY-NC and AAC&U (2009a, 2009b) CC BY-NC-SA 4.0

the QR grading rubrics in Tables 9 and 10, which incorporate phrasing from the AAC&U Written Communication rubric.

Conclusion

More than a decade of work by researchers and educators has built a strong foundation for QR assessment at the institutional level. The growth of QR programs and courses necessitates the development of valid and reliable assessment tools for classroom grading. The current study modified a widely used institutional QR assessment to develop and test two analytic grading rubrics. Both demonstrated good reliability. While additional research is needed, the new QR grading rubrics provide instructors and institutions with the opportunity to align teaching, grading, and learning with programmatic assessment. The importance of supporting institutional efforts with corresponding grading rubrics should not be understated. For "Alone we can do so little; [but] together we can do so much" (Keller 2014, para. 1).

References

Andrade, H. G. 2000. "Using Rubrics to Promote Thinking and Learning." *Educational Leadership* 57, 13–18. http://www.ascd.org/publications/educational-leadership/feb00/vol57/num05/Using-Rubrics-to-Promote-Thinking-and-Learning.aspx

Arter, J. A., and J. Chappuis. 2006. *Creating and Recognizing Quality Rubrics*. Pearson.

Arter, J. A., and J. McTighe. 2001. *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*. Corwin.

Association of American Colleges and Universities. 2009a. "VALUE Rubrics–Quantitative Literacy." https://www.aacu.org/initiatives/value-initiative/value-rubrics/value-rubrics-quantitative-literacy

Association of American Colleges and Universities. 2009b. "VALUE Rubrics Written Communication." https://www.aacu.org/initiatives/value-initiative/value-rubrics/value-rubrics-written-communication

Association of American Colleges and Universities. 2022. "Valid Assessment of Learning in Undergraduate Education (VALUE)." https://www.aacu.org/initiatives/value

Bell, A., R. Mladenovic, and M. Price. 2013. "Students' Perceptions of the Usefulness of Marking Guides, Grade Descriptors and Annotated

Exemplars." *Assessment and Evaluation in Higher Education* 38: 769–788.
https://doi.org/10.1080/02602938.2012.714738

Bobak, C. A., P. J. Barr, and J. O'Malley. 2018. "Estimation of an Inter-rater
Intra-class Correlation Coefficient that Overcomes Common Assumption
Violations in the Assessment of Health Measurement Scales." *BMC Medical
Research Methodology* 18, Article 93. https://doi.org/10.1186/s12874-018-
0550-6

Boersma, S., C. Diefenderfer, S. W. Dingman, and B. L. Madison. 2011.
Quantitative Reasoning in the Contemporary World, 3: Assessing Student
Learning. *Numeracy* 4 (2): Article 8. https://doi.org/10.5038/1936-4660.4.2.8

Bressoud, D. 2009. "Establishing the Quantitative Thinking Program at
Macalester." *Numeracy* 2 (1): Article 3. https://doi.org/10.5038/1936-
4660.2.1.3

Brookhart, S. M. 2013. *How to Create and Use Rubrics for Formative Assessment
and Grading*. ASCD. https://doi.org/10.4135/9781452218649.n15

Brookhart, S. M. 2018. "Appropriate Criteria: Key to Effective Rubrics."
Frontiers in Education 3: Article 22.
https://doi.org/10.3389/feduc.2018.00022

Brookhart, S. M., and A. J. Nitko. 2019. *Educational Assessment of Students*. 8th
ed. Pearson.

Cavanagh, T. M. 2014. "Cheating on Online Assessment Tests: Prevalence and
Impact on Validity" [Dissertation]. *Colorado State University*.
https://scholar.dominican.edu/cgi/viewcontent.cgi?article=1176&context=all-
faculty

Colombini, C. B., and S. Hum. 2017. "Integrating Quantitative Literacy into
Technical Writing Instruction." *Technical Communication Quarterly* 26 (4):
379–394. https://doi.org/10.1080/10572252.2017.1382259

Craver, K. W. 2014. *Developing Quantitative Literacy Skills in History and the
Social Sciences: A Web-Based Common Core Standards Approach*. Rowman
& Littlefield.

Daniels, R. A. 2022. "MKTG 4361—QR Writing Rubric Video." October 24.
Video, 10:33. https://youtu.be/nBy5SfpGRkQ

Daniels, R. A., K. A. Knowles, A. Lindner, and E. Naasz. 2022. "Spiral
Curriculum: Pedagogy for Developing a Quantitative 'Habit of Mind' in
Marketing Students." Manuscript submitted for publication.

Elrod, S. 2014. "Quantitative Reasoning: The Next 'Across the Curriculum'
Movement." *Peer Review* 16 (3): 4–8.
https://www.aacu.org/peerreview/2014/summer/elrod

Grawe, N. D. 2011a. "Beyond Math Skills: Measuring Quantitative Reasoning in
Context." *New Directions for Institutional Research* 2011 (149): 41–52.
https://doi.org/10.1002/ir.379

Grawe, N. D. 2011b. "The Potential for Teaching Quantitative Reasoning Across the Curriculum: Empirical Evidence." *International Journal for the Scholarship of Teaching and Learning* 5 (1): 1–12. https://doi.org/10.20429/ijsotl.2011.050114

Grawe, N. D., N. Lutsky, and C. J. Tassava. 2010. "A Rubric for Assessing Quantitative Reasoning in Written Arguments." *Numeracy* 3 (1): Article 3, 1–21. https://doi.org/10.5038/1936-4660.3.1.3

Gaze, E. C., A. Montgomery, S. Kiliç-Bahi, D. Leoni, L. Misener, and C. Taylor. 2014. "Towards Developing a Quantitative Literacy/Reasoning Assessment Instrument." *Numeracy* 7 (2): Article 4. https://doi.org/10.5038/1936-4660.7.2.4

Gray, J. S., M. A. Brown, and J. Connolly. 2017. "Examining Construct Validity of the Quantitative Literacy VALUE Rubric in College-Level STEM Assignments." *Research and Practice in Assessment* 12: 20–31. https://files.eric.ed.gov/fulltext/EJ1149590.pdf

Hubert, D. A., and K. J. Lewis. 2014. "A Framework for General Education Assessment: Assessing Information Literacy and Quantitative Literacy with ePortfolios." *International Journal of ePortfolio* 4 (1): 61–71. https://eric.ed.gov/?id=EJ1107860

Jastram, I., D. Leebaw, and H. Tompkins. 2014. "Situating Information Literacy Within the Curriculum: Using a Rubric to Shape a Program." *portal: Libraries and the Academy* 14 (2): 165–186. https://doi.org/10.1353/pla.2014.0011

Jensen, K. 1995. "Effective Rubric Design—Making the Most of This Powerful Assessment Tool." *The Science Teacher* 62 (5): 34–37. https://eric.ed.gov/?id=EJ509095

Keller, H. 2014. "The Power of Communication – Helen Keller." *Team Bonding*. February 10. https://www.teambonding.com/power-of-communication/

Koo, T. K., and M. Y. Li. 2016. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." *Journal of Chiropractic Medicine* 15 (2): 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lievens, F., and E. Burke. 2011. "Dealing with Threats Inherent in Unproctored Internet Testing of Cognitive Ability: Results from a Large-Scale Operational Test Program." *Journal of Occupational and Organizational Psychology* 84: 817–824. https://doi.org/10.1348/096317910X522672

McConnell, K. D., and T. L. Rhodes. 2017. "On Solid Ground—VALUE Report 2017." *Association of American Colleges and Universities.* https://www.aacu.org/publication/on-solid-ground-value-report-2017

McHugh, M. L. 2012. "Interrater Reliability: The Kappa Statistic." *Biochemia Medica* 22 (3): 276-282. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/

Miller, L. A., and R. L. Lovler. 2020. *Foundations of Psychological Testing: A Practical Approach.* SAGE.

Nordrum, L., K. Evans, and M. Gustafsson. 2013. "Comparing Student Learning Experiences of In-Text Commentary and Rubric-Articulated Feedback: Strategies for Formative Assessment." *Assessment and Evaluation in Higher Education* 38: 919–940. https://doi.org/10.1080/02602938.2012.758229

Numeracy Infusion Course for Higher Education. n.d. "*Quantitative Reasoning Across the Curriculum: A Multidisciplinary Approach to Teaching QR."* https://serc.carleton.edu/NICHE/qr_across_curriculum.html

Ober, T., K. Gjicali, E. Vianna, and P. Brooks. 2019. "*Quantitative Literacy Across the Curriculum*." March 30. http://teachpsych.org/widget/page-1784686/7252425

Panadero, E., and A. Jonsson. 2013. "The Use of Scoring Rubrics for Formative Assessment Purposes Revisited: A Review." *Educational Research Review* 9: 129–144. https://doi.org/10.1016/j.edurev.2013.01.002

"Recent Downloads." 2022. *Numeracy*. https://digitalcommons.usf.edu/numeracy/readership_map.html#content

Rhodes, T. L., and A. Finley. 2013. "Using the VALUE Rubrics for Improvement of Learning and Authentic Assessment*." Association of American Colleges & Universities.* https://commission.fiu.edu/helpful-documents/competency-based-courses-degrees/using-the-value-rubrics-for-improvement.pdf

Roohr, K. C., E. A. Graf, and O. L. Liu. 2014. "Assessing Quantitative Literacy in Higher Education: An Overview of Existing Research and Assessments with Recommendations for Next-Generation Assessment." *ETS Research Report Series* 2014 (2): 1–26. https://doi.org/10.1002/ets2.12024

Rosen, L. P., L. Weil, and C. von Zastrow. 2003. "Quantitative Literacy in the Workplace: Making It a Reality." In B. L. Madison and L. A. Steen (Eds.), *Quantitative Literacy: Why Numeracy Matters for Schools and Colleges* (pp. 43–52). Woodrow Wilson National Foundation. https://www.maa.org/external_archive/QL/pgs43_52.pdf

Rutz, C. A., and N. D. Grawe. 2010. "Pairing WAC and Quantitative Reasoning Through Portfolio Assessment and Faculty Development." *Across the Disciplines* 6*.* https://doi.org/10.37514/ATD-J.2009.6.1.04

Science Education Resource Center. 2018. "QR Across the Curriculum Profiles." *Carlton College.* May 7. https://serc.carleton.edu/sp/library/qr/across_curric.html

Steele, B., and S. Kiliç-Bahi. 2008. "Quantitative Literacy Across the Curriculum: A Case Study." *Numeracy* 1 (2): Article 3. https://doi.org/10.5038/1936-4660.1.2.3

Steen, L. A. 2004. "Everything I Needed to Know About Averages I Learned in College." *Peer Review* 6 (4): 4–8.

Sun, S. 2011. "Meta-Analysis of Cohen's Kappa." *Health Services and Outcomes Research Methodology* 11: 145–163. https://doi.org/10.1007/s10742-011-0077-3

Sundre, D. L. 2008. "The Quantitative Reasoning Test, Version 9 (QR-9)." *Madison Assessment.* https://www.madisonassessment.com/uploads/Manual-QR-2013.pdf

Sundre, D. L., and A. D. Thelk. 2010. "Advancing Assessment of Quantitative and Scientific Reasoning." *Numeracy* 3 (2), Article 2: 1–12. https://doi.org/10.5038/1936-4660.3.2.2

Taber, K. S. 2018. "The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education." *Research in Science Education* 48 (6): 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tavakol, M., and R. Dennick. 2011. "Making Sense of Cronbach's Alpha." *International Journal of Medical Education* 2: 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Wiggins, G. 2003. "'Get Real!': Assessing for Quantitative Literacy." In B. L. Madison and L. A. Steen (eds.), *Quantitative Literacy: Why Numeracy Matters for Schools and Colleges.* National Council on Education and the Disciplines.

Wolfe, J. 2010. "Rhetorical Numbers: A Case for Quantitative Writing in the Composition Classroom." *College Composition and Communication* 61 (3): 452–475. https://www.jstor.org/stable/40593335

Appendix A: Pre-Test Questions
Salary

Jose, who owns a company in San Antonio, needs to hire a new marketing assistant. Before he advertises the position, he needs to determine an appropriate salary. He wants someone with a bachelor's degree and 3 to 4 years' experience in marketing. When he puts this information into online salary calculator, the website generates the chart below:

10%	25%	50%(Median)	75%	90%
$31,300	$36,496	**$42,203**	$47,790	$52,877

- What does the chart tell Jose about the salary of marketing assistants in his area?
- What salary should Jose list when he advertises for a new marketing assistant? Why is your recommended amount appropriate?

Income Statement

Lori is considering investing in Company XYZ. The table below lists income statement information for the last 3 years.

	2020	2019	2018
Total Revenue	$ 6,600,730.00	$ 6,508,030.00	$ 6,545,257.00
Expenses			
Cost of goods sold	$ 799,567.36	$ 797,562.10	$ 795,564.52
Advertising	$ 78,987.00	$ 69,447.24	$ 68,191.24
Depreciation	$ 61,372.12	$ 39,139.89	$ 45,460.83
Rent	$ 98,996.00	$ 94,367.23	$ 90,921.66
Payroll taxes	$ 79,556.49	$ 64,563.19	$ 63,645.19
Salary and wages	$ 972,768.90	$ 924,658.59	$ 909,424.59
Total Expenses	$ 2,091,247.87	$ 1,989,738.24	$ 1,973,208.03
Net Income	$ 4,509,482.13	$ 4,518,291.76	$ 4,572,048.97

- What does the income statement data tell Lori about the financial situation of Company XYZ?
- Based on the data, should Lori invest in Company XYZ? Why or why not?

Price Elasticity

Jordan owns a local restaurant. His most popular item is a cheeseburger combo, which includes the burger, fries, and a soft drink. The regular price for the combo

is $5.95. In the past, Jordan has offered the combo at a promotional (sale) price of $4.50. Below are the quantities sold at both price points.

Variable	Regular	Sale
Price	5.95	4.50
Quantity sold	198	242

- Based on the data, what price point is best for the cheeseburger combo? Why?

Data Visualization

John owns a small company that sells medical equipment. Average monthly revenue for each of his sales representatives is listed below.

Teresa	$18,456
Jill	$42,340
Carl	$37,450
Mario	$45,534

- Create an appropriate chart to visually represent the information. You may use software (such as Excel) or draw a chart by hand.
- Then provide a written explanation of what conclusions John should draw from the data.

Appendix B: Examples of QR at Different Performance Levels

Numerical Evidence - Integrated appropriate numerical evidence in the narrative explanations

Rating of 0 – Not submitted
Did not submit the assignment
An assignment was not submitted.

Rating of 1 – None
No use of numerical evidence
Even though the sale price sold more, the regular price has a higher profit.

Rating of 2 – Poor
The use of numerical evidence is so poor that either it is impossible to evaluate the argument with the information presented or the argument is clearly fallacious.

The regular price point of $5.95 is better. At the regular price, Jordan can generate more revenue and use less product.

Rating of 3 – Adequate
The use of numerical evidence is sufficient to allow the reader to follow the argument. But there may be times when information is missing or misused. The combo for $4.50. The $4.50 combo made just $90 fewer than the $5.95 combo. The $4.50 combo brought more customers in, which increases your chances of making more sales in the future.

Rating of 4 – Good
The use of numerical evidence is good throughout the argument.
When Jordan sells his cheeseburger combos at the regular price of $5.95, he sells 198 which totals to $1,178.10 in sales. But, when he sells them at a discounted price of $4.50, he sells more combos at 242, which totals $1,089. Although at initial glance it seems the discounted price is the best based off quantity sold, the regular price point is the best because total revenue was $89 more than the sale price. This means that despite the regular price not selling as much compared to the sales price, Jordan is able to sell less and still come out with more total sales by sticking with his regular cheeseburger combo pricing.
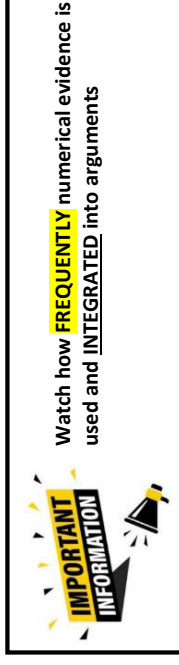
Rating of 5 – Exemplary
The use of numerical evidence is consistently of the highest quality.
The question asked, "Based on the data (not your opinion), what price point is best for the cheeseburger combo?" The answer depends on Jordan's goals for his business. From a revenue perspective, the regular price of $5.95 is best. When cheeseburger combos were offered at the regular price, he sold 198 combos and produced $1,178.10 in total revenue. At the sale price of $4.50, Jordan sold 242 burgers and generated $1,089.00. While 44 more burger combos were sold at the sale price, Jordan generated $89.10 less in total revenue when he discounted the price. If Jordan's goal is to maximize the profitability of his business, he should continue to sell the cheeseburger combos at the regular price.

If, on the other hand, Jordan's goal is to build his restaurant's brand as an affordable option, the sale price of $4.50 may be the best price point for the cheeseburger combo. The lower price allows customers to spend $1.45 less on their meal while receiving the same value. Because many consumers, especially those with large families, are price sensitive, the lower price point may attract new (and potentially repeat) customers to Jordan's restaurant.

Regardless of which option is selected, Jordan should carefully consider his cost of goods sold. The scenario did not provide information about the cost to make a cheeseburger combo or Jordan's net profit. If his cost of goods sold (e.g., bread, meat, pickles, potatoes, condiments, soft drinks, cups, etc.) his high, it is possible that Jordan is losing money at both the sale and regular price points. As a result, before deciding, Jordan should determine the cost of producing a cheeseburger combo. He should also carefully monitor his net profit (as well as total revenue) when making pricing decisions.

Watch how **FREQUENTLY** numerical evidence is used and **INTEGRATED** into arguments

Conclusion – Presented an appropriate, well-reasoned conclusion based on numerical evidence

Rating of 0 – Not submitted
No response was submitted
An assignment was not submitted.

Rating of 1 – No conclusions
Did not draw any conclusions
Jordan offered 2 prices for the cheeseburger combo.

Rating of 2 – Significantly underdeveloped conclusion
Presented a significantly underdeveloped conclusion with minimum supporting arguments; numbers are presented but not woven into argument
The regular price point of $5.95 is better. At the regular price, Jordan can generate more revenue and use less product.

Rating of 3 – Adequate conclusion
Presented a sufficient conclusion but lacked discussion of the numerical evidence that made a full evaluation of the argument difficult.
The best price is $4.50. The $4.50 combo made just $90 fewer than the $5.95 combo. The $4.50 combo brought more customers in, which increases your chances of making more sales in the future.

Rating of 4 – Good conclusion
Presented a good conclusion that aligned with the numerical evidence; may not have explored all possible aspects of the numerical evidence.
When Jordan sells his cheeseburger combos at the regular price of $5.95, he sells 198 which totals to $1,178.10 in sales. But, when he sells them at a discounted price of $4.50, he sells more combos at 242, which totals $1,089. Although at initial glance it seems the discounted price is the best based off quantity sold, the regular price point is the best because total revenue was $89 more than the sale price. This means that despite the regular price not selling as much compared to the sales price, Jordan is able to sell less and still come out with more total sales by sticking with his regular cheeseburger combo pricing.

Rating of 5 – Logical & well-reasoned conclusion
Presented a logical and well-reasoned conclusion; interpretation of the numerical evidence is complete and considers all available information.
The question asked, "Based on the data (not your opinion), what price point is best for the cheeseburger combo?" The answer depends on Jordan's goals for his business. From a revenue perspective, the regular price of $5.95 is best. When cheeseburger combos were offered at the regular price, he sold 198 combos and produced $1,178.10 in total revenue. At the sale price of $4.50, Jordan sold 242 burgers and generated $1,089.00. While 44 more burger combos were sold at the sale price, Jordan generated $89.10 less in total revenue when he discounted the price. If Jordan's goal is to maximize the profitability of his business, he should continue to sell the cheeseburger combos at the regular price.

If, on the other hand, Jordan's goal is to build his restaurant's brand as an affordable option, the sale price of $4.50 may be the best price point for the cheeseburger combo. The lower price allows customers to spend $1.45 less on their meal while receiving the same value. Because many consumers, especially those with large families, are price sensitive, the lower price point may attract new (and potentially repeat) customers to Jordan's restaurant.

Regardless of which option is selected, Jordan should carefully consider his cost of goods sold. The scenario did not provide information about the cost to make a cheeseburger combo or Jordan's net profit. If his cost of goods sold (e.g., bread, meat, pickles, potatoes, condiments, soft drinks, cups, etc.) his high, it is possible that Jordan is losing money at both the sale and regular price points. As a result, before deciding, Jordan should determine the cost of producing a cheeseburger combo. He should also carefully monitor his net profit (as well as total revenue) when making pricing decisions.

IMPORTANT INFORMATION

- **Watch for clear CONCLUSIONS and arguments**
- Look for context and nuance in the interpretation of numerical evidence.
- Be wary of conclusions based on assertions, opinion, or sweeping generalizations.

Writing – Demonstrated college-level writing (clarity, fluency, and writing mechanics)

Rating of 0 – Not submitted
No response was submitted

An assignment was not submitted.

Rating of 1 – Major writing errors distract
Writing errors distract significantly from ideas in the paper

the regular price they sold cheesebuger for 5.95 and quanstify sold w 198 meaning that the totl would be 1,178.10. Thiy sale price i4.50 with 242 sold bieng 1089. Meanng was a $89.10 differnce between teh different pirces.

Rating of 2 – Rushed or confusing
The writing is RUSHED; and/or several ideas are CONFUSING

Regular because moat quantity sold meaning more customers. Hire price is better becaoue moat quantity sold meaning more customers. Without no for sure, I would saw the sale chance of retaining more customers.

Rating of 3 – Rambles or wordy
The writing RAMBLES and/or is WORDY

I would never pay $5.95 for a burger combo. I got for the lowered price. Their lots of other places have options that are more affordable. plus we don't no the reputation of the restaurant. For all we know, there burgers are crappy! So, it really doesn't matter what the price is if the quality is terrible.

Rating of 4 – Clear
The writing expressed MOST ideas clearly

When Jordan sells his cheeseburger combos at the regular price of $5.95, he sells 198 which totals to $1,178.10 in slaes. But, when he sells them at a discounted price of $4.50, he sells more combos at 242, which totals $1,089. Although at initial glance it seems the discounted price is the best based off quantity sold, the regular price point is teh best because of total sales being $89 more than the sales price. This means that despite the regular price not selling as much compared to the sales price, Jordan is able to sell less and still come out with more total sales by sticking with his regular cheeseburger combo pricing.

Rating of 5 – Clear and fluent
CONSISTENTLY expressed ideas clearly and fluently throughout the paper

The question asked, "Based on the data (not your opinion), what price point is best for the cheeseburger combo?" The answer depends on Jordan's goals for his business. From a revenue perspective, the regular price of $5.95 is best. When cheeseburger combos were offered at the regular price, he sold 198 combos and produced $1,178.10 in total revenue. At the sale price of $4.50, Jordan sold 242 burgers and generated $1,089.00. While 44 more burger combos were sold at the sale price, Jordan generated $89.10 less in total revenue when he discounted the price. If Jordan's goal is to maximize the profitability of his business, he should continue to sell the cheeseburger combos at the regular price.

If, on the other hand, Jordan's goal is to build his restaurant's brand as an affordable option, the sale price of $4.50 may be the best price point for the cheeseburger combo. The lower price allows customers to spend $1.45 less on their meal while receiving the same value. Because many consumers, especially those with large families, are price sensitive, the lower price point may attract new (and potentially repeat) customers to Jordan's restaurant.

Regardless of which option is selected, Jordan should carefully consider his cost of goods sold. The scenario did not provide information about the cost to make a cheeseburger combo or Jordan's net profit. If his cost of goods sold (e.g., bread, meat, pickles, potatoes, condiments, soft drinks, cups, etc.) his high, it is possible that Jordan is losing money at both the sale and regular price points. As a result, before deciding, Jordan should determine the cost of producing a cheeseburger combo. He should also carefully monitor his net profit (as well as total revenue) when making pricing decisions.

-
- **Watch for CLARITY and FLUENCY**
  - Consider how writing mechanics may hinder clarify and fluency (e.g., misspelled words, grammar/ punctuation problems, capitalization errors, as well as common writing errors like sentence fragments, subject-verb disagreement, run-on sentences, etc.).

IMPORTANT INFORMATION

Visualization Representation – Converted numerical information into a visual format (e.g., chart, graph, infographic, etc.)

Rating of 0 – Not submitted

No response was submitted

An assignment was not submitted.

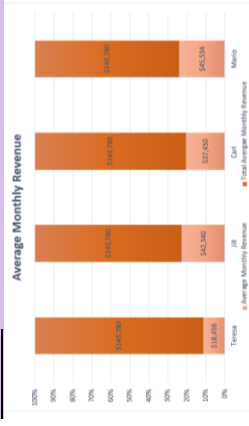## Rating of 1 – No visual representation of data

Did not visually represent any numerical information in a chart

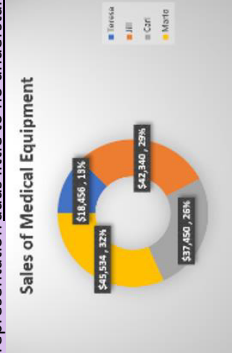| Sales Representative	Average Monthly Revenue
Teresa	$18,456
Jill	$42,340
Carl	$37,450
Mario	$45,534

Note: When coding, a table should not be considered a chart.

Rating of 2 – Inappropriate or inaccurate

Converted numerical information but the resulting visual representation is inaccurate

Average Monthly Revenue

Notice …

- The y-axis is labeled for percentages while the columns display dollars.
- The total bars have the same dollar amount but are different lengths.

Rating of 3 – Partially appropriate or accurate

Converted numerical information accurately but the resulting visual representation adds little to no understanding
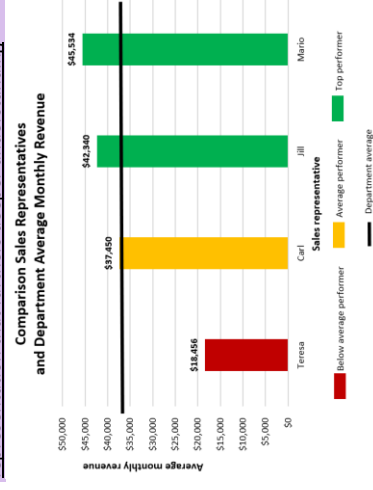
Sales of Medical Equipment

Notice …

- The title of the chart is confusing.
- The pie slices are not sorted, so it is hard to see which categories are the largest and smallest.
- The data labels (which show both values and percentages) are confusing and hard to read.

Rating of 4 – Competent visual representation

Clearly and accurately converted numerical information into a visual representation that furthers **some understanding**

Notice …

- A chart title is not provided.
- The axes are not labeled.
- The columns are not sorted.

Rating of 5 – Skilled visual representation

Skillfully converted numerical information into appropriately labeled visual representation that furthers deeper understanding

Comparison Sales Representatives and Department Average Monthly Revenue

Notice …

- Clear chart title
- Labeled axes
- Labeled values
- Sorted columns
- Intentional use of color.
- Column chart is appropriate for frequency data

IMPORTANT INFORMATION

- **Watch for the chart's ACCURACY and depth of UNDERSTANDING** - Do you clearly understand the main points of the chart?
- Look for clear, concise chart titles, axes labels, data labels, and intentional use of chart format (pie charts for percentages; column or bar charts for frequencies, etc.) and color.