

# UC Davis

## Journal of Writing Assessment

### Title

The Effects of Student-Fashioning and Teacher-Pleasing in the Assessment of First-Year Writing Reflective Essays

### Permalink

<https://escholarship.org/uc/item/3c26v2hh>

### Journal

Journal of Writing Assessment, 14(1)

### Authors

Pruchnic, Jeff  
Barton, Ellen  
Trimble, Thomas  
[et al.](#)

### Publication Date

2021

### DOI

10.5070/W414155459

### Copyright Information

Copyright 2021 by the author(s). This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

## The Effects of Student-Fashioning and Teacher-Pleasing in the Assessment of First-Year Writing Reflective Essays

by **Jeff Pruchnic** Wayne State University, **Ellen Barton**, Wayne State University, **Thomas Trimble**, Wayne State University, **Sarah Primeau**, University of Illinois at Chicago, **Hillary Weiss**, Wayne State University, **Nicole Guinot Varty**, Wayne State University, **Tanina Foster Moore**, Karmanos Cancer Institute/Wayne State University

The use of reflective essays has become a key artifact of outcome-based writing assessment in the field of writing studies (White, 2005). However, scoring reflective essays may be influenced by textual features irrelevant to most outcomes and assessment rubrics. Two problematic features are teacher-pleasing, which Yancey (1996) called the “schmooze factor,” and student-fashioning, which Miura (2018) related to identity formation. In this article, we present two mixed methods studies to examine the effects of these particular textual features on the direct assessment of first-year writing (FYW) reflective essays. In the first pilot study, we identified four textual features relevant to teacher-pleasing and student-fashioning. In the second quasi-experimental study, we created a sample of FYW essays with and without these features. Two assessment teams then scored the essays in order to determine whether these features had statistically significant effects on assessment scores. The empirical results of these linked studies indicate these features did not have significant effects in a direct assessment of FYW reflective essays. However, in the discussion and conclusion, we argue our mixed methods approach offers a feasible and efficient set of research methods to examine specific textual features in the direct assessments of student writing.

Keywords: writing assessment, first year writing (FYW), reflection, reflective essays, mixed methods

---

*I hate it when they praise the teacher.*  
—Anonymous Assessment Rater, 2019

Having students assess their own learning across a semester, while long an important practice in higher education writing pedagogy, has become increasingly central to writing assessment of the past few decades. One of the reasons self-assessment of this type has become so popular is because it lies at the intersection of two other very important trends in writing studies scholarship and practices during the same time: (a) the identification of reflection as an important skill for student writers (Jankens & Trimble, 2019; Moon, 2004; Taczak, 2016; Yancey, 1998a), and (b) the use of portfolios in the direct assessment of student writing (White, 2005).

However, the very features that make reflective and self-assessment writing so useful might also leave assessment scoring vulnerable to influence by other textual features irrelevant to most learning outcomes and assessment rubrics. One of the most problematic of these features is the tendency for students to insert “teacher-pleasing” statements into reflective writing, such as “Dr. [Name] has done an excellent job.” By teacher-pleasing, students may be tempted to praise their instructor directly and/or to provide the description of their learning experience desired by their instructor, regardless of how accurate that description may be. Yancey (1996) has given this tendency and its potential impact the memorable title of the “schmooze factor.”

At the other end of the spectrum, another problematic tendency is for students to provide exaggerated descriptions of their own writing abilities (or lack thereof) and/or the progress they have made in their class, such as “I am an adept writer.” As Conway (1994), Emmons (2003), and White (2005) have detailed, these “student-fashioning” statements in reflective writing are a common occurrence and are often tied to students’ desires to accommodate the audience of the teacher, with the possible consequence of displacing the purpose of an evaluative reflective essay related to course outcomes. In this case as well, there is a danger that students’ strategies of constructing themselves as particular kinds of writers in explicit descriptions of their abilities might unduly influence the assessment of their performance of those skills in reflective writing assessment.

In this article, we present two mixed methods studies to examine the effects of these features on the direct assessment of student writing. In the first pilot study, we identified four textual features relevant to teacher-pleasing and student-fashioning that might occur in the first-year writing (FYW) reflective assignment we use as our assessment artifact. In the second quasi-experimental study, we developed instances of these features and introduced them under controlled conditions into the texts of student reflective essays. We then had two assessment teams score the original and altered versions of these essays to determine whether the presence of these features had significant effects on assessment scores. The results of these linked studies suggest writing programs need not worry about these particular features skewing the results of a direct assessment of FYW reflective essays, and our mixed methods approach offers a feasible and efficient set of research methods to examine the effects of textual features in direct assessments of student writing.

## Literature Review

### Reflective Letters as Assessment Artifacts

Since the emergence of portfolio assessment in the early 1990s (Elbow & Belanoff, 1986), portfolio reflective letters have been an increasingly important genre in FYW courses (Conway, 1994; Haswell, 2001; Sommers, 2011; Yancey, 1998a). Not only does portfolio assessment serve to support teaching in the composition classroom, it also works productively with theoretical understandings of what writing is and does, such as process theories, rhetorical theory, and contemporary linguistic and literacy theories (Adler-Kassner & O’Neill, 2010). The use of portfolios and reflective cover letters has become an elegant way for writing instructors and administrators to gather information about students’ writing process, their writing progress over time, and their metacognitive steps as they navigate the complex task of learning how to write at the college level.

Many scholars have pointed to portfolio reflective letters in particular as a site of student-ownership over the assessment of their writing (Andrade & Valtcheva, 2009), an artifact that can serve as “the self-assessment that bridges” the internal cognition we call reflection and the classroom- and program-based work of assessment (Yancey, 1998b, p. 13). Yancey (1998b) sees such self-assessment as a crucial step for students to construct themselves and their writing processes and, in so doing, to become less dependent on external evaluation and

grades. She presents her schema of self-knowledge, content knowledge, task knowledge, and judgement to frame the self-assessment she asks students to do in reflective letters. The fourth criteria in this schema, judgement, is the area where students must not only know their own writing well enough to be accurate in their judgments, but they must also “process it” so they can apply knowledge to future writing tasks. Yancey acknowledges this category does not fit neatly into her self-assessment schema. Yet, she still advocates for its use in prompting students’ self-assessment of their own writing, asking educators to shift the dialectic from formative versus summative assessment to internal versus external assessment. The self-assessment students do in reflective letters, she argued, can ultimately collapse that dichotomy.

Not only is the portfolio reflective letter a site of student self-assessment, it also performs assessment functions within courses and programs (White, 2005). White (2005) introduced portfolio reflective letters as a key to scoring portfolios reliably in small, as well as large-scale assessments. When assessments are performed at the program level, it is crucial students understand their own written work in the portfolio as evidence of their learning rather than a springboard into a narrative of progress (White, 2005, p. 591). However, White (2005) maintains that, not only are students able to evaluate their own writing and the course itself, but, With their writing in front of them, they are less likely to praise or blame the teacher for what they did; more often than one would think they ask themselves what they got out of the course and why they didn’t work harder to learn more. (p. 592) Altogether, White’s Phase 2 assessment seems to provide a solution to messy, overly affective or under-critical reflective letters.

White (2005) is clear to point out his Phase 2 scoring is far from a panacea but nonetheless argued it a “major improvement for portfolio assessment in the vast majority of instances” (p. 594). He maintains the reflective letter is itself a genre that provides its share of challenges “even as it simplifies measurement,” noting that, even when provided with the framework of Phase 2 scoring, students may still produce personal narratives or “flattering course evaluations” not in line with the assigned task of synthesizing and providing evidence of learning (White, 2005, p. 594).

In sum, whether situated as a metacognitive bridge and/or a key to reliability for measurement in assessment, the reflective letter of a writing portfolio is a loci of focus for researchers, instructors, and administrators intent on developing best practices in the assessment of FYW reflective essays.

### **Teacher Pleasing in Reflective Letters**

Composition research has pointed to the issue of teacher-pleasing, also known as “schmoozing” (Weiser, 1997) and “the schmooze factor” (Yancey, 1996) in reflective letters. Yancey (1996) defined the schmooze factor as “the temptation to reward students who tell us what we naturally enough want to hear: that they are learning, that they are taking risks...that they have never experienced such pleasure in learning before” (p. 100). The schmooze factor, these researchers found, may be a strategy students use in reflective writing to attempt to persuade their audience favorably.

Sommers et al. (1993) and Zubizarreta (2009) also referred to the effect the schmooze factor has on instructors and raters as “glow,” which Weiser (1997) defined as “the positive effect a particularly strong piece of writing may have on the rating of a portfolio” (p. 300). In Sommers et al.’s (1993) study, some raters read the reflective letter at the beginning of the portfolio, and this may “shed a glow (or shadow) over the remainder of the portfolio” (p. 21). They argued this opening letter caused some raters to score a portfolio higher, even though the portfolio “dropped off in quality” (Sommers et al., 1993, p. 21).

However, Lindenman et al. (2018) found that, even though students sometimes include the schmooze factor in their reflective memos, “they do sometimes enact this progress in their revisions (they walk the walk)” (p. 590). In other words, some students who include teacher-pleasing statements may produce accurate evaluations instead of simply using these statements to try to get a higher grade.

### **Student-Fashioning in Reflective Letters**

In studies on the use of portfolios in the writing classroom, Yancey (1996) and White (2005) focus on the ways students benefit from self-evaluation in reflective letters. For example, White (2005) points out the reflective letter component of Phase 2 portfolios “supports the key principle behind portfolios, that students should be involved with reflection about and assessment of their own work” (p. 583). Although composition scholars recognize the contribution of reflective writing to student learning, they also question ways it may limit student learning. In a study of the self in discourse, Faigley (1989) argued writing teachers judge successful first-person writers on their ability to “create...the illusion of a unified and knowing self that overviews the world around it” (p. 408). He points out the most frequently occurring way of demonstrating the kind of rational knowledge of selves favored by English teachers is “by characterizing former selves as objects for analysis” (Faigley, 1989, p. 411). Similarly, Schendel and O’Neill (1999) argue in their study of self-evaluative student writing, when students are instructed to analyze their growth as writers in a reflective letter for their portfolio.

The self the students construct is part of what is evaluated. The self-assessment works both ways: students have to assess their writing and their selves in creating the portfolio while readers assess the students’ self-assessments and the selves they construct. (p. 211)

Both Emmons (2003) and Conway (1994) highlight some of the potential problems created by cover letter assignments that seem to require students to construct a “self” to be assessed by instructors. Emmons offers a critique of reflective letter assignments based on a study of student responses. In her examination of students’ reflective letters, she identified the narrative of progress as a particular discursive strategy students take up in portfolio reflective letters and essays to respond to prompts asking for descriptions of personal growth as writers. She argues: By rewarding internally focused reflection, our assignments miss the chance to direct students toward a wider view of writing as participation in the work of particular communities. Additionally, with their attention focused on personal growth, students may be unlikely to recognize that their

progress as writers could proceed in a variety of directions depending on the discursive situations they encounter. (Emmons, 2003, p. 44)

Ultimately, Emmons (2003) suggests self-assessment prompts be revised to ask students to “consider the contexts within which [their] progress has been made and against which that progress will ultimately be tested” in order to push students from an internal reflective gaze to an external analysis (p. 45).

Conway (1994) acknowledges reflective cover letters are an opportunity for instructors to learn more about the processes of their students’ writing, but she also expresses concern about the way cover letters may “portray the character and commitment of the writer... [and] convey assumptions about the reader” (p. 84). In an analysis of her students’ portfolio cover letters and her responses to them, Conway (1994) found that, while she had positive feelings in response to cover letters that described the positive impact she had as a teacher and/or changes in their “writing practices and attitudes,” the same letters also demonstrated knowledge of “what it means to think like academic writers” (p. 85). In other words, while students may be putting on a “mask adopted for the purposes of writing a strong cover letter,” she argues it is not wrong or false for them to do so (Conway, 1994, p. 85). After briefly reviewing cover letters that evoked negative feelings, Conway (1994) points out what distinguishes successful cover letters from unsuccessful ones is successful students are “taking responsibility for their own choices” (p. 87), which is, in fact, a move that may require students to construct themselves a particular way in their cover letter. As a result, she argues writing instructors are evaluating their students’ constructions of themselves in addition to their students’ writing in cover letter assignments, and instructors need to be more transparent about this with students. Conway also calls for more scholarly discussions of cover letters and how instructors evaluate them.

### **Purpose of the Research**

As noted by White (2005), Yancey (1996), Conway (1994), and others, very little empirical research has systematically examined the possible effects of students’ self-fashioning and teacher-pleasing features in the assessment of FYW reflective essays. The purpose of the two studies presented here is to use a qualitative-quantitative mixed methods approach to determine the effects of these two textual features on raters’ scores in a large-scale assessment of FYW reflective essays.

### **Methods**

In this section, we first describe our mixed methods approach for the assessment of FYW reflective essays. We then describe the specific research methods of the qualitative-quantitative pilot study (a qualitative rich feature discourse analysis and a test for reliability) and the quasi-experimental study (quantitative methods and statistical analysis).

## **IRB Approval**

All of our assessment studies have been conducted by our Composition Research Committee (CRC). Our university Institutional Review Board (IRB) reviewed this research protocol and determined it was Exempt under 45-CFR-46 categories #1 (research involving normal educational settings and practices) and #2 (education research with de-identified data). Given that the CRC did not collect data directly from students, the IRB did not require a consent process.

## **Site and Context**

Both studies reported on here took place at a large Midwestern urban research university with a diverse undergraduate enrollment of over 27,000 students. Each fall semester, approximately 1,200 students enroll in more than 60 sections of FYW taught by full-time composition faculty and trained graduate teaching assistants (GTAs). Our FYW course uses a common syllabus and an assignment sequence strongly influenced by the Council of Writing Program Administrators' (2014) outcomes statement. Our four FYW learning outcomes are writing, reading, research, and reflection, and the last assignment of the course is a reflective essay inspired by White's (2005) Phase 2 assessment model, in which students evaluate the degree to which they have met the course's learning outcomes, developed with references to written work composed across the semester. (See Appendix A for our FYW learning outcomes and Appendix B for our common reflective essay assignment.)

## **Mixed Methods Approach to FYW Assessment**

Following White (2005), our composition program uses the reflective essay as the main artifact of assessment for our FYW course. Although many FYW programs use consensus scoring as part of small-scale Phase 2 assessments (Haefner, 2011), we use a more robust empirical approach following White et al. (2015), who maintain:

We need to be clear about our reason for advocating empirical techniques of program assessment: Preventing over-simplified measures of writing programs—measures that fail to support instruction—is simply not possible without the use of sophisticated quantitative and qualitative processes ... The use of descriptive and inferential statistics is more effective in administrative meetings than the rhetoric often employed. (p. 114)

Recently, Comer and White (2016) conducted a massive quantitative analysis of an online FYW course, also concluding that mixed methods are essential: "Our experience suggests that Big [Assessment] Data has limited potential without also considering qualitative data" (p. 347).<sup>1</sup> With these expectations in mind, our assessment protocol for this research uses an innovative mixed

---

<sup>1</sup> For useful overviews of mixed methods research, see Creswell and Creswell (2018), Broad (2012), and Haswell (2012). For examples of mixed methods assessment studies with varying topics, design, and sample sizes, see Bourelle et al. (2013), Condon & Kelly-Riley (2004), Elliot et al. (2016), and Pagano et al. (2008).

methods approach to conduct large-scale direct assessment of FYW reflective essays that combines qualitative analysis to examine textual features of interest and quantitative scoring to generate statistically representative data.

Driven by a programmatic commitment to assess a statistically representative number of student artifacts, our quantitative approach to assessment scoring came from an unlikely and, at first, counterintuitive place—a method from behavioral psychology called thin-slice methods (Ambady et al., 2000). Thin-slice methods were popularized as “fast cognition” in Gladwell’s (2005) best-seller, *Blink*, and analyzed in Kahneman’s (2011) award-winning *Thinking, Fast and Slow*. Originally developed to decrease coding time and burden in observational studies of lengthy face-to-face social and institutional interactions, thin-slice methods select and assess relatively small representative “slices” of longer interactions for multiple raters to score with a common rubric. In comparison to full observational coding of entire texts and interactions, such as medical appointments, teaching sessions, and social media profiles, thin-slice methods have proven to be surprisingly reliable. For example, in one of the best known thin-slice studies, raters could predict end-of-term teacher evaluation scores by assessing and scoring a set of three 10-second silent video slices of a teaching session with more reliability than raters assessing and scoring the entire teaching sessions (Ambady & Rosenthal, 1993). Thin-slice methods have been evidenced in a variety of domains, including assessment in higher education (Kulik, 2001). As far as we know, we are the first research team to use thin-slice methods in the domain of higher education writing assessment.

To create thin-sliced data for our assessment of FYW reflective essays, we followed Ambady et al.’s (2000) recommendation that, if a full text or interaction has a definite beginning, middle, and end, the representative thin slices should come from these segments. We thus selected our thin slices from the beginning, middle, and end of FYW reflective essays, which happen to be the traditional categories of essay structure: the first paragraph (introduction), one full paragraph from the middle page of the essay (body paragraph), and the final paragraph (conclusion). For our assessment protocol’s qualitative component, we used rich feature analysis—a method of discourse analysis developed for composition studies (Barton, 2004). Rich feature analysis inductively or deductively looks for textual features that point to the relationship between a text and its context. Rich features have both linguistic integrity (i.e., they are structural features of language and discourse, so they can be defined and then coded, categorized, counted, and otherwise analyzed empirically) and contextual value (i.e., they can be conventionally connected to matters of function, meaning, analysis, interpretation, and significance). Rich feature analysis is aligned with other textual analysis methods in writing studies, including Haas et al. (2012), Huckin (1992), and Grant-Davie (1992).

To establish proof of concept for our mixed methods approach to the assessment of FYW reflective essays, we published our theoretical and empirical arguments in the *Journal of Writing Assessment* (Pruchnic et al., 2018). Our quantitative results showed that assessment raters reading and scoring an entire reflective essay and raters reading and scoring a thin-sliced essay were correlated; that is, both assessment teams scored the essays similarly to an extent: Pearson’s  $r = 0.46$ , a low positive correlation, but trending toward high positive (0.50). According

to Cicchetti (1994), however, the thin-sliced team was notably more reliable (0.76, excellent reliability), while the full essay team's reliability was (0.60, good reliability, but trending strongly toward fair reliability (0.59). Further, time for scoring was significantly different: The full essay team took 11 hours and 45 minutes for the assessment while the thin-sliced team cut that time in half (5 hours and 20 minutes). Finally, on the thin-slice team, five raters scored each essay, but on the full essay team only two raters scored each essay.

For our qualitative analysis, we conducted a rich feature analysis of essays that did or did not achieve the reflection outcome. We then offered related recommendations to our curriculum committee to develop a more robust pedagogy of reflective writing in FYW, including more direct instruction on the genre of reflective writing as an academic argument and adding short post-assignment reflections throughout the course, which we have incorporated in our common syllabus, thereby closing the assessment loop with evidence-based change (Walvoord, 2010). Turning to the field of writing studies, we argued the affordances of thin-slice methods offer excellent reliability and ample evidence of validity in order to ask and answer quantitative questions about textual features in FYW assessment. We also argued the affordances of rich feature analysis offer deep qualitative contextualization of empirical findings in order to pose evidence-based curricular and pedagogical changes. We concluded our approach allows the field to achieve a high quality large-scale direct assessment of textual features in student writing when that is warranted in a particular assessment context.<sup>2</sup>

### **Methods for the Pilot Study**

The research reported here aimed to determine the effects of two textual features on assessment scoring: student-fashioning and teacher-pleasing. For our pilot study, we first sought to confirm the observations in the assessment literature that FYW reflective essays include features in which students make general evaluations of their own abilities as writers (self-fashioning) and their favorable evaluations of their instructors (teacher-pleasing). The specific research questions of the pilot study were the following:

Were student-fashioning and teacher-pleasing features present in FYW reflective essays?  
Can assessment raters reliably identify these features in FYW reflective essays?

### **Sample for the Pilot Study**

The data for the pilot study came from a 2017 corpus of FYW reflective essays. Course instructors submitted 1,377 reflective essays, and the CRC team then created a representative, randomized sample of 291 essays. Our project manager, a member of the CRC team, read 150 of the 2017 corpus essays and selected 18 essays she determined had student-fashioning and teacher-pleasing features for the purpose of developing a coding schema to identify these features in FYW reflective essays in our assessment. (She discarded 14 essays because they included both types of features, which was not part of our research design.)

---

<sup>2</sup> In this JWA article, we were emphatically not making a claim that our mixed methods approach, including the use of quantitative thin-slice methods, is right for every assessment context (see the discussion and conclusion below).

## Coding Schema

The members of the CRC team independently read the 18 reflective essays to identify student-fashioning and teacher-pleasing features. We then used rich feature analysis in group discussions to compare, discuss, develop, and categorize four discrete features of student-fashioning and teacher-pleasing (Table 1). We created two self-fashioning categories to capture writers' explicit evaluations of themselves as good or bad writers. We also created two teacher-pleasing categories to capture student writers' explicit comments about their instructors as good or bad teachers.

**Table 1**  
*Self-Fashioning and Teacher-Pleasing Coding Schema*

Code	Description	Examples
1	Good writer	I have always thought I have been a pretty decent writer. I'm already a great writer.
2	Bad writer	I am not the best writer. I did not think of myself as a good writer.
3	Good teacher	I had a wonderful teacher. Dr [Name] has done an excellent job.
4	Bad teacher	No examples

## Participants

Six members of the CRC participated in the pilot study: three full-time faculty (one professor and two lecturers) and three GTAs.

## Norming

Our assessment coordinator, also a member of the CRC, normed the teams on the outcome-based coding schema, using examples from the 18 essays selected by the project manager.

## Materials

Members of the CRC team received the coding schema and the set of thin-sliced essays in the sample (n = 259).

## Data Collection

Five raters scored each essay. For each essay, raters scored yes to indicate the presence of each feature of our coding schema (good writer, bad writer, good teacher, bad teacher). The project manager then calculated the means of the five scores to assign the final score for each essay in the sample.

## Data Analysis

To analyze our data, we first measured inter-rater reliability (IRR) to determine the degree to which CRC raters were able to reliably identify student-fashioning and teacher-pleasing features in the sample of FYW reflective essays. We used Hallgren's (2012) intra-class correlation coefficient (ICC), which measures the degree of agreement among multiple raters' scores. To interpret the reliability data, we followed Cicchetti's (1994) classifications: excellent ( $> 0.75$ ), good (0.60 to 0.74), fair (0.40 to 0.59), and poor ( $< 0.40$ ).

We then calculated the total number of essays with student-fashioning and teacher-pleasing as well as the frequency distributions of each code in the coding schema.

## Results and Discussion of the Pilot Study

Our CRC team achieved an overall reliability of 0.79, which Cicchetti (1994) classifies as excellent, indicating raters were reliable in their identification of student-fashioning and teacher-pleasing features in FYW reflective essays. This level of reliability provides evidence of the validity of the rubric and scoring.

In the sample of 259 essays, 162 (62.5%) had no student-fashioning or teacher-pleasing features, and 97 (37.4%) included student-fashioning or teacher-pleasing features. Over one third of the reflective essays included these features, indicating these are relatively common textual features as the FYW assessment literature suggests.

Table 2 presents the frequency distributions of the four codes in the coding schema.

**Table 2**

*Frequencies of Student-Fashioning and Teacher-Pleasing Features (N = 99)*

Codes	Description	<i>n</i> (%)
1	Good writer	17 (17.5)
2	Bad writer	43 (44.3)
3	Good teacher	37 (38.1)
4	Bad teacher	0 (0)

The most frequent feature in the reflective essays was a negative self-fashioning as a bad writer, perhaps an indicator of a narrative of progress in the genre of FYW reflective essays (Emmons, 2003; White, 2005). Interestingly, no essays included any bad teacher features, which may indicate some student writers are reluctant to criticize their instructors in a course assignment although no bad teacher features may instead indicate some student writers were

accurate in evaluating their instructors, as suggested by White (2005), Lindenman et al. (2018), and Emmons (2003).

In sum, our answers to the pilot research questions were positive for the presence and identification of self-fashioning and teacher-pleasing features in our sample of FYW reflective essays. However, even though the thin-sliced sample was large and the identification of these features was reliable, the frequencies of the individual codes were too small to determine any statistically significant effects on the scores of essays containing student-fashioning or teacher-pleasing features compared to essays containing no such textual features.

### Methods for the Quasi-Experimental Study

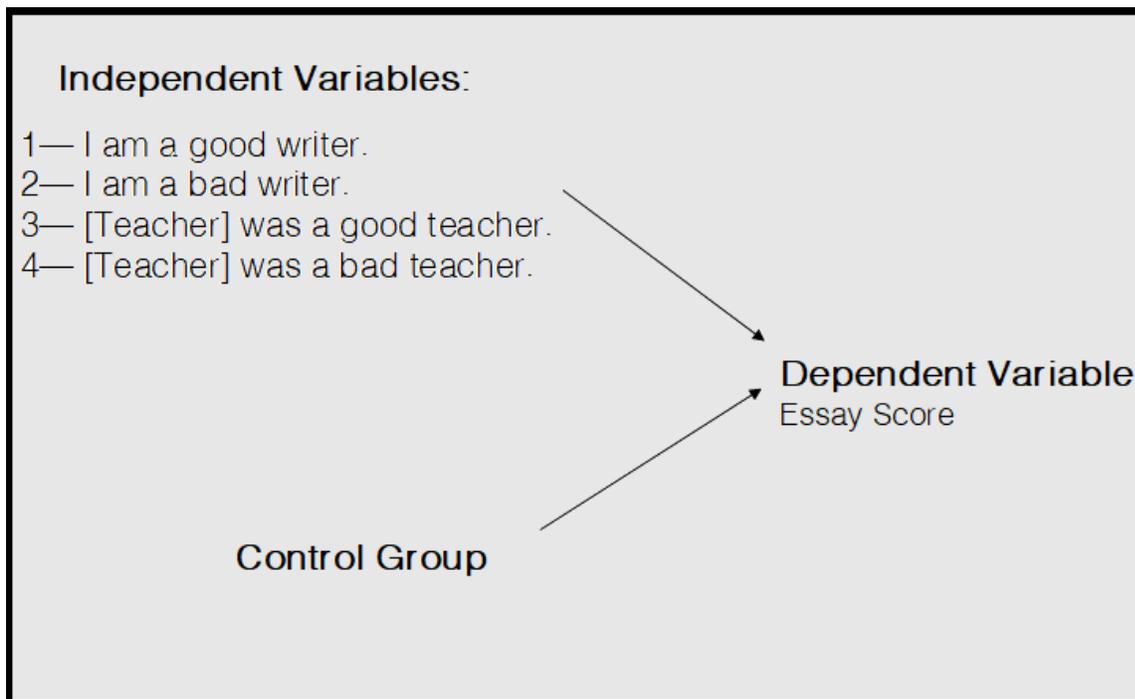
For our second study in this research, we developed a quasi-experimental research design. The specific research question of the quantitative study was the following:

1. Did student-fashioning and teacher-pleasing features have significant effects on the overall mean scores in an assessment of FYW reflective essays?

To answer this question, we created a sample that would allow the CRC team to compare raters' scores for two versions of the same essays. One set of essays was in their original form with no features. Another set of the same essays was seeded with student-fashioning and teacher-pleasing features. For the quasi-experimental design, the textual features of self-fashioning and teacher-pleasing were our independent variables, and the overall mean score was the dependent variable.

#### Figure 1

*Independent and Dependent Variables Research Design*



This research design is quasi-experimental because we constructed a sample that was not a randomized, representative sample of the entire population of FYW reflective essays (Creswell & Creswell, 2018, p. 250). Instead, we used a power analysis to create our sample.

### Sample for the Quasi-Experimental Study

The data for the quasi-experimental study came from a 2019 corpus of 788 FYW reflective essays. The CRC team’s statistician conducted a power analysis to determine the required sample size to detect a statistical effect in the sample: A power of 0.80 indicates the probability a study will obtain a statistically significant effect eight times out of 10 (Creswell & Creswell, 2018, p. 151). For our study, the power analysis determined a sample size of 136 essays.

### Materials for the Quasi-Experiment Study

The CRC project manager first read reflective essays from the 2019 corpus until she was able to find 136 thin-sliced essays that did not contain any instances of the independent variables of self-fashioning and teacher-pleasing. These 136 essays served as our control group in the sample.

The project manager then copied the set of 136 essays and inserted student-fashioning and teacher-pleasing variables. These seeded thin-sliced essays served as our experimental group in the sample. Table 3 lists the distributions of these essays.

**Table 3**  
*Distribution of Independent Variables (N = 136)*

Features	n	Team 1		Team 2	
		Control	Experimental	Control	Experimental
Good writer	34	17 no features	17 features	17 no features	17 features
Bad writer	34	17 no features	17 features	17 no features	17 features
Good teacher	34	17 no features	17 features	17 no features	17 features
Bad teacher	0				

### Participants

We recruited 10 assessment raters who were not members of the CRC team: six full-time faculty lecturers, three GTAs, and one part-time faculty member. We split participants into two 5-person teams.

### Rubrics

We chose to assess two of our FYW learning outcomes – writing and reflection – as the most likely outcomes to be affected by the presence of student-fashioning and teacher-pleasing. For

these outcomes, we used rubrics previously developed and tested for reliability and validity in previous assessment research (Barton et al., in press; Pruchnic et al., 2018). In developing our rubrics, we make a distinction between our WPA-influenced course outcomes, which are complex and comprehensive (see Appendix A), and our specific assessment research outcomes, which are focused and empirical.

For reflection, our assessment outcome was the following: “Use written reflection to evaluate one’s own learning and writing.” We assessed two specific traits of reflection in the rubric: (a) argument, defined as thesis, claims, and relation to course outcomes, and (b) evidence, defined as examples, analysis, experiences, and discussion. Our rubric used a 6-point Likert scale to assess the reflective essays (6 = high, 1 = low). Our reflection rubric is presented in Appendix C.

For writing, our assessment outcome was the following: “Use conventions of academic discourse.” We assessed two specific writing traits in the rubric: (a) effective paragraph organization (defined as topic sentence and paragraph development), and (b) editing (defined as grammar, punctuation, spelling, and awkward sentences). Our rubric included a 6-point Likert scale to assess the reflective essay (6 = high, 1 = low). Our writing rubric is presented in Appendix D.

Both rubrics had a decision rule directing raters to record the lower score across the two traits: For example, in the reflection assessment, a score of 5 for argument and a score of 4 for evidence would receive a final score of 4; in the writing assessment, a score of 3 for paragraph organization and a score of 5 for editing would receive a final score of 3. We asked raters to record the lower score of the two traits to identify textual features in essays that may not be scored as achieving the outcome (1, 2, 3 in our rubrics).

### **Norming**

As part of our earlier assessment research, we developed a best-practices norming process for assessing FYW thin-sliced reflective essays (Barton et al., in press). Our assessment coordinator first introduced reliability as one of the primary goals of assessment and defined reliability as a measure of consistency across raters. He went on to introduce the rubrics for reflection and writing, explaining they were developed using a “storming and norming” process by the CRC (Colombini & McBride, 2012). He also discussed thin-slice methods as part of our mixed methods approach to FYW assessment. The facilitator instructed raters to focus solely on the language of the rubric to score their essays, deferring consideration of other aspects of reflection and writing (e.g., voice, narrative, readiness for the next course, or personal heuristics), acknowledging this is a difficult task for experienced composition instructors. The facilitator then discussed two thin-sliced essays didactically, modeling scoring to each one of the rubrics. He then asked raters to score an essay using the reflection and writing rubrics and posted the scores on a whiteboard. The group used visual inspection to display the reliability of scores by looking for non-contiguous scores: If two adjacent scores varied by more than one point, the facilitator led the group in a short discussion, again emphasizing the language of the rubrics. This process was repeated with another essay.

## **Scoring**

Upon completion of the norming process, raters were instructed to use the rubric to score their set of thin-sliced reflective essays independently, without consultation with other raters or any discussion at the table.

## **Data Collection**

Each of the 136 seeded essays was scored by five raters. The CRC project manager calculated the mean of the five raters' scores to assign a final mean score for reflection and a final mean score for writing for each of 136 essays in the sample.

## **Data Analysis -- Reliability**

We used Hallgren's (2012) ICC to measure consistency across raters' scores. To interpret the ICC results for raters' scores on both the writing and reflection learning outcomes, we again followed Cicchetti's (1994) classifications: excellent reliability ( $> 0.75$ ), good reliability (0.60-0.74), fair reliability (0.40-0.59), poor reliability ( $< 0.40$ ).

## **Data Analysis – Statistical Tests**

To determine whether the presence of student-fashioning and teacher-pleasing features had significant effects on raters' overall means for the reflection and writing outcomes, our statistician selected and conducted three statistical tests. First, she calculated the R squared value, which indicates the variance accounted for in the experimental groups. Second, she performed an analysis of variance (ANOVA), which indicates whether the set of independent variable assignment was significant as a group. Third, she performed a regression analysis by calculating regression coefficients to determine if any one independent variable affected the mean score of each assessed outcome.

## **Results of the Quasi-Experimental Study**

### **Reliability and Validity**

Following Cicchetti (1994), our raters demonstrated excellent reliability in scoring for the reflection outcome (0.79). Our raters demonstrated good reliability for the writing outcome (0.74) but trending strongly to excellent reliability (0.75). As in the pilot study, these levels of reliability provide evidence of the validity of the rubric and scoring.

In addition to reliability-based validity, evidence of validity for the assessment methods used in this study have been tested and affirmed in previous studies. We have confirmed strong correlation between the scores of full essays and the scores of paragraphs from the same essays isolated by thin-slice methods, as well as the validity of our rubric and our norming process; we have also successfully used assessment findings derived from these methods to create and implement curricular and pedagogical changes (Barton et al., in press; Pruchnic et al., 2018). Similarly, while only a single project from each student is used for the purposes of program assessment in our methods, this project (the reflective essay) requires students to describe and provide examples from work they have written across all required genres in the

course, and our program has previously tested and affirmed strong correlation between the assessment scores of the reflective essay and full portfolios of work by the same student.

### **Statistical Analysis**

To determine whether student-fashioning and teacher-pleasing features had significant effects on raters' overall means for our reflection outcome, our statistician calculated an R squared value of .045, which indicates 4.5% of the variance across scores could be attributed by independent variable assignment. Her ANOVA analysis, which relates to how the mean scores across independent variable groups varied, resulted in an F test value of 2.079, which indicates essay scores based on assignment were not significantly different for writing across the independent variable groups. Finally, her regression analysis, which indicates if any one independent variable group is affecting mean scores, resulted in t test values that were not statistically significant.

Her results when analyzing scores for our reflection rubric were similar. Analysis yielded an R squared value of .028, which indicates 2.8% of the variance across scores can be attributed by variable assignment for reflection. Our F test, which resulted in a calculated value of 1.274, was also not significant. Finally, our regression analysis generated t test values which were also not significant.

Overall, none of the differences were statistically significant, indicating the presence of student-fashioning and teacher-pleasing features had no significant effects on raters' overall means for the reflection and writing outcomes.

## **Discussion and Conclusion**

### **Discussion**

These results answered our qualitative-quantitative research questions by determining student-fashioning and teacher-pleasing features in reflective essays do not have significant effects in our assessment of FYW reflective essays. The immediate implication of our research is that, at least for the purposes of large-scale assessment, there is no reason to develop a curricular or pedagogical intervention related to these all-too-human writing strategies, as suggested by White (2005), Yancey (1996), Lindenman et al. (2018), and Conway (1994).

There are, however, three important methodological implications of our research for FYW assessment. First, we believe our research speaks to the value of publishing studies that report non-significant results. That is, even as we acknowledge that studies with non-significant results present rhetorical challenges for both writers and publishers, knowing that X does not cause Y can be both important and worthy of circulation for the field.

Second, we believe this research contributes to the growing literature on mixed methods research in the assessment of writing in higher education (White et al., 2015). In particular, we believe our research demonstrates the value and potential of working with statisticians who can provide feedback about what kinds of questions can and cannot be answered in statistical tests,

advice on advanced mixed methods designs, and technical expertise to analyze data of importance to the field of writing assessment. This may seem like an obvious point; however, our experience demonstrates that, even while it may not be possible for all writing researchers to perform advanced statistical analyses themselves, having a basic vocabulary with which to converse with statisticians and the work they do can be of immense value to assessment researchers and administrators.

Third, we believe our research demonstrates the feasibility and utility of using mixed methods to answer specific questions about textual features in FYW reflective essays and other writing assignments. Our mixed methods approach is pragmatically aimed at providing efficient and high-quality evidence for programmatic decisions about curriculum and pedagogy through a combination of statistically representative samples, specific assessment outcomes, well-defined rubrics normed for reliability, contextualized norming protocols, quantitative statistical analysis, and qualitative rich features analysis. As noted above, however, we are not making a claim that our mixed methods with statistical analysis is right for every assessment context. Other programs may wish to conduct small-scale assessments using thin-slice methods to identify and describe interesting textual features with basic counting, if desired, and to use qualitative analysis of multiple types, including rich feature analysis, thematic analysis, and grounded theory. In sum, we believe our mixed methods approach expands the ability of researchers and administrators to develop and conduct assessment studies that do not need the considerable investment of time and resources involved in full essay scoring.

## **Conclusion**

This examination of the effects of student-fashioning and teacher-pleasing features in an FYW assessment of reflective essays has four possible limitations. First, this study was conducted in a single site, so the results have no institutional comparisons at this time. Second, this study examined only two of our four writing outcomes in FYW, reflection and writing, so the results are partial across our outcomes. Third, this study was conducted in the context of a highly structured FYW composition program: We use a common syllabus with a well-articulated assignment sequence, which allows advanced quantitative analyses because all students are writing in the context of the same outcomes and the same assignment. Fourth, we have not yet broadened our research to consider issues of diversity and inclusion in FYW assessment. We hope to address these four limitations in future research. First, we are pursuing opportunities to work collaboratively with other composition programs interested in developing mixed methods approaches in FYW assessment. Such collaborative research would provide more high-quality evidence for the validity of using thin-slice methods in FYW assessment. Second, we plan to develop outcomes and rubrics for assessing our other FYW outcomes, reading and researching. As our research progresses, we hope to bring our mixed methods approach to other courses offered by our and other composition programs, from dual enrollment and teacher training to graduate seminars and qualifying examinations. As in the studies presented here, we will use our mixed methods by conducting rich feature analyses to discover the ways students talk about these outcomes in their reflective essays in order to design mixed methods assessment studies. Third, we feel the need to know more about how composition programs in general regard standardization in their curriculum and/or pedagogy. This is a

sensitive issue in the field because some researchers and programs prize variation (e.g., expressive pedagogy), but little is known about how and why composition programs utilize standardization in their FYW curriculum, pedagogy, and assessment. We plan to launch a nationwide survey on this issue because more and more writing studies researchers are interested in designing mixed methods for assessment, with mixed methods defined as quantitative and qualitative research methods aimed at validity, reliability, and change based on a thorough understanding of the writing of their entire FYW student body.

Fourth and finally, we have just begun to incorporate social justice in our mixed methods approach to FYW assessment. Sarah Primeau (2020) designed an inclusive FYW assessment, seeking the perspectives of faculty and students to learn more about our “habits of judgement” (Inoue, 2019). To avoid the inevitable hierarchical dominance of one assessment team with both faculty and students, Primeau conducted two small-scale assessment teams at different times, conducting one assessment for faculty and another assessment for students. Both teams received the same rubric and an outcome-oriented norming before scoring a set of FYW reflection essays independently. The student team also participated in a discussion of their experience and their recommendations for future FYW courses and assessments. In their discussion, students suggested an intriguing curricular intervention, starting with the observation that most FYW reflective examples are typically from highly rated essays. They then proposed a norming and scoring classroom exercise with multiple examples of essays across the entire rubric, which, they argued, would help FYW students understand teachers’ focus on specific learning outcomes and their complex expectations for assessment.

## **Appendix A**

### **FYW Learning Outcomes**

#### **Reading**

Use reading strategies in order to identify, analyze, evaluate, and respond to arguments, rhetorical elements, and genre conventions in college-level texts and other media.

#### **Writing**

Compose persuasive academic genres, including argument and analysis, using rhetorical and genre awareness.

Use a flexible writing process that includes brainstorming/inventing ideas, planning, drafting, giving and receiving feedback, revising, editing, and publishing.

#### **Researching**

Use a flexible research process to find, evaluate, and use information from secondary sources to support and formulate new ideas and arguments.

#### **Reflecting**

Use written reflection to plan, monitor, and evaluate one's own learning and writing.

## **Appendix B**

### **FYW Reflective Essay Assignment (ENG 1020)**

**Introduction/Rationale.** The Reflective Essay is a 1000-1250-word (4-5 pages) essay in which you make a case for how well you have achieved the goals of the course. To do so, you must look back over the work you produced during the semester in order to find, cite, and discuss evidence of achievement in each of the four learning outcome categories (reading, writing, research, and reflection). It is critical that your Reflective Essay includes concrete examples and discussion of what you have been able to do as a result of your work in the course.

While your discussion of achievements with respect to ENG 1020 learning outcomes is perhaps the most important goal in the Reflective Essay, the written expression of these achievements can be strengthened when it is integrated into a broader narrative that describes where you are coming from and who you are as a student. In this narrative, you may discuss, for example, how you learned and used various reading strategies in the course, or you may describe, for example, how your ability to perform effective research increased.

In sum, the Reflective Essay should make claims about your success with respect to ENG 1020 learning outcomes and support these claims with compelling evidence of achievement in order to demonstrate what you have learned and what you can do as a result of your work in the course. In this way, a successful Reflective Essay will inspire confidence that you are prepared to move forward into your next composition courses and into the larger academic discourse community.

**Assignment Prompt.** In this assignment, you will evaluate your growth as an English 1020 student, using your choice of experiences and work on the projects to support your claims. In an essay of 4-5 pages, make an argument that analyzes your work in ENG 1020 in relationship to the course learning outcomes listed on the syllabus for the course. Explain what you have achieved for the learning outcomes by citing specific passages from your essays and other assigned writings for the course, and by explaining how those passages demonstrate the outcomes. Also, consider describing the process you used to complete this work and any background information about yourself, as listed above, that might help us better understand the work you did this semester in working toward the course learning outcomes.

You will want to choose the learning outcomes and knowledge that have developed most strongly and importantly for you. If you think there is little evidence of your growth in a particular learning outcome, no problem: just articulate why in your final essay. You should address all of the learning outcomes, but you may choose which ones you focus on. Your main claim (or thesis statement) should identify specific characteristics that you believe your experiences and work in English 1020 (which you'll use your body paragraphs to talk about) represent. The body paragraphs of your essay should develop your main claim with evidence from your major works and experiences in this course. As you choose evidence and sub-claims to make about your major assignments, you will develop your paragraphs by drawing upon the process of completing the assignment to support the claim.

In a nutshell, this assignment asks you to take a critical look at your work from this semester and talk about it in terms of your knowledge of yourself as a learner and thinker.

**Appendix C**  
**FYW Reflection Rubric**

Assessment learning outcome: Use written reflection to evaluate one's own learning and writing.						
	6	5	4	3	2	1
	Excellent	Good		Limited		None
Argument (thesis, claim, relation to course outcomes).	6	5	4	3	2	1
Evidence (examples, analysis, experiences, discussion).	6	5	4	3	2	1

## Appendix D

### FYW Writing Rubric

Assessment writing outcome: Use conventions of academic discourse.						
	6	5	4	3	2	1
	Excellent	Good		Limited		None
Paragraph Organization (topic sentence, paragraph development)	6	5	4	3	2	1
Editing (grammar, punctuation, spelling, awkward sentences)	6	5	4	3	2	1

## References

- Adler-Kassner, L., & O'Neill, P. (2010). *Reframing writing assessment to improve teaching and learning*. Utah State University Press.
- Ambady, N., Bernieri, F., & Richeson, J. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In M. Zanna (Ed.), *Advances in experimental social psychology*, 32, 201-271.
- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, 64(3), 431-441.
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, 48(1), 12-19.
- Barton, E. (2004). Linguistic discourse analysis: How the language in texts works. In C. Bazerman & P. Prior (Eds.), *What writing does and how it does it* (pp. 57-82). Routledge.
- Barton, E., Pruchnic, J., Boeder, R., Grogan, J., Primeau, S., & Foster Moore, T. (in press). Thin-slice methods and contextualized norming: Innovative methodologies for austere times. In C. Donohye, Blewett, K., & Monroe, C. (Eds.), *The expanding universe of writing studies: Higher education writing research*. Peter Lang.
- Bouelle, T., Rankins-Robert, S., Bouelle, A., & Roen, D. (2013). Assessing learning in redesigned online first-year composition courses. In H. McKee & D. DeVoss (Eds.), *Digital writing: Assessment and evaluation*. Computers and Composition Digital Press. [https://ccdigitalpress.org/book/dwae/12\\_bouelle.html](https://ccdigitalpress.org/book/dwae/12_bouelle.html)
- Broad, B. (2012). Strategies and passions in empirical research. In L. Nickoson & M. Sheridan (Eds.), *Writing studies research in practice: Methods and methodologies* (pp. 197-209). Southern Illinois University Press.
- Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290.
- Colombini, C., & McBride, M. (2012). "Storming and norming": Exploring the value of group development models in addressing conflict in communal writing assessment. *Assessing Writing*, 17(4), 191-207.
- Comer, D., & White, E. (2016). Adventuring into MOOC writing assessment: Challenges, results, and possibilities. *College Composition and Communication*, 67(3), 318-359.
- Condon, W., & Kelly-Riley, D. (2004.) Assessing and teaching what we value: The relationship between college-level critical thinking and writing abilities. *Assessing Writing*, 9(1), 56-75.
- Conway, G. (1994). Portfolio cover letters, students' self-presentation, and teachers' ethics. In D. Daiker, J. Sommers, & Gail Stygall (Eds.), *New directions in portfolio assessment* (pp. 83-92). Boynton/Cook.

Council of Writing Program Administrators. (2014). *WPA outcomes statement for first-year Composition (3.0)*.

Creswell, J., & Creswell, D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage.

Elbow, P., & Belanoff, P. (1986). Using portfolios to increase collaboration and community in a writing program. *WPA: Writing Program Administration*, 9(3), 27-40.

Elliot, N., Rudniy, A., Deess, P., Klobucar, A., Collins, R., & Sava, S. (2016.) ePortfolios: Foundational measurement issues. *Journal of Writing Assessment*, 9(1).  
[journalofwritingassessment.org/article.php?article=110](http://journalofwritingassessment.org/article.php?article=110)

Emmons, K. (2003). Rethinking genres of reflection: Student portfolio cover letters and the narrative of progress. *Composition Studies*, 31(1), 43-62.

Faigley, L. (1989). Judging writing, judging selves. *College Composition and Communication*, 40(4), 395-412.

Gladwell, M. (2005). *Blink: The power of thinking without thinking*. Little, Brown.

Grant-Davie, K. (1992). Coding data: Issues of validity, reliability, and interpretation. In G. Kirsch & P. Sullivan (Eds.), *Methods and methodology in composition research* (pp. 248-270). Southern Illinois University Press.

Haas, C., Takayoshi, P., & Carr, B. (2012). Analytic strategies, competent inquiries, and methodological tensions in the study of writing. In L. Nickoson & M. Sheridan (Eds.), *Writing studies research in practice: Methods and methodologies* (pp. 51-62). Southern Illinois University Press.

Haefner, J. (2010). A practical, pithy guide to quantitative scoring assessment as a SLAC. [https://sun.iwu.edu/~jhaefner/practical\\_pithy\\_guide.pdf](https://sun.iwu.edu/~jhaefner/practical_pithy_guide.pdf)

Hallgren, K. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23-34.

Haswell, R. (Ed.). (2001). *Beyond outcomes: Assessment and instruction within a university writing program*. Greenwood.

Haswell, R. (2012). Quantitative methods in composition studies: An introduction to their functionality. In L. Nickoson & M. Sheridan (Eds.), *Writing studies research in practice: Methods and methodologies* (pp. 185-196). Southern Illinois University Press.

Huckin, T. (1992). Context-sensitive test analysis. In G. Kirsch & P. Sullivan (Eds.), *Methods and methodology in composition research* (pp. 84-104). Southern Illinois University Press.

Inoue, A. (2019). Classroom writing assessment as an antiracist practice: Confronting White supremacy in the judgments of language. *Pedagogy*, 19(3), 373-404.

Jankens, A., & Trimble, T. (2019). Using taxonomies of metacognitive behaviors to analyze student reflection and improve teaching practice. *Pedagogy*, 19(3), 433-454.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus, and Giroux.

- Kulik, J. (2001). Student ratings: Validity, utility, and controversy. *New Directions for Institutional Research*, 2001(109), 9-25.
- Lindenman, H., Camper, M., Jacoby, L., & Enoch, J. (2018). Revision and reflection: A study of (dis)connections between writing knowledge and writing practice. *College Composition and Communication*, 69(4), 581-611.
- Miura, C. (2018, November 11). Self-fashioning and the performance of identity in first-year writing [Paper presentation]. The 116th Annual Conference of the Pacific Ancient and Modern Language Association, Bellingham, WA. <http://www.pamla.org/2018/proposals/self-fashioning-and-performance-identity-first-year-writing>
- Moon, J. (2004). *A handbook of reflective and experiential learning: Theory and practice*. Routledge.
- Pagano, N., Bernhardt, S., Reynolds, D., Williams, M., & McCurrie, M. (2008). An inter-institutional model for college writing assessment. *College Composition and Communication*, 60(2), 285-320.
- Primeau, S. (2020). Developing and assessing a cultural rhetorics oriented first-year composition course [Doctoral dissertation, Wayne State University]. [https://digitalcommons.wayne.edu/oa\\_dissertations/2422/](https://digitalcommons.wayne.edu/oa_dissertations/2422/)
- Pruchnic, J., Susak, C., Grogan, J., Primeau, S., Torok, J., Trimble, T., Foster, T., & Barton, E. (2018). Slouching toward sustainability: Mixed methods in the direct assessment of student writing. *Journal of Writing Assessment*, 11(1).
- Pruchnic, J., et al. (2021). Correlating what we know: A mixed methods study of reflection and writing in first-year composition [Manuscript submitted for publication].
- Schendel, E., & O'Neill, P. (1999). Exploring the theories and consequences of self-assessment through ethical inquiry. *Assessing Writing*, 6(2), 199-227.
- Sommers, J. (2011). Reflection revisited: The class collage. *Journal of Basic Writing*, 30(1), 99-129.
- Sommers, J., Black, L., Daiker, D., & Stygall, G. (1993). The challenge of rating portfolios: What WPAs can expect. *WPA: Writing Program Administration*, 17(1-2), 7-29.
- Taczak, K. (2016). Reflection is critical for writers' development. In L. Adler-Kassner & E. Wardle (Eds.), *Naming what we know: Threshold concepts of writing studies* (pp. 78-81). Utah State University Press.
- Walvoord, B. (2010). *Assessment clear and simple: A practical guide for institutions, departments, and general education* (2nd ed.). Jossey-Bass.
- Weiser, I. (1997). Revising our practices: How portfolios help teachers learn. In K. Yancey & I. Weiser (Eds.), *Situating portfolios: Four perspectives* (pp. 293-301). University Press of Colorado and Utah State University Press.
- White, E. (2005). The scoring of writing portfolios: Phase 2. *College Composition and Communication*, 56(4), 581-600.

White, E., Elliot, N., & Peckham, I. (2015). *Very like a whale: The assessment of writing programs*. University Press of Colorado and Utah State University.

Yancey, K. (1996). Dialogue, interplay, and discovery: Mapping the role and the rhetoric of reflection in portfolio assessment. In R. Calfee & P. Perfumo (Eds.), *Writing portfolios in the classroom: Policy and practice, promise and peril* (pp. 83-102). Routledge.

Yancey, K. (1998a). *Reflection in the writing classroom*. Utah State University.

Yancey, K. (1998b). Getting beyond exhaustion: Reflection, self-assessment, and learning. *The Clearing House: A Journal of Educational Strategies, Issues, and Ideas*, 72(1), 13-17.

Zubizarreta, J. (Ed.). (2009). *The learning portfolio: Reflective practice for improving student learning* (2nd ed.). Jossey-Bass.