

A New Rubric Format with a Focus on “Areas for Improvement” (AFIs)

Judy R. Wilkerson, Ph.D.

Professor, Research and Assessment
Florida Gulf Coast University
10501 FGCU Blvd.
Fort Myers, Florida 33965
USA

Abstract

An original rubric format is described along with the rationale for its development. The format includes an “expected performance” accompanied by potential errors, or “Areas for Improvement,” (AFIs). Ratings are derived from a set of proficiency levels that specify the number and severity of AFIs. A case study illustrates the use of this rubric in a teacher education measurement course and documents the evolution of the rubric and the resulting changes in student performance. Results indicate a change in student performance from nearly all students requiring remediation to a near 100% first time success rate. Instructional improvements based on AFI data and rubric modifications over a five-semester period are provided. The rubric format worked as intended and is recommended for expanded use. Additional research should focus on the utility and replicability of the format.

Keywords: Assessment, Accountability, Continuous Improvement, Program Evaluation, Quality Assurance

1: Introduction

Many of the rubrics currently used to make summative decisions about student progression do not provide students or teachers with information about the specifics of what they can and cannot do well. While some are analytic with multiple ratings summed, and others are holistic leading to a single score, neither format really provides a detailed analysis for student or program improvement. 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 score. This prevents instructional personnel from using specific embedded sub-traits to foster improvement. Of course, other options for teaching and learning are available, but an important opportunity is missed because of this issue. Figure 1 provides an example of this problem with the writing rubric used in K-12 education in many states, including Florida, Ohio, Oregon, Michigan, Mississippi, Missouri, and Texas (Wilkerson, 2019).

Figure 1

Excerpt from Multi-State Writing Rubric for Writing -- Evidence and Elaboration Trait

Multi-State Writing Rubric Trait of Evidence and Elaboration			
4	3	2	1
The response provides thorough and convincing support/evidence for the writer's opinion that includes the effective use of sources, facts, and details. The response includes most of the following:	The response provides adequate support/evidence for the writer's opinion that includes the use of sources, facts, and details. The response includes most of the following:	The response provides uneven, cursory support/evidence for the writer's opinion that includes ineffective use of sources, facts, and details. The response may include the following:	The response provides minimal support/evidence for the writer's opinion, including little if any use of sources, facts, and details. The response may include the following:
<ol style="list-style-type: none"> 1. Relevant evidence integrated smoothly and thoroughly with references to sources 2. Effective use of a variety of elaborative techniques, demonstrating understanding of the topic and text 3. Clear and effective expression of ideas, using precise language 4. Academic and domain-specific vocabulary clearly appropriate for the audience and purpose 5. Varied sentence structure, demonstrating language facility 	<ol style="list-style-type: none"> 1. Generally integrated evidence from sources, though references may be general, imprecise, or inconsistent 2. Adequate use of some elaborative techniques 3. Adequate expression of ideas, employing a mix of precise and general language 4. Domain-specific vocabulary generally appropriate for the audience and purpose 5. Some variation in sentence structure 	<ol style="list-style-type: none"> 1. Weakly integrated evidence from sources and erratic or irrelevant references 2. Repetitive or ineffective use of elaborative techniques 3. Imprecise or simplistic expression of ideas 4. Inappropriate or ineffective domain-specific vocabulary 5. Sentences possibly limited to simple constructions 	<ol style="list-style-type: none"> 1. Minimal, absent, erroneous, or irrelevant evidence from the source material 2. Expression of ideas that is vague, unclear, or confusing 3. Limited or inappropriate language or domain-specific vocabulary 4. Sentences limited to simple constructions

Note. Numbers were added to facilitate reading.

The USA is not alone in this problem. The European PISA rubrics follow a similar analytic pattern, as is illustrated in Figure 2, which includes the two highest levels of the PISA reading rubric.

Figure 2

Two Levels of the 7-Level PISA Rubric for Reading Proficiency

Level	Lower score limit	Characteristics of tasks
6	698	Tasks at this level typically require the reader to make multiple inferences, comparisons and contrasts that are both detailed and precise. They require demonstration of a full and detailed understanding of one or more texts and may involve integrating information from more than one text. Tasks may require the reader to deal with unfamiliar ideas, in the presence of prominent competing information, and to generate abstract categories for interpretations. <i>Reflect and evaluate</i> tasks may require the reader to hypothesise about or critically evaluate a complex text on an unfamiliar topic, taking into account multiple criteria or perspectives, and applying sophisticated understandings from beyond the text. A salient condition for access and retrieve tasks at this level is precision of analysis and fine attention to detail that is inconspicuous in the texts.
5	626	Tasks at this level that involve retrieving information require the reader to locate and organise several pieces of deeply embedded information, inferring which information in the text is relevant. Reflective tasks require critical evaluation or hypothesis, drawing on specialised knowledge. Both interpretative and reflective tasks require a full and detailed understanding of a text whose content or form is unfamiliar. For all aspects of reading, tasks at this level typically involve dealing with concepts that are contrary to expectations.

When writing or reading is scored using a rubric like either of these, the student's work has to show consistent evidence of all sub-traits (numbered or not) to achieve a specific rating. No room is left for a student to be partially at a level four and partially at a level three or to even have one specific area that is at lower level. The rater just makes his/her best judgment of the closest rating. Obviously, there is no feedback on sub-traits that were lower (or higher) than the selected rating.

This problem spans the educational system not only globally in elementary and secondary schools but also in post-secondary education in the USA. The VALUE rubrics used by the American Association of Colleges and Universities (AACU, n.d.) provide yet another example. While much simpler and shorter, they compile multiple features in each rating, again limiting specific feedback.

From a policy perspective, failing to provide helpful feedback is non-responsive and problematic. I presented a solution of using a new rubric format, designated as "AFI Rubrics," with "AFI" standing for "areas for improvement" (Wilkerson et al, 2018, Wilkerson, 2019a, Wilkerson, 2019b). This designation highlights the focus on celebrating successful work while focusing on any specific aspects that could be better.

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. Both institutions of higher education (IHEs) and elementary and secondary schools need to find ways to demonstrate the strong success of their work through student learning (acquisition of required knowledge and skills) while helping both students and teachers to find ways to do what they do well even better. These two conflicting paradigms of formative vs. summative improvement and quality assurance vs. quality improvement present a need for rubrics that meet all four purposes (Wilkerson, 2019a).

The purpose of this research is to illustrate the evolution of an AFI rubric, its data-driven changes, and its impact on instructional improvement, using a case study of one of the major assessments (one of two critical tasks) in one teacher education course, Foundations of Measurement (EDF 6432), at my institution and in my classes.

The technique for "unpacking" the rubrics will be illustrated. Results will be discussed along with both student and instructional improvements that evolved over multiple semesters as the technique was tested. The overarching research question was: "How does the AFI rubric facilitate the provision of actionable feedback to students while identifying needed instructional improvements?"

The significance of this technique is that it provides an enhanced opportunity for student self-assessment and remediation. Faculty can guide improvement at the student, course, and even program levels based on specific data from the rubrics.

2: The Case

The case study centers on an assessment task within a master's level Foundations of Measurement course at a public university in the southeastern United States. The assessment task is a "critical task" in the College of Education. Critical tasks are defined as the most important assignments – ones that graduates must be able to perform successfully on the job (Wilkerson & Lang, 2007).

The complex task requires students to develop standards-based items, both selected and constructed response, for a traditional test, formative assessments, an affective assessment, and a performance-based task. Students also analyze data from a simulated class. In its current form, students work in teams of three, and, as part of the overall project, create items for a unit provided to them from an online Florida resource (CPALMS), a series of lessons and assessments sponsored by the Florida Department of Education. Students are provided with item writing rules and practice their skills on a formative assessment prior to completing the summative task.

For this study example I summarize a portion of the most frequently missed work in the entire task – the writing of selected and constructed response items that meet item-writing guidelines. The rules are simple and straightforward, but many students prefer not to spend the time it takes to follow them. Likely, they hope I will just "forget about it." So, in the past, they have been unhappy that they had to re-do their work. The combination of a template format (instructional and assessment aid) and the new rubric format (for assessment, feedback and improvement purposes) is presented herein. Figure 3 shows the relevant segment of the template the students complete. The example is for the segment on writing binary items. To ensure that each member works on every segment, their assigned items are labeled TM1, TM2, and TM3 for each of the three team members. They succeed (or not) together, since the score is based on all three items. I encourage them to collaborate and review each other's work to preserve a good score.

Figure 3

Template with Directions for Writing Binary Items

Part C: Traditional Test	
<p>C1: Selected Response Items Write three binary items (with at least one item for which the response is false), three multiple choice items, and one matching set with at least six responses to be used. Remember to write directions, provide the correct answers to your items and to make them look like a real test, with a place for your students to respond. For each item type, each team member (TM) must contribute an item; however, the items are graded as a set, so points for mistakes on an individual item are deducted from the total points for everyone. Review each other’s work to assume maximum team points!</p>	
<p>C1a: Binary items: <u><i>Follow all item writing rules; points are deducted if you miss any of the rules.</i></u></p>	
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 5px;">Directions to Students (TM1):</td> </tr> </table>	Directions to Students (TM1):
Directions to Students (TM1):	
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 5px;"> Binary Item 1 (TM1): Knowledge or skill(s) tested: Item: </td> </tr> </table>	Binary Item 1 (TM1): Knowledge or skill(s) tested: Item:
Binary Item 1 (TM1): Knowledge or skill(s) tested: Item:	
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 5px;"> Binary Item 2 (TM2): Knowledge or skill(s) tested: Item: </td> </tr> </table>	Binary Item 2 (TM2): Knowledge or skill(s) tested: Item:
Binary Item 2 (TM2): Knowledge or skill(s) tested: Item:	
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 5px;"> Binary Item 3 (TM3): Knowledge or skill(s) tested: Item: </td> </tr> </table>	Binary Item 3 (TM3): Knowledge or skill(s) tested: Item:
Binary Item 3 (TM3): Knowledge or skill(s) tested: Item:	
<p>Etc.</p>	

In addition to receiving and using the template, students receive the rubric and are asked to self-assess. The rubric is formatted using the AFI technique, which is presented next.

3: 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 our expectation that students will do excellent work and only exceptions are noted so that they can improve where that excellence is not demonstrated. 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. This solves the challenge of raters having to interpret ambiguous terms like, “sometimes,” “may,” “some but not all,” and so forth, and raters do not have to select a rating for work that has some attributes of one rating and other attributes of other ratings. Any of the pre-determined behaviors that are inadequate are then classified as “minor” or “major” deficiencies, and the number and classification of AFIs drives the decision about the earned proficiency level. Figure 4 presents the example of an expected performance with its associated AFIs from the version used in summer 2019 and summer 2020).

Figure 4

AFI Rubric: Expected Performance with Derived AFIs

Five 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.	<ol style="list-style-type: none"> 1. There are 5 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.
--	--

Note: This version was used in Summer 2019 and Summer 2020.

Student work is scored on a four-point rating scale indicating levels of proficiency ranging from excellent (4 points) to unsatisfactory (1 point). The proficiency levels, or ratings are determined based on the number and severity of identified areas for improvement. The scale is:

- 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 found. The student has the skills necessary to be an effective teacher or leader and will likely remediate the AFIs independently with practice.
- 2 = Needs Improvement: 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 areas of improvement found. The student has limited skills necessary to be an effective teacher or leader and will likely need support on this element.
- 1 – Unsatisfactory: The student has attempted the work but there are major deficiencies with at least two major AFIs cited. This deficit needs to be fixed prior to satisfactory completion of the task.
- 0 – Missing.

This is a radically different rubric style, which blends the analytic rubric with a form of checklist. It does not rely on paragraph descriptions of increasingly deficient performance that attempt to reduce proficiency levels into conglomerates of behaviors that may or may not cover all potential results.

Figure 5 illustrate a hypothetical result for a student who made multiple errors in writing the items based on the item-writing rules taught and provided.

Figure 5

Example Results for Binary Items Segment (Spring 2021 version – current)

#	Criterion	Expected Performance	AFIs	AFIs Cited		Pts.
				Minor	Major	
4	C1a: Binary Items	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.	19. There are 3 binary response items. 20. The cognitive skill tested is specified correctly. 21. The cognitive skill is aligned with the item. 22. Directions are clear. 23. Item(s) are written in accordance with proper item writing rules. a. <u>Never use a negative word in a t/f item (e.g., not or never).</u> b. <u>Never provide two questions in one, especially if one is true and the other is false.</u> c. Provide a place for them to respond and tell them how and where to do it. d. Never ask them to write “t” or “f”; you will get a “combo” version when they don’t know – sort of a “T” and sort of an “F.” e. Write some items that are true and others that are false. f. Do not have true statements consistently longer or shorter than false (or vice versa). g. Don’t be tricky or mean. 24. <u>Answers are provided.</u>	24	23a/b	2

This fictitious team did provide the required three items that tested a specific cognitive skill, and they wrote a clear set of directions. However, their items had major problems in terms of following the item writing rules. They violated two rules (a and b). They also forgot to answer their own questions. In my opinion, missing these two rules would be a major problem that needed to be fixed before they could move forward, so I would have asked them to adjust their items. I was less concerned about their forgetting to answer their own questions. Classifying the latter as a minor error

allowed them to have a “needs improvement” rating rather than an “unsatisfactory” rating, which I felt was more appropriate and supportive.

4: Research Method

This study uses a case study design in which the researcher analyzes a single event comprehensively in the real-world context (Baxter & Jack, 2008; Creswell & Poth, 2018; Stake, 1995, 2006; Yin, 2018). Case studies take on a variety of formats (Creswell & Poth, 2018), including, among others the intrinsic case, which must be well-defined, comprehensive, and have a unique significance that focuses on a particular issue or problem. In this study, the problem is the design and use of rubrics that provide sufficient feedback for improvement of both student and instructional improvement, making it an instrumental case study. Because data were collected from multiple individuals (and course sections) over a period of time (spring 2019 through summer 2020 semesters), this study may be classified as a multiple-case-study design. Multiple case studies include more than two individual cases analyzed regarding a similar phenomenon, which are literal replications (Duff & Anderson, 2016; Yin, 2018); it is the design used in this study with the specific classes (and students within those classes) taught considered to be replications. However, each case is important on its own because it both fits into a collection of cases but also has distinct features (Stake, 2006).

There is also an aspect of action research in this design. Action research, as defined by Watts (1985) is a process that allows participants to examine their own education practice systematically and carefully, and that is a clear focus of this study, which is designed to improve my practice. Action research is typically carried out within one’s own environment and designed to inform and change personal practices, it addresses important questions that deal with significant educational matters. McFarland and Stansell (1993) describe the origins of the term which was derived by Lewin, a social psychologist in the 1940s for work that blended investigation with action to solve problems. Pragmatism, as defined by Frey (2018) guided this work because it “sidesteps the contentious issues of truth and reality” (Feilzer 2010, p. 8), and “focuses instead on ‘what works’ as the truth regarding the research questions under investigation” (Tashakkori & Teddlie 2003b, p. 713).

Ferrance (2000) outlines five phases of inquiry in action research: (1) identification of problem area. (2) collection and organization of data; (3) interpretation of data, (4) action based on data, and (5) reflection. Each of these is discussed in the next section along with the presentation of the results of the case.

5: Results

5.1: Step 1 -- Identification of Problem Area

This step was completed in the previous presentations and the viewpoint published in *Quality Assurance in Education* (Wilkerson et al, 2018, Wilkerson, 2019a, Wilkerson, 2019b). This work extends the theory into practice in a specific post-secondary context (teacher education) and adds results.

5.2: Step 2 --Collection and Organization of Data

Using the rubrics previously described, data were collected for three semesters (spring 2019, summer 2020, and fall 2020), and the research results are being confirmed in spring, 2021. The data are in the form of completed rubrics, as well as individualized feedback on the assignments themselves. For the example presented in this paper, the binary choice criterion is presented. In the first version of the AFI rubric, deficiencies were stated in negative terms – what was not done correctly. Note the difference between this earlier version and subsequent ones.

Figure 6

AFI rubric (binary item criterion) used initially in spring 2019.

<p>Three binary response items are written in accordance with proper item writing rules. At least one item has a correct response of false. The cognitive skill tested is specified and is aligned with the items. Directions are clearly written for the students at the 10th grade level. Answers are provided.</p>	<ol style="list-style-type: none"> 1. There are less than 3 binary response items. 2. Item(s) are not written in accordance with proper item writing rules. 3. There is no item for which the correct response is false. 4. The cognitive skill tested is not specified correctly. 5. The cognitive skill is not aligned with the item. 6. Directions are vague. 7. Directions are not written for students at the 10th grade level. 8. Answers are not provided.
--	--

5.3: Steps 3 and 4 -- Interpretation of Data and Actions Based on Data

5.3.1: Phase 1 -- Fall 2014 and fall 2016 (Baseline)

Baseline data were provided prior to the use of the AFI model in fall 2014 and fall 2016, using a simple points-based rubric. In both semesters almost all of the students needed to revise their work but eventually succeeded. Many needed to revise multiple times. In this earlier phase, students had to critique a test of their choice, using the item writing guidelines from the textbook author (W. James Popham). Although guidelines were explicit, many did not remember to include the tests they were reviewing, did not select a test that met task specifications (test was required to have both selected and constructed response items), and/or did not use the required guidelines from the textbook. I was heavily engaged in writing detailed feedback and faced with multiple revisions. Beyond that, the tests they found were typically acceptable, so the exercise did not really help them to find and fix test problems. It was clear that the task needed significant revision. I needed to do the following:

1. Standardize the items analyzed.
2. Provide an opportunity for them to apply the guidelines in writing their own items.
3. Reduce the amount of feedback I needed to write.
4. Change the format of the assignment to make it less free-form.

5.3.2: Phase 2 -- Spring 2019

When asked to teach the course again in spring 2019, I had already developed and tested the AFI model internally in the College's successful CAEP accreditation, so deciding to explore it further in my own courses was a natural evolution of my work. I transitioned to a template with a separate rubric, returning both the template with comments and the rubric to them as feedback. Students were to practice item critiques and development, using a resource "handout" which was included in their module but not integrated into their materials.

Results were better but still not as strong as expected. Students generally did not refer to the handout, and half needed to correct their work; however, my time in providing feedback reduced significantly. I realized I needed to make the guidelines more "visible."

More important, though, is that I had identified some patterns in what they were missing. They performed generally well on standards alignment, formative assessments, and dispositions assessments, but they were generally lacking in item critiquing/writing and performance task development. Of the eight students that semester, errors in item writing were made as follows:

1. Completion: 3 students
2. Multiple choice: 3 students
3. Matching: 4 students
4. Binary 3 students

The students making the errors varied by item type; however, every student made errors in at least one item type. Again, clearly, they were not mastering this skill easily or sufficiently, although my ability to analyze the data and provide feedback was much improved. The problem remained on the instructional side, and clearly an improvement was needed in the preparatory work (formative assessments using "unit templates".) Based on spring 2019 data, it was clear that I needed to do the following: (1) make the item-writing rules more accessible, (2) find an alternative to their working alone and missing so many of the instructions, and (3) restructure formative assignments.

5.3.3: Phase 3 -- Summer 2020

I embedded the rules into the practice work (unit templates) and required them to highlight the rules that were violated. Their work on the critical task improved. While half of the students still struggled with development of performance tasks, of 12 teams, only one made errors in item writing. They were also required to work in teams.

In each semester prior to fall 2020, at least half of the students needed to make revisions, but by fall 2020, only one out of seven teams needed to revise their work, with the technique working better and the instructional materials improved. That team, however, experienced various interpersonal issues which contributed to their product-based deficiencies. I focused on additional areas in need of improvement.

5.3.4: Phase 4 -- Spring 2021 (ongoing)

One more change has been made to the rubric. The item writing criteria are currently embedded directly in the rubric to facilitate identification and use. They are also embedded into the template with a "stop" warning that tells them to review their work using the rubric. The example for a hypothetical student in figure 4 is the version currently in use.

Note that this hypothetical student omitted correct answers in the items written (classified as a minor AFI) but also violated two rules (classified as a major error). This resulted in a “needs improvement” rating for two points (out of four) only.

5.3.5: Step 4: Action Based on Data

Based on the results of the field test in 2014-2016, I did the following:

1. Changed to a template format for student submissions.
2. Provided the items to evaluate and re-write and added an opportunity to create new items.
3. Changed to the AFI format to make feedback less time intensive.

Based on the results of spring 2019, I did the following:

1. Re-organized the work into teams.
2. Embedded the item writing guidelines into the template so they could not miss them.
3. Transitioned to a positively phrased set of AFIs.

Based on the fall 2020 results, I embedded the item writing criteria directly into the rubric for ease of reference.

5.3.6: Step 5: Reflection

The AFI model is easy to write, and, if made highly visible and combined with self-assessment and “can’t miss it” instructions, it does appear to reduce the need to have students re-do their work to master the skills. More specific quantitative data, including retention of initial errors, will help to refine the results and target specific trouble spots. In terms of program improvement, there is a need to keep students more focused on reading and following directions and not racing to the finish line.

6: Conclusions and Limitations

The AFI model is working well in this course (and others that I teach) to guide students in self-assessment that leads to improved initial performance. As both the format of the rubric and the instructional techniques used to teach the content were refined, student performance improved, while I spent less time scoring and re-scoring. The use of highly formatted templates that embed content, combined with scoring on that content within the rubric, reduced the number of students making significant errors and needing to re-do their work. Initially, item writing errors were present in the work of almost all students but, after refinement of the rubric and the accompanying template, almost no students made item writing errors. However, the lengthy list of AFIs can be intimidating. Faculty who think about using the model are concerned that it will take longer to score. In my experience, though, it takes me much less time to grade because I know what I expect, as do the students, and our work becomes extremely focused and efficient.

7: Significance and Recommendations

The AFI rubric process is useful for providing meaningful feedback to students that results in performance changes over time. It also helps users to identify the most common problems in student work so that remediation needs are easier to identify and instructional improvement can be targeted. Additional research on student perceptions and experiences is recommended. A survey regarding student experiences using the format should be the next step in this research.

8: References

- Arter, J. and McTighe, J. (2001), *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*, Corwin Press/Sage Publications, Thousand Oaks, CA.
- Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report*, 13(4), 544-559. Retrieved from https://nsuworks.nova.edu/tqr/vol13/iss4/2?utm_source=nsuworks.nova.edu/tqr/vol13/iss4/2
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches*. Los Angeles, CA: Sage.
- Duff, P. & Anderson, T. (2016). Case study research. In J. D. Brown & C. Coombs (Eds.), *Cambridge guide to language research* (pp.112-119). Cambridge, England: Cambridge University Press.
- Feilzer, M.Y. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research* (4)1, 6-16. doi.org/10.1177/1558689809349691.

- Ferrence, E. (2000). Themes in Education: Action Research. Northeast and Islands Regional Educational Laboratory at Brown University. Providence, RI.
- Frey, B., ed. (2018). Pragmatism, *The Sage Encyclopedia of Educational Research, Measurement, and Evaluation*. Thousand Oaks, CA.
- McFarland, K.P., & Stansell, J.C. (1993). Historical perspectives. In L. Patterson, C.M. Santa, C.G. Short, & K. Smith (Eds.), *Teachers are researchers: Reflection and action*. Newark, DE: International Reading Association.
- Nitko, A.J. and Brookhart, S.M. (2007), *Educational Assessment of Students*, Pearson, Upper Saddle River, NJ.
- OECD (2017). PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving, revised edition, PISA, OECD Publishing, Paris.
<http://dx.doi.org/10.1787/9789264281820-en>.
- Popham, W.J. (1997), "What's wrong – what's right – with rubrics?", *Educational Leadership*, Vol. 44 No. 2, pp. 72-75.
- Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- Stake, R. E. (2006). *Multiple case study analysis*. New York, NY: Guilford Press.
- Tashakkori, A. & Teddlie, C. (2003). *Handbook of Mixed Methods in Social & Behavioral Research*. Thousand Oaks: Sage.
- Watts, H. (1985). When teachers are researchers, teaching improves. *Journal of Staff Development*, 6 (2), 118-127.
- Wilkerson, J. (2019a), "Rubrics meeting quality assurance and improvement needs in the accreditation context", *Quality Assurance in Education*, Vol. 28 No. 1, pp. 19-32. <https://doi.org/10.1108/QAE-04-2019-0045>.
- Wilkerson, J.R. (2019b). Using EPP Embedded Assessments Formatively and Summatively for Improvement and Accountability Purposes: A New Rubric Format That Works! Paper presented at the annual meeting of the American Association of College of Teacher Education, Louisville, KY.
- Wilkerson, J. Martelli, D., Johnston, V., Zhang, J., Wang, C., Gilbert, S., Parfitt, C., Kratt, D., LaPaglia, K., Greene, J. (2018). *Faculty, students, and administrators working together to solve the CAEP psychometric and evaluation challenge through a novel rubric approach*. Panel session presented at the annual meeting of the Eastern Educational Research Association, Clearwater, Florida, February 8, 2018.
- Yin, R. K. (2018). *Case study research and applications: Design and methods*. Los Angeles, CA: Sage.