

ADVANCES IN GLOBAL EDUCATION AND RESEARCH

GLO CER '21

VOLUME 4

Editors:

Dr. Wayne B. James

Dr. Cihan Cobanoglu

Dr. Muhittin Cavusoglu



Co-Editors

Dr. Wayne James, University of South Florida, USA

Dr. Cihan Cobanoglu, University of South Florida, USA

Dr. Muhittin Cavusoglu, Northern Arizona University, USA

ADVANCES IN GLOBAL EDUCATION AND RESEARCH: VOLUME 4

ISBN 978-1-955833-04-2

****Authors are fully responsible for corrections of any typographical, copyrighted materials, technical and content errors.***

Co-Editors

Dr. Wayne James, University of South Florida, USA

Dr. Cihan Cobanoglu, University of South Florida, USA

Dr. Muhittin Cavusoglu, Northern Arizona University, USA

ISBN 978-1-955833-04-2

© USF M3 Publishing 2021

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use. The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This imprint is published by USF M3 Publishing, LLC

The registered company address is University of South Florida, 8350 N Tamiami Tr, Sarasota, FL 34243 USA.

Assistant Editor

Dr. Alia Hadid, University of Rhode Island, USA

Editor Assistants

Zahra Alrushdy, Bahcesehir University, Turkey

Gokhan Sener, Necmettin Erbakan University, Turkey

Abraham Terrah, University of South Florida, USA

****Authors are fully responsible for corrections of any typographical, copyrighted materials, technical and content errors.***

Rubrics That Systematically Identify Areas for Improvement

Judy R. Wilkerson¹ and W. Steve Lang²

¹The College of Education
Florida Gulf Coast University, United States

²The College of Education
University of South Florida, United States

Abstract

Commonly used analytic and holistic formats for rubrics have significant limitations in their potential to yield meaningful feedback to students. This inhibits students' potential to improve their work. Consistent with the recommendation of the Gordon Commission for the Future of Assessment in Education that the focus of assessment shift from accountability to improvement, this research presents one easily-implemented step toward that goal. An alternative rubric format that is designed to provide meaningful feedback to students in order to trigger necessary learning and improvement is illustrated, and feedback from students on its utility is documented. The alternative rubric format, which focuses on "Areas for Improvement" or "AFIs," has shown initial success in targeting instructional improvements, but student perceptions of utility have not been sought previously. This article explores those student perceptions. The sample included 38 masters' level students from one institution who had completed complex product assessments, and 14 students from a different university who had completed only a simple discussion forum, totalling 50 graduate students. Results indicate that graduate level teacher education students are highly satisfied with the approach regardless of whether it is applied to major product assessments or simple tasks. This indicates that the format can be used for multiple types of performance and product assessments of varying levels of complexity. Respondents have also indicated support for attempting this technique in their own K-12 classrooms – a future area for exploration.

Keywords: rubrics, quality improvement, assessment

Recommended Citation: Wilkerson, J. R., & Lang, W. S. (2021). Rubrics that systematically identify areas for improvement. In W. B. James, C. Cobanoglu, & M. Cavusoglu (Eds.), *Advances in global education and research* (Vol. 4, pp. 1–12). USF M3 Publishing. <https://www.doi.org/10.5038/9781955833042>

Introduction

We believe that, with a simple shift in perception, we can begin to re-envision assessment not just as a tool for measuring students and sorting them along a hierarchy for the purposes of holding students, teachers, and even entire school systems accountable, but instead as a tool to help us understand the processes of teaching and learning and to improve the quality of our educational interventions—so that we may enable all students to develop to the full extent of their native-born ability (Gordon & Rajagoalan, 2016, p.2).

The Gordon Commission on the Future of Assessment in Education was created to consider the nature and content of American education in this century and how assessment could be used effectively to support teaching, learning, and human development (Gordon Commission, 2013). The Commission's work set the stage for new forms of assessment that improve learning rather

than focusing on whether or not learning occurred by quantifying (well or poorly) how much of it happened.

Many have recognized that the high-stakes testing and accountability movement has resulted in a lack of attention to improvement of instruction resulting from regular assessment of student knowledge and skills. Brookhart (2013) and Elliott and Kahl (2014) argue strongly for formative use. Baker and Gordon (2014) reaffirmed the need for a transformation of assessment purpose and use to address this troublesome issue. Glover et al (2016) noted that summative data is collected late in the academic year and therefore of little use for formative purposes. Maki (2017) is among those who have made a strong case for actionable feedback in post-secondary, as well. The issue is also important for meeting accreditation requirements related to both quality assurance and quality improvement, and this is a challenge that spans the educational system at all levels (Wilkerson, 2019b).

Solomon (as cited in Burke, 2011) concluded that: “It is the analytical potential of rubrics, their ability to pinpoint specific gaps or deficiencies, that may be most useful to the school improvement process” (p. 63). In the literature review that follows, the types and formats of rubrics will be described, followed by some design challenges faced by instructors at all educational levels who wish to use rubrics to as one vehicle to transform assessment from its focus on accountability to feedback for improving learning.

Literature

Rubrics and Rubric Formats

In assessment methods other than traditional tests (e.g., projects, case analyses, essays, portfolios, and live performances), students’ work cannot be evaluated with complete objectivity. Rubrics have been used as an alternative and effective approach to achieve reliable and valid scores regarding students’ performances since the early 1990s (Montgomery, 2002; Reddy, 2011), and the literature on rubrics has grown exponentially. Dawson (2017) traced the history of scholarship on rubrics, noting the publication of hundreds of papers each year with the 5000th paper in 2013. In a recent quick search in ERIC (Proquest) with “rubric” in the title and “design,” “development,” or “format” also in the title, nearly 80,000 results surfaced. There is no dearth of literature on rubrics, but there remain limited sources on specific formats of use across disciplinary and educational level lines.

The Standards of Educational and Psychological Testing (AERA, APA, NCME, 2014) define rubrics as “the established criteria, including rules, principles, and illustrations, used in scoring constructed responses to individual tasks and clusters of tasks” (p. 223). A “rubric” in education literature is commonly understood as an assessment tool that is used to describe and score observable qualitative differences in performances. It captures the essence of performance in academic tasks by “listing the criteria, of what counts, and describing levels of quality from excellent to poor” (Andrade & Du, 2005, p. 1).

There are also many textbook definitions of rubrics, and they usually describe the necessity or process for developing a coherent set of criteria and levels of performance that ostensibly will lead to high quality judgments and feedback (Arter & McTighe, 2001; Nitko & Brookhart, 2008),

whether they are used formatively or summatively and whether they are formatted in holistic or analytic style. Students' need to envision improvements based on descriptive feedback in order to maximize their learning (Arter & McTighe, 2001; Brookhart, 2007; Earl, 2013; Elliott & Kahl, 2014; Stiggins, 2001, 2014; Panadero & Johnson, 2013; Popham, 1997; Wiggins, 1998). Maki (2017, cited in Wilkerson, 2019b) wrote:

Across all conditions, feedback—especially when it is prompt and focused on specific, clearly defined behaviors—is the most powerful tool teachers, coaches, mentors, advisers, and students themselves have at their disposition to improve performance (p. i).

Holistic rubrics result in a single score or rating for an entire product or performance, resulting in a general impression about the quality of work. Analytic rubrics, on the other hand, yield an overall score that is subdivided into specific criteria, each of which is judged and scored separately and then summed. Holistic rubrics are thought to be best for judging simple products or performances, allowing for a quick snapshot of overall quality when detailed analysis is not viewed as important. The result is that they are less useful for diagnostic purposes (Wilkerson et al, 2018; Wilkerson, 2019a,b; Wilkerson, 2021).

Analytic rubrics provide more specific feedback and are generally recommended for classroom, while holistic rubrics are typically used in standardized testing, particularly in the area of writing assessments (Arter & McTighe, 2001; Nitko & Brookhart, 2008; Popham, 1997; Wilkerson, 2019b). The authors of the Joint Standards (AERA, APA, NCME, 2014) note “because the analytic procedure can provide information on a number of critical dimensions, it potentially provides valuable diagnostic information, lending itself to evaluating strengths and weaknesses of test takers” (p. 79). Reddy and Andrade (2010) concluded:

Rubrics are often used by teachers to grade student work, but many authors argue that they can serve another, more important, role as well: when used by students as part of the formative assessment of their works in progress, rubrics can teach as well as evaluate (p. 437).

What rubrics can do and what they actually do are two different things; the design challenge is discussed next.

The Rubric Design Challenge

One of the major flaws related to writing rubrics is lack of clarity in meaningfully differentiating between performance levels; descriptions of rubric development that specify how levels should be written are rare (Reddy, 2011). The analytic rubric format, used most typically in the classroom throughout the K-20 system, provides for every characteristic (or sub-criterion) included in a criterion to rise and fall in quality levels from one proficiency level (or cell in a matrix) to the next (Wilkerson, 2019a,b; Wilkerson, 2021). In other words, all of the sub-criteria in a cell rated as excellent are excellent; all begin to degrade in the next rating level; all degrade even more as ratings become lower until finally everything is either missing or of poor quality. This process ignores the reality that performance on sub-criteria may not be consistently high or low. Some characteristics may be good, while others may be bad. Words like “or,” “may,” or “most” are used, adding even more difficulty to raters needing to decide which deficits occurred and which did not.

Not only do these deficits in the rubrics become a difficulty for raters, but students, too, do not know what was actually weak or good in their work. As the problem compounds, aggregated data (e.g., group means) provide even less insight on specific areas for improvement, so both instructional personnel and their students cannot make valid decisions about what they need to target for improvement (Wilkerson et al 2018; Wilkerson, 2019a,b; Wilkerson, 2021).

Humphry and Heldsinger (2014) concluded that while rubrics are often seen as providing rich information about complex skills, the design features present widespread threats to validity. While these researchers sought a resolution to the problem that focusses on large scale testing and a better (more valid) score for accountability related decisions, their solution is not applicable to the classroom assessment process and does little to help with the feedback issue or assessment FOR learning, as recommended by the Gordon Commission (2013). They sought only to vary the number of proficiency levels (or ratings) per criterion without addressing the content of those levels -- the problem of all high, medium, or low within a given rating category.

Multiple assessment text authors have suggested strategies to improve the effectiveness of rubrics, and each of these attends to the leveling problem just described. Linn and Gronlund (1995) were first almost three decades ago, suggesting that assessment developers list potential minor and major flaws. This idea did not take hold then, and it was tested in an early phase of this research. The term “flaw” was perceived as too negative, so faculty resisted the idea of focusing on deficits entirely, as it was too negative an approach, discouraging students. Linn and Gronlund (1995) did not address the issue of how to differentiate minor from major, which was another important problem in terms of reliability.

Linn and Gronlund’s (1995) idea re-surfaced almost a decade later. Mertler (2003) suggested that the rating process begin with a definition of expectations for exemplary performance. suggesting that “errors” rather than “flaws” be identified and major/minor distinctions be stated explicitly. The general idea of specificity in the rubric, though, was quite similar. Again, almost another decade later, Thorndike and Thorndike-Christ (2010), suggested replacing lengthy trait description definitions with a list of specific behaviors. Stiggins (2017), too, in his discussions of rubrics, noted the need to identify which details were mastered, adding the positive focus of mastery.

A generally set of excellent ideas remained untested in the literature for nearly a quarter of a century from 1995 to 2019. Wilkerson (2019a,b) broke that cycle, suggesting “Areas for Improvement” (AFIs) as a new rubric format. She used Mertler’s (2003) idea of starting by articulating expectations of excellence in “the expected performance” statements. Major and minor deficiencies were noted as the AFIs, with the problem of differentiating minor and major (Linn & Gronlund, 1995; Mertler, 2003) resolved by quantifying them in the proficiency levels. The lengthy trait descriptions were omitted (Thorndike & Thorndike-Christ, 2010), used for only the highest level of performance (Mertler, 2003) and then broken down into specific behaviors (Thorndike and Thorndike-Christ, 2010). The negativity of the “flaws” problem was resolved with the more pro-active “areas for improvement,” consistent with the quality improvement focus in the accreditation field (Wilkerson, 2019b). The utility for program improvement was tested by Wilkerson (2021) with the finding that systematic analysis of frequently occurring AFIs generated modifications in instructional design which, in turn, resulted in improved performance across students in subsequent semesters. The next question that remains unanswered is whether or not

students find AFIs to be of individual benefit in targeting their own successful performances. That is the subject of this research.

Problem Statement and Purpose

Rubrics typically used to make summative decisions about student progression rarely provide students or teachers with information about potential pitfalls before work is submitted or about deficits in learning demonstrated in submitted work. Whether rubrics are analytic (using multiple ratings summed for a total score) or holistic (a single rating or score), they do not yield sufficient detail to improve student work in a meaningful way. Instead, they limit the diagnosis and feedback opportunities so necessary for success (Arter & McTighe, 2001; Nitko & Brookhart, 2007; Popham, 1997), since multiple sub-traits are always embedded in a rating or score. This deficit causes users to miss an important opportunity for student success, and the problem is experienced throughout the K-20 educational system. The use of AFIs has potential for solving this problem, and, although there is preliminary evidence that instruction can be improved prior to assessment, there have been no studies of whether the use of AFIs is perceived by students as helpful in preventing errors or deficiencies in their performance.

The purpose of this research is to describe perceptions of master's level students enrolled in measurement and research courses as part of their master's degrees. Two different levels of complexity were tested – complex assignments (critical tasks in the degree program) and a simple assignment (a routine discussion forum). The overarching research questions were:

- Do practicing teachers find the AFI rubric useful?
- Is the acceptance rate different for minor and major assessments?

Methods

This research is one segment of a larger study of the design and utility of a new rubric technique. Here, the specific methodology is survey. It is relatively simple survey, designed only to get the initial perceptions of students experiencing the method regarding its utility and to obtain feedback on strengths, weaknesses, and suggestions for improvement. Since this rubric format is quite different from those typically used, it is important to track whether or not it works well for students. Selected response items (true/false) and constructed response items were both used (Fink, 2017). Selected response items were analyzed with the number and percent for both choices obtained; constructed response items were analyzed using descriptive coding to identify themes (Saldana, 2014). Sample, data collection, and empirical model will be discussed.

Sample

The survey was administered to 38 master's candidates at one university and 38 master's candidates at the second university. Students were taking similar courses in educational measurement and research. Students at the first university based their responses on the completion of complex assessment tasks in Foundations of Research (EDF 6481, n=32) and Foundations of Measurement (EDF 6432, n=6), while students at the second university based their responses on posts to a graded discussion forum in Foundations of Research (EDF 6481, n=14). All students were enrolled in a master's program in education; students in the first university's EDF 6481 were

enrolled in an accelerated seven-week semester, while the others were enrolled in the traditional 15-week semester.

Data Collection

The following five true/false questions were asked for both populations.

- 1: The lists of "Areas for Improvement" (AFIs) on the rubrics helped me to complete the critical task.
- 2: I checked my work against the AFIs prior to submission.
- 3: If I lost points on a criterion, the AFIs helped me to see what I missed and why. (Do not respond if you had a perfect score.)
- 4: If I had to revise my work, the AFIs helped me know what to fix. (Do not respond if you did not need to revise your work.) (OR "The AFIs will help me to understand course expectations better for the next discussion or activity.")
- 5: I would consider using the AFIs in rubrics I create for my classroom.

Three constructed response items were only used in the first university only, since the second researcher has not yet used the AFI technique for critical task assessment. Question #8 was used at both universities. The questions were:

- 6: What did you like most about the AFI rubrics?
- 7: What did you like least about the AFI rubrics?
- 8: What recommendations to improve the AFI rubrics do you have?
- 9: If you want to try using the AFIs in your classroom, would you be willing to work with Dr. Wilkerson in the future on this? If so, what are you thinking about doing and how would you like to do it together?

Empirical Model: The AFI Technique

The AFI technique deviates from the traditional analytic rubric format by specifying only the expected performance at its highest level regardless of the number of proficiency levels. In that respect, it models faculty expectations that students will do excellent work and succeed; only exceptions are noted. The major difference in this rubric format is that the intermediate and lowest levels are not articulated. Instead, the expected performance is deconstructed into a list of traits that can be highlighted if they are inadequate and classified as minor or major in relation to their contribution to mastery demonstration. The number and severity of AFIs drives the decision about the earned proficiency level.

In this study, the AFI model was tested on two types of assessments. At one university, students completed two complex summative assessments, designated as "critical tasks" (Wilkerson and Lang, 2007); proficiency level ratings were necessary because of the many criteria assessed. At the other university, students completed a simple weekly assignment – a posting in a discussion forum. Only points needed to be provided.

For the critical tasks, proficiency levels and their descriptors were as follows:

- 4=Excellent: The student has met all aspects of the expected performance, demonstrating mastery of the concept. There may be one minor area for improvement, but no major areas of improvement were found. The student has the skills necessary to be an effective teacher or leader.
- 3=Good. The student has met most aspects of the expected performance, demonstrating sufficient mastery of the concept. There may be 2 or 3 minor areas for improvement but no major areas of improvement were found.
- 2=Needs Improvement: The student has met some aspects of the expected performance, demonstrating limited mastery of the concept. There may be 2 or 3 minor areas for improvement but no major areas of improvement were found. The student has the skills necessary to be an effective teacher or leader and will remediate the AFIs independently with practice.
- 1=Unsatisfactory. The student has met some aspects of the expected performance, demonstrating limited mastery of the concept, but there are multiple problems in the work. There are 4 or more minor areas for improvement or one major area of improvement found. The student has limited skills necessary to be an effective teacher or leader and will likely need support on this element.
- .0=Missing. (No response provided.)

The rubrics are provided to students in a seven-column format, which includes the criterion name (for easy reference purposes), the expected performance (a paragraph), the criteria (listed and numbered), the AFIs (minor and major), a weighting factor, and points. In the AFI column, the numbers of the criteria are listed under the relevant minor/major column.

In both the Foundations of Measurement class, one of the two critical tasks requires students to create assessment instruments, including binary items for a traditional test. The criterion is labeled or named “Binary Items.” The expected performance reads: “Three binary response items are written. The cognitive skill tested is specified and is aligned with the items. Directions are clearly written for the students. The items follow proper item writing rules. At least one item has a correct response of false. Answers are provided.” That expected performance is then broken down into six specific criteria that can be listed as AFIs. They are:

- 1: There are 3 binary response items.
- 2: The cognitive skill tested is specified correctly.
- 3: The cognitive skill is aligned with the item.
- 4: Directions are clear.
- 5: Item(s) are written in accordance with proper item writing rules.
- 6: Answers are provided.

An example of the simpler rubric at the second institution illustrates a complete AFI rubric (based on points only); whereas, the previous example was just a segment of a much larger task and rubric. For this simpler tasks students receive directions followed by a list of AFIs. The example directions are: “In the discussion forum, after reading the assigned article, give your definition of the ‘Lake Wobegon Effect,’ and report if you have witnessed or experienced this effect in Florida. Comment on other group members' posts when they are insightful to you. Be sure to

check the list of “areas for Improvement” to make sure you get all 15 of your points. 1-2 points will be deducted per AFI, depending on severity of the error.’ The list of AFIs is:

- 1: An appropriate definition of the Lake Wobegon Effect is provided.
- 2: The definition makes use of the concepts presented by Cannell.
- 3: An experiential account is provided.
- 4: The experiential account is well related to the Effect.
- 5: There is at least one comment regarding other members’ posts.
- 6: The comment is original and does not duplicate what others have written in previous comments.
- 7: The comment is technically accurate.

Findings for Research Question #1

The first research question asked if practicing teachers find the AFI rubric useful. The results from the large critical tasks and the short discussion forum are presented separately and then compared.

Critical Task Survey Results

Across two different masters level classes with two different pairs of critical tasks, Foundations of Measurement (EDF 6432) and Foundations of Research (EDF 6481), a total of 38 students responded; one enrolled student did not respond. The questions and number of students responding “true” are listed below.

- 1: The lists of “Areas for Improvement” (AFIs) on the rubrics helped me to complete the critical task. (35 out of 38 responded true)
- 2: I checked my work against the AFIs prior to submission (36 out of 38 responded true).
- 3: If I lost points on a criterion, the AFIS helped me to see what I missed and why. (Do not respond if you had a perfect score.) (33 out of 38 responded true; Most had perfect or near perfect scores.)
- 4: If I had to revise my work, the AFIs helped me know what to fix. (Do not respond if you did not need to revise your work.) (21 responded true; note that only 14 revised their work.)
- 5: I would consider using the AFIs in rubrics I create for my classroom. (30 out of 38 responded true.)

For the constructed response items, the responses were coded descriptively with themes identified. Results follow for each question:

Question 6: What Did You Like Most About the AFI Rubrics?

Of the 38 students, 33 responded and five did not. After an initial coding in eight categories, three response themes or clusters emerged from the question about what students liked most about the AFIs. Of these respondents, 14 commented on how the AFIs explained feedback and grades including where points were deducted, provided guidance on how to make corrections, and helped them learn from their mistakes. Another 14 comments clustered around clear expectations, step-by-step and specific guidance on how to present their work in ways that were easy to understand,

precise, and organized. Five students specifically wrote about socio-emotional issues, noting the opportunity to self-check, self-reflect, and self-evaluate, with two students noting that the feedback was positive and gave them peace of mind.

Question 7: What Did You Like Least About the AFI Rubrics?

Of the 38 students, 27 responded and 11 did not. The need for coding for this question was limited. Of those 27 students who responded, 18 either indicated “not applicable,” or “nothing that I can think of,” or “great as is.” Three commented that they did not understand what they were supposed to do or that it was initially overwhelming or a lot to process. The remainder of comments were un-related to each other, including four comments about the lack of one-to-one correspondence between the number of points available and the number of AFIs, the AFIs that were repeated across criteria (e.g., followed item writing guidelines for each type of item), that students could get different grades, and that some categories were narrow.

Question 8: What Recommendations to Improve the AFI Rubrics Do You Have?

Of the 38 students, 28 responded and 10 did not. Of those who responded, 19 indicated “not applicable,” “nothing,” or “none.” There was no pattern to the remaining responses. Two students commented on the need for more time (this was in the seven-week course), point values for each AFI (2), more variety and simplification, relevance to content, and embedding the AFIs in the assignment (which was done in green boxes.)

Question 9: If You Want to Try Using the AFIs in Your Classroom, Would You Be Willing to Work With Dr. Wilkerson in the Future on This? If so, What Are You Thinking About Doing and How Would You Like to Do It Together?

Of the 38 students, 26 responded and 12 did not. While some responded not applicable (5) and others expressed concern that AFIs might not be appropriate for young children or SWDs (3); ten expressed an interest in working with these authors to implement at their school.

Overall, the results of this survey are very positive, with 92% of the 33 respondents agreeing that the AFI rubrics helped them to complete the critical tasks and 95% agreeing that they checked their work against the AFIs prior to submission. Only 8% of those who lost points on the total score or had to revise their work found the AFIs as not helpful in understanding what they missed, why they missed it, or what they needed to do to revise their work. They commented that the AFIs helped to clarify instructor expectations by providing step-by-step guidance, noting how AFIs helped them to see the value in self-assessment – an important goal at any age. There was little that they could name that was problematic, and they offered very few recommendations for improvement. Less than 10% found the AFIs to be problematic in any way, and 26% expect to use them in their own classrooms, although a number of students are in classes where they are not as helpful (e.g., for young children and students with disabilities). Some were currently unemployed.

One constructed response question was provided: “What recommendations to improve the AFI rubrics do you have?” Of the 12 students responding to the survey, seven responded to this question. Individual students stated that the AFI rubrics were understandable, solid, appreciated, and very helpful. One student commented that s/he really like them and wished more teachers

would use them. Students also commented individually that the wording was a little confusing, more detail would be needed in some assignments, aesthetic appeal was missing, and there should be a point value per AFI.

Discussion Forum Survey Results

Upon completion of the first discussion forum for the semester, students at the second university were scored using a simplified version of the AFI rubric technique and then surveyed on their perceptions in order to test this process at a different university and with a different type of task. The true/false questions posed were the same as for the critical tasks in the first university with one exception. Item 4, presented in italics, was modified slightly to fit this course. At the time of this writing, 14 of the 38 students had responded to the survey. Support for the process was nearly unanimous, as presented below.

- 1: The lists of “Areas for Improvement” (AFIs) on the rubrics helped me to complete the discussion forum. (14 out of 14 responded true)
- 2: I checked my work against the AFIs prior to submission. (14 out of 14 responded true).
- 3: The AFIs helped me to see what I missed and why. (14 out of 14 responded true).
- 4: The AFIs will help me to understand course expectations better for the next discussion or activity. (14 out of 14 responded true.)
- 5: I would consider using the AFIs in rubrics I create for my classroom. (13 out of 14 responded true)

One constructed response question was provided: “What recommendations to improve the AFI rubrics do you have?” Of the 12 students responding to the survey, seven responded to this question. Individual students stated that the AFI rubrics were understandable, solid, appreciated, and very helpful. One student commented that s/he really like them and wished more teachers would use them. Students also commented individually that the wording was a little confusing, more detail would be needed in some assignments, aesthetic appeal was missing, and there should be a point value per AFI.

Findings for Research Question #2

The second research question asked if the level of acceptance is different for minor and major assessments. At the first University, because a different type of task was used and questions allowed for students not to respond based on their assessment results (i.e., need to revise), the number of students responding “false” is the most appropriate comparison. In the first university, the percent of false responses remained 3% or less for all questions other than use in their own classroom, which was 7%. This higher percent was explained in the constructed response questions based on perceptions that the technique would be less appropriate for young children and Students With Disabilities. Hence, support for the process was nearly unanimous in the first university and unanimous in the second university. No difference was detected, and no statistical inference test was applied.

Conclusions and Recommendations

Not surprisingly, students crave a clarity in the grading that they do not get in the traditional rubric format or the simple points assignment process. The AFI format was found to be highly useful to the graduate students surveyed -- both as learners and teachers especially as tested in an online environment. It guides students in self-assessment that leads to improved initial performance. Although untested in this research, the AFI technique likely has more impact on students' understanding and internalization of course objectives than just reading the objectives alone. Many have expressed interest in testing it in their own K-12 classrooms, particularly at the middle and high school levels. This strategy for rubric design appears to have promise in higher education, as well.

In the future, teachers who test the AFI format should systematically track the response patterns of their students and determine if the format was helpful to them in targeting improvements both at the student and program levels. A variety of school settings in elementary, middle, secondary, and post-secondary should be used to accomplish this task.

References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). Standards for educational and psychological testing. Washington, DC: Author.
- Andrade, H. & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, 10:5, <http://PAREonline.net/getvn.asp?v=10&n=3>
- Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Thousand Oaks, CA: Corwin Press/Sage Publications.
- Baker, E. L., & Gordon, E. W. (2014). From the assessment OF education to the assessment FOR education: Policy and futures. *Teachers College Record*, 116(11), 1-24. Retrieved from <https://www.tcrecord.org/content.asp?contentid=17624>,
- Brookhart, S. (2013). *How to create and use rubrics for formative assessment and grading*. Alexandria VA: Association for Supervision and Curriculum Development.
- Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Upper Saddle River, NJ: Pearson.
- Burke, K. (2011). *From standards to rubrics in six steps: Tools for assessing student learning*. Thousand Oaks, CA: Corwin Press.
- Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education* (42)3, 347-360, <http://dx.doi.org/10.1080/02602938.2015.1111294>.
- Earl, L. M. (2013). *Assessment as learning: Using classroom assessment to maximize student learning* (2nd ed). Thousand Oaks, CA: Corwin Press.
- Elliott, E., & Kahl, S. (2014). *The domain of assessment literacy and its status in teacher education programs*. Washington, DC: American Association of Colleges of Teacher Education.
- Fink, A. (2017). *How to conduct surveys: A step-by-step guide* (6th ed.). Los Angeles: Sage.
- Gordon Commission (2013). *A public policy statement*. New York, NY: Gordon Commission on the Future of Assessment in Education.
- Gordon, E.W., and Rajagoalan, K. (2016) The Gordon Commission and a vision for the future of assessment in education. In Gordon, E.W., and Rajagoalan, K. (Eds). *The testing and learning revolution: The future of assessment in education*. (pp. 1-8). Springer. DOI: 10.1057/9781137519962_
- Glover, T.A.; Reddy, L.A.; Ketler, R.J.; Alexander, L.A.J. (2016). Improving high stakes decisions via formative assessment, professional development, and comprehensive educator evaluation: The school system improvement project. *Teachers College Record* (118)14.
- Humphry, M., & Heldsinger, S. A. (2014). Common structural design features of rubrics may represent a threat validity. *Educational Research* (43)5, pp. 253-263 DOI: 10.3102/0013189x14542154.

- Linn, R., & Gronlund, N. (1995). *Measurement and assessment in teaching* (7th ed.) Upper Saddle River, NJ: Prentice-Hall.
- Maki, P. L. (2017). *Real time student assessment: Meeting the imperative for improved time to degree, closing the opportunity gap, and assuring student competencies for 21st-century needs*. Sterling, VA: Stylus Publishing.
- Mertler, C. (2003). *Classroom Assessment: A Practical Guide for Educators*. London: Routledge.
- Montgomery, K. (2002). Authentic tasks and rubrics: Going beyond traditional assessments in college teaching. *College Teaching* (50)1, 34-40.
- Nitko, A. J., & Brookhart, S. M. (2007). *Educational assessment of students*. Upper Saddle River, NJ: Pearson.
- Panadero, E., & Jonsson, A (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144. doi:10.1016/j.edurev.2013.01.002
- Popham, W. J. (1997). What's wrong – what's right – with rubrics? *Educational Leadership*, 44(2), 72-75.
- Reddy, M.Y. (2011) Design and development of rubrics to improve assessment outcomes: A pilot study in a Master's level business program in India. *Quality Assurance in Education* (19)1, pp. 84-104.
- Saldana, Johnny (2014). Coding and analysis strategies. *The Oxford Handbook of Qualitative Research* (1st ed.). <https://DOI:10.193/oxvodb/9780199811755.016.001>.
- Stiggins, R. (2001). *Student-involved classroom assessment* (3rd ed.) Upper Saddle River, NJ: Prentice-Hall.
- Stiggins, R. (2017). *The perfect assessment system*. Arlington, VA: Association for Supervision and Curriculum Development.
- Thorndike, R.M., and Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education* *8th ed.) Boston, MA: Pearson.
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass Publishers.
- Wilkerson, J.R., and Lang, W.S. (2007). *Assessing teacher competency: Five standards-based steps to valid measurements using the CAATS model*. Thousand Oaks: Corwin Press.
- Wilkerson, J. R. (2019a). Using EPP embedded assessments formatively and summatively for improvement and accountability purposes: A new rubric format that works! Paper presented at the meeting of the American Association of College of Teacher Education, Louisville, KY.
- Wilkerson, J. (2019b). "Rubrics meeting quality assurance and improvement needs in the accreditation context", *Quality Assurance in Education* (28)1, 19-32. <https://doi.org/10.1108/QAE-04-2019-0045>.
- Wilkerson, J.R. (2021). A new rubric format with a focus on "Areas for Improvement" (AFIs). *Journal of Education and Social Policy* (8)2.
- Wilkerson, J., Parfitt, C., Greene, J., Kratt, D., Martelli, D., LaPaglia, K., Johnson, V., Gilbert, S., Zhang, J. (2018). Solving the rubric conundrum: High ratings and accountability for almost all graduates vs. actionable data and continuous improvement for CAEP (and others). Paper presented at the annual meeting of the Eastern Educational Research Association, Clearwater, Florida, February 8, 2018.