

## Grading Contracts: Assessing Their Effectiveness on Different Racial Formations

Asao B. Inoue

I've used grading contracts for several years now. I initially chose to adopt contracts because they solve at least three problems I have with grading writing. One, grades are deceptive; not only do they replace real feedback on student writing with a one-dimensional, somewhat arbitrary symbol, but that symbol often is perceived by the student to stand in for how well he or she is doing. Two, grades create false hierarchies that are counterproductive to a collaborative and educative learning environment, making some students feel bad about themselves as writers when they should not and prematurely halting revision in other students (Bleich, 1997; Elbow, 1993). Three, grades tend to create a need in students for more grades, often at the expense of formative and more authentic response (Elbow, 1993; Kohn, 1993). The presence and expectation of grades tend to construct an ill-fitting kind of motivation for the writing classroom, one based on extrinsic rewards that keep students from learning. I prefer to encourage intrinsic rewards for writing.

These grade-related problems are not new. They and other issues have been identified in the writing assessment literature. Beyond the literature on portfolios (Belanoff & Dickson, 1991; Black et al., 1994), which often discusses the delaying of grades (Hamp-Lyons & Condon, 2000, p. 34), Huot, for instance, advocates for a distinction from grading in our reading and evaluation practices, calling for "instructive evaluation" (2002, p. 69). Similarly, Bleich (1997) argues against grading from an historical perspective and for "descriptive evaluation" (pp. 29-31). And, of course, there's the literature that questions the reliability of grades. In one famous study, 300 papers were given to 53 judges (roughly half were from academia and half held non-academic positions, such as editors, lawyers, and writers). The researchers found very low reliability—a median correlation of .31—in the grades given to the papers (Diederich, 1974, p. 6). Bowman (1973) found similar results on assessment in business writing courses, with the same piece of writing receiving a wide range of grades from senior and junior teachers (pp. 28-29). (See also Dulek & Shelby, 1981).

In a summary of the research on grading, O'Hagan explains that while grading emerged in the US around 1850, "studies as early as 1912 questioned

the validity of grading, suggesting that in writing instruction [...] grades were far too subjective” (1997, p. 4). While O’Hagan claims the subjectivity of grading is an issue of validity, I would say it falls under the category of reliability, or the consistency of grades. The common complaint that grades are too “subjective” to be useful seems equally an issue of reliability or consistency as it does in validity. Regardless, classroom grading practices have had many critics, and I consider myself one of them.

In attempting to solve these grading problems, I moved to a grading contract, but after several years, I wondered: how effective has my grading contract been for my students? Does it work better for some students than others? Like Kelly-Riley in this collection, I wanted to know how various racial formations were faring on our assessments. In this chapter, I assess the effectiveness of grading contracts on three racial formations, Asian Pacific Islanders (APIs), African Americans, and whites, in Fresno State’s First Year Writing (FYW) program.

I ask about grading contract effectiveness not just to understand its impact on various racial formations in classrooms but to illustrate the way any assessment technology may have differential effects on various students. As Hanson (1993) has argued convincingly about testing generally, grading technologies are never neutral systems. Like any other “test,” grading produces the very traits, characteristics, and categories it claims to measure in students and their writing (Hanson, pp. 284, 287–88). So in part, I wonder what biases my grading contract has. Furthermore, grading contracts, like all grading technologies, make obvious that we form agreements with students to produce evaluations of their writing and course grades, agreements that have governing criteria, criteria that may have biases unknown to teachers (or students), and potentially differential effectiveness for various students. In short, the effectiveness of our grading technologies, whatever they may be, may not be evenly distributed among all students.

### **Elements Of Effectiveness In Grading Contracts**

In many ways, it is unfair to measure the effectiveness of grading contracts against conventional grading systems. Contracts are so different to teachers and students that their “effectiveness” may look quite different from conventional grading technologies. However, according to the literature on contracts in composition, contracts do possess at least three features of effectiveness. In 1973, Mandel offered a contract system in which the more written projects a student completed, the higher the course grade that student received. Mandel explained that his contract allowed him not to grade student writing through judgments of quality, which he found detrimental to student learning (p. 623);

instead, he graded on quantity (p. 628). A few years later, Knapp (1976) discussed a similar contract system, one also based on quantity. Papers were given a simple binary judgment: acceptable or unacceptable. If unacceptable, the student took her feedback and revised for submission the next week. Each successive paper demanded more of the student, and the more papers the student completed, the higher her grade in the course. Knapp distinguishes an important feature not present in Mandel's contract: every essay gets a binary decision, which is unlike Mandel's system in which "faith in students" to achieve and challenge themselves drives drafting and writing efforts (Mandel, 1973, p. 629). Despite this important difference in how drafts are judged, both share one important feature in their contracts. Grades are calculated primarily by the quantity of work produced, which I argue is crucial to how effective both contracts can be. And so, grading contract effectiveness can be measured in part by the quantity of work produced by students.

Of course a focus on quantity is not paramount to disregarding quality. Mandel lets quality organically spring from the cycle of drafting and conferences, while Knapp uses that binary judgment, a judgment mainly about quality, to ensure that students are ready to move to the next writing assignment. Additionally, many who believe grading contracts are a better choice often voice the opinion that the more a student practices writing, the better she will get, arguing that quality is a function of quantity.

The focus on quantity of work over quality of work to determine grades is a hallmark of most grading contract systems. Bauman's grading contract (1997, pp. 164–65) is a good example. Bauman's main concern, however, is with the contract's ability to motivate students in the right ways. She argues that grades motivate students toward the wrong ends in a writing class, toward extrinsic rewards (p. 166), something that Danielewicz and Elbow mention about Elbow's contract (2009, p. 247). Drawing on Paris and Turner (1994), Bauman explains that motivation is not simply a "characteristic of people or a property of events"; it is "derived from contextual transactions" (Bauman, 1997, p. 167). People and their contexts create motivation (p. 167). Grading contracts' emphasis on quantity (and not grades), Bauman argues, creates an environment for such intrinsic motivation. Motivation as a measure of effectiveness of contracts, however, is difficult to assess. Motivation may show up in a variety ways, but most agree that it is a reaction to the learning environment.

Spidell and Thelin's (2006) study of student resistance to grading contracts offers some insight on motivation. While Spidell and Thelin's methods did not allow them to separate data by racial formation, they found several general trends in student surveys and interviews about contracts in their mostly white student population: (1) many students resented contracts because

they were used to working in point systems (p. 40), which quantified efforts (p. 41); (2) a “perception of increased responsibility led to anxiety and resistance” to contracts (p. 42)—i.e., it felt like too much work; (3) some students felt that the contract leveled the grading curve too much (p. 44); (4) “students felt the contract made the course more difficult than necessary” (p. 48); and (5) while some students found that the contract made expectations clearer and motivated them to do work, others felt it may have offered clear objectives for grades but did not motivate students to write better (p. 50). Spidell and Thelin’s study helps us see a third feature of the effectiveness of any grading technology: effectiveness can be measured by students’ reactions to the grading contract.

In sum, the effectiveness of grading contracts, if judged on their own terms, has three features, which I’ll use in this study. Effectiveness is a measure of (1) the quantity of work produced, (2) the quality of writing produced in class, and (3) student reactions to and acceptance of the contract itself. I’d like to note two things about the literature on grading contracts reviewed above. First, there is no discussion of differential impact or effectiveness of contracts on different racial formations. Second, there is no quantitative research published on the effectiveness of contracts. This chapter attempts to fill both of these gaps.

### Fresno State’s Grading Contract

Fresno State’s FYW program uses a grading contract that is adopted from Shor (1996) and Danielewicz and Elbow (2009). These versions of grading contracts were chosen because they seemed to offer the best chances for our diverse populations to succeed as writers. Shor’s and Danielewicz and Elbow’s contracts both use the main elements of contracts mentioned above. In addition to focusing on quantity of work to produce course grades and cultivating intrinsic motivation for doing work in the class by grading less or not at all in the semester, Shor’s contract is negotiated with students, and attempts to do the things that Shiffman is looking for in a grading system: share power and redistribute authority self-consciously (Shor, 1996, p. 20; Danielewicz and Elbow, 2009, p. 245).

Danielewicz and Elbow offer an instructive way to compare the similarities between Shor’s contract and theirs. They explain that Shor’s contract’s emphasis on resisting the “culture of capitalism” is similar to their contract’s focus on resisting a “culture of grading and assessment” (Danielewicz & Elbow, 2009, p. 248). The two cultures function similarly on the individual: conventional grading “helps induce student compliance by obscuring analogous structures of unfairness” (p. 248). Finally, Danielewicz and Elbow explain that their

contract encourages extrinsic motivation, in the form of doing a certain amount of work to get a course grade of “B,” but this extrinsic motivation leads to intrinsic motivation (p. 257). They argue that contracts motivate students in the opposite direction from what Bauman promotes, but intrinsic motivation is the goal, just as Bauman argues. Thus by doing the work, students become more interested and engaged. Practice makes for engagement and interest, for motivation, and thus better writing. The contract’s direct critique of conventional grading’s tendency to obscure “structures of unfairness” seemed a perfect solution for the difficulties that two particular racial formations—Asian Pacific Islanders (who are mostly Hmong), and African Americans—historically have had in the FYW program.

In a nutshell, all contracts in Fresno State’s FYW program are negotiated in the first week or two of classes. Our contracts justify grades, typically a “B,” by the amount of work done, with little attention to the quality of writing, except in the crudest sense of judging whether basic requirements are met, which is akin to Knapp’s binary distinction (acceptable or not). Our contract language describes meeting the contract’s expectations as turning in writing “in the manner and spirit it is asked” of students. The contract also attempts to limit the teacher’s power over student writing and revising by limiting the range and potency of judgments possible that affect student course grades. The logic is that fewer teacher judgments of quality and fewer distinctions of quality in writing will allow students to have opportunities to make and articulate decisions as writers, even ones teachers do not agree with, then discuss those decisions in portfolio reflection letters. Thus, all assignments are typically acceptable or not acceptable, meaning they either have met basic assignment requirements or they have not (e.g., due time and date, word count, addressing particular texts or questions, etc.).

### Methods And Data

To address the effectiveness of grading contracts, I gathered data from three sources: (1) anonymous exit surveys of FYW students in the writing program; (2) final portfolio ratings from the same group of students; and (3) course grade distributions of the same students. All three data sources come from English 5B students from the Spring 2009 and Spring 2010 semesters in Fresno State’s FYW program. English 5B is the second and final course in the stretch option, one option students may choose to fulfill their university writing requirement. A student’s 5A and 5B teachers are typically the same instructor, and 5A uses a mandatory grading contract and is a credit/no credit course. Additionally, the FYW program is a Directed Self-Placement (DSP) program similar to the one discussed by Royer and Gilles (1998; 2003) at

Grand Valley State University. Roughly 69% of all incoming first-year students chose the stretch option in the 2008–09 AY, while 56.5% choose it in 2009–10 AY. I’ve written more about Fresno State’s DSP, stretch courses, curricula, program portfolio, students, and our program assessment efforts elsewhere (Inoue, 2009a).

### Surveys

The anonymous surveys were conducted online, and asked students to answer the following questions about grading contracts in their classes: (1) “How effective overall did you find the grading contract to be in your writing class?” (2) “How happy are you with the grading contract as a student in the class?” and (3) “Do you prefer a grading contract over traditional grading systems (where grades are placed on each assignment) in courses like this one?” Additionally, two open-ended responses were solicited: (1) “Please explain below what you like about the grading contract in courses like writing courses, or how the contract helped you as a student or writer,” and (2) “Please explain below what you do not like about grading contracts in courses like writing courses.” I later grouped similar open-ended responses into themes. This allowed me to both quantify the kinds of open-ended responses received and consider qualitatively their content.

As seen in Table 1, when student enrollment statistics for Fall 2008 and Fall 2009 (only fall enrollment is published by the university) are compared to students who completed the spring 5B exit surveys, students of color are somewhat overrepresented in the survey. In both years, white students are underrepresented when compared to overall university enrollment figures.

	Spring 2009 survey		Spring 2010 survey	
	N	%	N	%
African American	27	6.25%	43	8.69%
Native American	1	0.23%	3	0.61%
Asian Pacific Is.	100	23.15%	126	25.45%
Hispanic	186	43.06%	209	42.22%
White	118	27.31%	113	22.83%
Total	432	100.0%	494	100.0%

Table 1. FYW Exit Survey Participants By Race

### Portfolio Ratings

The program portfolio consists of 10 pages of polished work (often 2–3 pieces), usually coming from the main projects dictated by the program’s curriculum, one previous draft of each document included, and a letter of reflection (not counted in the 10 pages). Portfolio ratings are generated by blind readings conducted in the summer by teachers in the program. The only portfolio ratings we collect for the stretch program are from the English 5A midterm and English 5B final portfolios; thus these are the ratings reported in this study, constituting a semester and half of instruction using a grading contract. Each portfolio is rated on a linear scale from 1 to 6 on five of the program’s eight outcomes (with 1–2 being inadequate quality, 3–4 being proficient quality, and 5–6 superior quality):

- **READING/WRITING STRATEGIES:** Demonstrate or articulate an understanding of reading strategies and assumptions that guide effective reading, and how to read actively, purposefully, and rhetorically;
- **REFLECTION:** Make meaningful generalizations/reflections about reading and writing practices and processes;
- **SUMMARY/CONVERSATION:** Demonstrate summarizing purposefully, integrate “they say” into writing effectively or self-consciously, appropriately incorporate quotes into writing (punctuation, attributions, relevance), and discuss and use texts as “conversations” (writing, then, demonstrates entering a conversation);
- **RHETORICALITY:** Articulate or demonstrate an awareness of the rhetorical features of texts, such as purpose, audience, context, rhetorical appeals, and elements, and write rhetorically, discussing similar features in texts;
- **LANGUAGE COHERENCE:** Have developed, unified, and coherent paragraphs and sentences that have clarity and some variety. (California State Fresno, 2009, n.p.)

### Grade Distributions

The grade distributions were gathered and processed by the university’s Office of Institutional Research, Assessment, and Planning, the same office that completed the statistical work on the portfolio ratings and enrollments.<sup>1</sup>

### Participants

Asian Pacific Islanders (API) at Fresno State consist mostly of Hmong students, of which about half speak Hmong at home and half English.<sup>2</sup> In this



same formation, just less than 50% identify their parents as having “less than a high school education.” Approximately 26% reported that their parents have a high school diploma or GED. Fewer than 15% reported that their parents have some college or an Associates degree, and about 5% say their parents have a Bachelor’s degree. In Fall 2009, 86.1% of all regularly admitted, first-year API students were designated as needing remediation.<sup>3</sup>

African American students in the FYW program report that they speak English at home (97.78% of the time), with Spanish as the only other language spoken at home (2.22% of the time). Their parents are generally more educated than the Hmong population, identifying 6.82% with less than a high school education, 27.27% with a high school diploma, 34.09% with an associates degree, 22.73% with a Bachelor’s, and 9.09% with a Master’s. In Fall 2009, 78.3% of all regularly admitted, first-year African Americans were designated as needing remediation by the university, a marginally better rate than APIs.

Finally, white students primarily speak English at home (92.11% of the time), with Spanish (2.63%) and other languages (5.26%) also spoken. Most of their parents have more than a high school diploma (24.3%), with 28.04% identifying an Associates, 21.50% a Bachelor’s, 17.76% a Master’s, and 4.67% a PhD, EdD, or MD degree as the highest level of parental education. This racial formation comes from the most educated households, and has the lowest remediation rates in Fall 2009, at 59.8% needing remediation.

### **Results: Effectiveness As Quantity, Student Acceptance, And Quality**

Effectiveness of grading contracts in English 5B is measured in this study by considering three data sources: (1) exit surveys, which help determine student acceptance and response to the contract; (2) English 5A midterm and English 5B final portfolio ratings, which help determine quality of writing and development or growth in writing<sup>4</sup>; and (3) English 5B grade distributions, which help determine the quantity of work done, since all contracts are for “B” grades.

#### **Asian Pacific Islanders (API)**

In English 5A midterm and 5B final portfolios, API students performed better than the mean of all students in every category in both years (Tables 2 and 3). They did not have the biggest difference in ratings, but like all students, they moved from just above “poor” (overall average 2.97 and 2.52) to “acceptable” quality (overall average 3.68 and 3.54) in most dimensions in both years. Interestingly, the one dimension in which Hmong students did receive statistically significantly higher average ratings than all other racial



formations on the final portfolio in both years is a critical program outcome, “summary and conversation” (difference of .81 and 1.25). Focusing on quantity of work done may be helping API students engage better with academic texts and incorporate them appropriately in their own writing. More writing seems to equate to better summary and engagement with texts in their writing. API students also met the contract’s workload obligations at about the same level as all students, 79.1% receiving “As” and “Bs” in 2009 and 80.1% receiving the same grades in 2010; however, in the year that fewer of all categories of students met contract obligations (2010), more API students met their obligations. Whatever changes occurred in classrooms in 2010, they did not affect API students negatively. In fact, those changes seemed to help them. The quantity of work our contracts demanded could have been a factor in APIs’ abilities to engage appropriately with academic texts in their writing and show significant improvement along this dimension in both years.

Additionally, APIs were the most consistently accepting of the contract (Table 4). APIs found the grading contract about as effective as all students in both years (81.48% and 77.78%), and were “happy” or “very happy” with the contract in about the same rates as all students (73.75% and 71.56%). However, it is their preference for the contract, higher than any other formation in both years (81.48% and 70.64%), that stands out. While APIs’ preference rates dropped at a similar rate as all other groups in 2010, they did not seem to match how many API students met their contracts, since more met their contracts in 2010 than 2009. So the changes that occurred in 2010 did not adversely affect APIs’ ability to meet their contract obligations (the quantity of work).

In their open-ended responses on surveys, APIs had mostly positive comments, with 129 positive and 116 negative comments entered. They had the most positive comments in three related themes, “expectations and clarity,” “motivation and staying on track,” and the most interesting one, “freedom to write without grades.” There were 34 comments made concerning “expectations and clarity.” Similar to Spidell and Thelin’s (2006) findings, most respondents explained that the contract allowed them to know the expectations for a grade in the class: “I like how the grading contract gives us certain days to miss a class and how many assignments we can miss to receive a certain grade.” There were 22 “motivation and staying on track” comments, which praised the contract for keeping them on task throughout the semester: “this grading contract helped me to be on task about doing my homework and essays.”

**TABLE 2. FALL 2008 5A MIDTERM AND SPRING 2009 5B FINAL PORTFOLIO MEAN RATINGS (1-6 SCALE)**

\* Results are based on Paired Samples T Tests (N<=20) or Wilcoxon Signed Rank Test (N<20) with significance level 0.05. Significant differences are in bold.

	All Students (n=130)		Asian Pacific Islander (n=21)		African American (n=14)		Latino/a (n=52)		White (n=32)	
	5A mid	5B fin	5A mid	5B fin	5A mid	5B fin	5A mid	5B fin	5A mid	5B fin
overall	2.88	3.53	2.97	3.68	2.83	3.73	2.75	3.51	3.01	3.41
strategies	2.62	3.35	2.81	3.48	2.64	3.64	2.48	3.37	2.63	3.16
reflection	2.73	3.54	2.90	3.81	2.71	3.79	2.56	3.48	2.81	3.44
summary	2.98	3.61	2.90	3.71	2.88	3.64	2.71	3.56	3.19	3.47
rhetoric	2.98	3.50	3.14	3.57	2.93	3.57	2.85	3.50	3.16	3.44
lang.co	3.08	3.67	3.10	3.81	2.93	4.00	3.00	3.65	3.25	3.53
			<b>diff.*</b>		<b>diff.*</b>		<b>diff.*</b>		<b>diff.*</b>	
			.65		.70		.90		.76	
			.72		.67		1.00		.88	
			.81		.90		1.07		.92	
			.62		.81		.71		.67	
			.52		.43		.64		.65	
			.59		.71		1.07		.65	
										<b>diff.**</b>
										.40
										.53
										.63
										.28
										.48

\*\* Regular ANOVA is used to test if there are significant differences in students' improvement among racial groups. The reference group is White students. The level of significance is 0.05. Significant differences are in bold.

**TABLE 3. FALL 2009 5A MIDTERM AND SPRING 2010 5B FINAL PORTFOLIO MEAN RATINGS (1-6 SCALE)**

\*\* Regular ANOVA is used to test if there are significant differences in students' improvement among racial groups. The reference group is White students. The level of significance is 0.05. Significant differences are in bold.

	All Students (n=156)		Asian Pacific Islander (n=36)		African American (n=15)		Latino/a (n=60)		White (n=31)	
	5A mid	5B fin	5A mid	5B fin	5A mid	5B fin	5A mid	5B fin	5A mid	5B fin
overall	2.47	3.49	2.52	3.54	2.28	3.27	2.39	3.42	2.67	3.63
strategies	2.24	3.36	2.36	3.47	1.80	2.87	2.20	3.30	2.39	3.55
reflection	2.22	3.44	2.47	3.47	2.00	2.93	2.08	3.42	2.32	3.65
summary	2.50	3.54	2.53	3.78	2.33	3.67	2.40	3.35	2.81	3.48
rhetoric	2.21	3.38	2.28	3.36	2.20	3.20	2.13	3.37	2.35	3.52
lang.co	3.17	3.72	2.94	3.61	3.07	3.67	3.12	3.65	3.48	3.97
			<b>diff.**</b>		<b>diff.**</b>		<b>diff.**</b>		<b>diff.**</b>	
			1.02		1.02		.99		1.03	
			1.12		1.11		1.07		1.10	
			1.22		1.00		.93		1.33	
			1.04		1.25		1.33		.95	
			1.17		1.08		1.00		1.23	
			.55		.67		.60		.53	
										.96
										1.16
										1.32
										.68
										1.16
										.48

\*\* Regular ANOVA is used to test if there are significant differences in students' improvement among racial groups. The reference group is White students. The level of significance is 0.05. Significant differences are in bold.

**TABLE 4. ENGL 5B EXIT SURVEY ON GRADING CONTRACTS**

	All Students (n=343)		Asian Pacific Islander (n=81)		African American (n=22)		Latino/a (n=140)		White (n=90)	
	Sp 09	Sp 10	Sp 09	Sp 10	Sp 09	Sp 10	Sp 09	Sp 10	Sp 09	Sp 10
effectiveness	79.24%	77.83%	81.48%	77.78%	72.73%	92.50%	81.43%	76.44%	74.75%	76.67%
happiness	72.51%	71.88%	73.75%	71.56%	57.14%	75.00%	75.54%	72.51%	70.30%	70.97%
preference	73.76%	63.46%	81.48%	70.64%	72.73%	62.50%	73.19%	63.01%	68.32%	58.24%

Finally, there were 22 comments under the “freedom to write without grades” theme, which, unlike Spidell and Thelin’s (2006) findings, tended to express appreciation and praise for the contract’s ability to keep grades off their writing and assignments. The comments in this theme tended to be longer than those in any other theme, and longer than comments by any other racial formation. Typical comments were as follows:

It allows me to write and not be afraid that my thoughts or perspective will be marked down if I do not meet the writing standards. The contract helps me concentrate more on my ideas and purpose rather than on grammar errors. It makes me feel like a real and independent writer.

For API students, it seems clear: the emphasis on quantity instead of quality gave them confidence and motivation to write, which they preferred. Contracts produced quality writing that was deemed proficient in their portfolios, as well as marked differences (growth) along all dimensions measured in portfolios, most significantly the key program outcome, “summary and conversation.” And since most APIs met their contract obligations and found that the contracts freed them to write without grades, the focus on quantity over quality in order to produce intrinsic motivation to write and write better appears to be most effective for APIs. In these ways, grading contracts were most effective for APIs at Fresno State.

### **African Americans**

In 2009, African Americans’ English 5A midterm ratings on most dimensions (overall rating 2.83) were similar to the mean ratings of all students (overall rating 2.88), but their English 5B final ratings were among the highest (overall rating 3.73) (Table 2). Meanwhile in 2010 portfolios, they received some of the lowest mean ratings in many dimensions of all groups at both English 5A midterm (overall rating 2.28) and English 5B final (overall rating 3.27) (Table 3). Still, African Americans in both years moved from generally poor quality to proficient quality, and like their Hmong counterparts in 2009, African Americans in 2010 showed the most growth among all formations along “summary and conversation,” (statistically significant growth in 2010, with a difference of 1.33). In fact, in 5B final portfolios, both racial formations were rated higher in this dimension than any other racial formation (API at 3.78 and African Americans at 3.67).

Interestingly, African Americans’ grade distributions achieve fewer “As” and “Bs” than any other group in both years, suggesting that they had trouble meeting the quantity required by the contract and getting the work done. Only 73.1% (2009) and 62.8% (2010) of African Americans completed their

contracts (or received an “A” or a “B”). The fact that more African Americans did not meet the workload (quantity) requirement of their grading contracts may account for their generally lower portfolio scores (quality) in 2010. As mentioned earlier, 2010 appeared to have stricter contract obligations for everyone. Regardless of what affected writing quality or fewer African Americans meeting their contract obligations, the ubiquitous use of grading contracts in our FYW program suggests one thing about our African American racial formation: they have a harder time generally meeting the workload expectations of the contracts. Despite this difficulty, their writing ends up being of proficient quality, and like the API formation, African Americans do best in the key outcome of “summary and conversation,” showing statistically significant change in 2010, the stricter year. Ultimately, for African Americans more work did seem to equate to higher quality and more growth in writing, but at the cost of more African Americans not meeting the workload requirements.

African Americans’ open-ended responses were mostly positive, with 42 positive and 37 negative comments. The positive comments contained three strong themes, with 11 concerning “expectations and clarity,” 8 concerning “relieves pressure,” and 12 concerning “effort over quality.” Like APIs, comments in the “expectations and clarity” theme praised the contract’s ability to make “it more clear on what we were expected to do throughout the class.” The comments themed as “relieves pressure” were similar to the API comments themed as “motivation and staying on track,” except the focus in African American comments tended toward the contract’s way of relieving the pressure created by constantly anticipating grades: “it puts less strain on my grades, and I can just focus on the content of my work, and not whether or not it’ll effect my grades.” African American comments in the “effort over quality” category focused on how the emphasis on effort makes a writer feel less like a bad writer: “you never feel like you totally failed something, but instead you can see your strengths or weaknesses more.”

African Americans found contracts effective and had some of the highest rates (92.5% in 2010) of satisfaction. In 2010, they were also happiest with contracts with 75% approving of contracts but least happy of all formations in 2009 at only 57% (Table 4). Perhaps the somewhat erratic findings can be explained by the fewer number of responses from African Americans, which is due to Fresno State’s low enrollment of African American students.

### Whites

In 2009, white students’ portfolio ratings showed little growth (overall difference .40), but averaged in the “acceptable” range of ratings at both English 5A midterm and 5B final (Table 2). They generally started higher in the qual-

ity of their work (overall 3.01) than any other formation, but ended lower than all other racial formations (overall 3.41). In this same year, 89.2% of all whites met their contract obligations, which is noticeably higher than all students (80.1%), and a higher percentage than any other racial formation achieved. In 2010, they also generally started with higher 5A midterm ratings (overall 2.67) and ended with generally higher 5B final ratings (overall 3.63) than any other formation; however, whites achieved the least amount of growth in quality (.96) between 5A midterm and 5B final, just as in the previous year (.40). In 2010, they also completed their contract obligations at again one of the highest levels, with 79.6% getting “As” and “Bs.” The only group in 2010 who did more work was their API counterparts.

Interestingly, for the white formation, it is unclear as to whether quantity equated to quality, or much growth in quality. In 2009, they received lower English 5B final portfolio ratings (overall 3.41) than all other formations and achieved less growth than all others (overall difference .40), but more whites completed their contracts’ workload expectations that year than any other formation (89.2%). Meanwhile in 2010, whites had higher English 5B final ratings than any other formation (overall 3.63), stronger growth (but still less than all other racial formations, with overall difference .96), yet fewer white students met the workload expectations of the contract than had the previous year (79.6%). These findings suggest that there may not be nearly as strong of a connection between the contract’s focus on quantity of work producing higher quality of work, or producing intrinsic motivation for writing.

White students found the contract as effective as most other racial formations in both years (74.75% and 76.67%), although the rates are generally a little lower. Their happiness with the contract was also on par with all other racial formations (70.30% and 70.97%). However, white students’ preferences for grading contracts were the lowest of all racial formations in both years (68.32% and 58.24%).

White students’ open-ended comments included 117 positive and 109 negative comments. The two strongest positive themes were “relieves pressure” (29 comments) and “expectations and clarity” (21 comments). Most comments in the “relieves pressure” theme focused on allowing more time to develop writing without the pressure of grades and taking risks without being penalized by grades: “it alleviates the pressure to do well on each assignment. I think there is some flexibility to try to take risks with writing strategies [...] without being penalized.” The positive comments by white students in the theme of “expectations and clarity” were similar to those expressed by APIs and African American students.

The biggest difference in comments was in negative comments by white students in the themes, “no grades/how well am I doing?” and “unfair to those who traditionally do well already.” The following was typical of whites’ negative comments and fits in both themes, which matches findings by Spidell and Thelin (2006):

I do not like knowing whether I stood in an A or B range for both english 5a and 5b throughout the semesters, I believe that the teacher should be more accurate and grade with the normal system, this way I would know where I stand [...] I feel like it is a lot easier for students to cheat their way to an A, it should be something that you have to work for, which should not be solely depend on a person's quality of portfolio, it should be based off of the work throughout the semester.

The seeming contradiction in the student’s response refers to not receiving grades on individual assignments and the dependence on the final portfolio to determine an “A” grade. The source of many white students’ negative comments seemed to be frustration centered on the ambiguity of grades.

White students’ comments on the theme of “unfair to those who traditionally do well already” further suggested a deep attachment to traditional grading systems. Perhaps whites’ higher rates of “As” and “Bs” are one indication that many in this formation expected those grades, and expected them to be given out less frequently to others.

In the end, all three measures of effectiveness (portfolio ratings, exit surveys, and grade distributions) were inconsistent. Grading contracts did seem to be marginally effective in producing growth in white students’ writing quality but least effective in generating student acceptance. Despite the higher rates of completing the contract’s workload obligations, and high rates of perceived effectiveness and happiness with the contract, white students did not prefer the contract at as high a rate as any other racial formation. Contracts did not seem to harm white students, however, as they produced sufficient quality of writing and growth with respect to the FYW program’s outcomes.

## Conclusions

Based on the findings of this study, grading contracts at Fresno State are most effective for APIs, somewhat effective for African Americans, and marginally effective for whites. By far, contracts were most effective for our Asian Pacific Islanders, who are mostly Hmong with low parental education levels, often speak Hmong in their homes, and are mostly identified as remedial. Contracts produced similar quality and development, even along the same dimensions in their writing, for African Americans, who have higher levels of parental education, speak English at home, and have high rates of remedial

status. African Americans, however, had more difficulty meeting all the contract obligations. The quantity of work expected in contracts appeared to be more difficult to complete for more African American students than any other racial formation, which may account for their lower preference rates in 2010. Finally, for the white student formation, who spoke mostly English at home, had high levels of parental education, and the lowest remedial rates, the contract proved to be least effective. While they still met program expectations in terms of the quality of their writing, they rated lower than all others in final portfolios, and had smaller differences in growth in their writing over a semester and half of instruction with the contract. Additionally, while they had the highest rates of completing the contract, they preferred it at the lowest rates, and had more (relatively speaking) negative responses in surveys.

These conclusions suggest that grading contracts like the ones used at Fresno State, and those promoted by Shor (1996) and Danielewicz and Elbow (2009), tend to be more effective for students who are predisposed to seeing—or can be convinced to view—grades as unhelpful, destructive, or harmful to their learning. At Fresno State, these students tend to be of color, have other languages spoken in their homes, and come from homes with parents who tend to have not gone to college. The majority are identified as needing remediation. Finally, the students for whom grading contracts were most effective were those who either see grades as punishment, as limiting their choices and decisions in writing, as producing pressure to get things right, as reducing freedom to write, or as de-motivating in some way, or those who see themselves as not being good writers already. The API formation at Fresno State best fit this profile.

My findings about the effectiveness of grading contracts for API, African American, and white students suggest that any grading technology may very well affect various racial formations differently. The effectiveness of any grading technology often hinges on the assumptions that the technology makes about the nature of quality writing, the relationship of quality to the workload of the course, the assumptions students must accept in order for the grading technology to function properly, and with whom that technology is interacting (about whom it makes decisions and who makes decisions). What I have not been able to inquire about in this study is teachers' identities, which surely influenced the effectiveness of the grading contracts. The gendered and racialized ethos of teachers may have affected how some students responded to the contract. For instance, most of our FYW teachers are white, middle-class female teaching associates (TAs) in their early to mid-twenties.

In the end, the present study clearly suggests that the effectiveness of grading contracts in classrooms is unevenly distributed among racial formations,



and it sheds light on potential biases within Fresno State's grading contract approach. The most important, I think, is a bias towards student effort rather than quality of writing. While quality is a measure of effectiveness, it is not *the only measure of effectiveness* in our contract grading system. In fact, when quality is the primary bias in a grading technology, then I think we doom many students to failure or near failure. Grading technologies with quality biases tend to oppress many students, but especially poorer students and students of color. So maybe it's not simply that grades harm students by placing them into hierarchies, but they harm them by restricting their freedom to write, or taking their sense of self worth away, because their writing just will not be judged "high quality." Fairer grading technologies would seem to be those with biases that allow all students the ability to achieve the full range of grades. Contracts do this by rewarding effort and labor, and these biases seem more fitting for a truly democratic and diverse society.

## Notes

---

<sup>1</sup> I want to thank Tina Leimer, the Director of IRAP and two of her Research Analysts, Hongtao Yue, and Dmitri Rogulkin, for their statistical help.

<sup>2</sup> All information on languages spoken at home and parents' education level comes from anonymous entry surveys in English 5A for Fall 2009. Student demographic numbers, GPAs, SAT Comp scores, and remediation numbers come from CSUF's Office of Institutional Research, Assessment, and Planning office.

<sup>3</sup> While CSU, Fresno uses a DSP model, all CSU students must take the EPT before enrolling, which the university uses to determine remedial status (still). The EPT is administered by ETS and consists of one 45-minute essay section and two 30-minute multiple-choice sections. All remediation statistics by race were obtained from California State University, 2009, n. p.

<sup>4</sup> Most students take English 5A and 5B with the same instructor. English 5A uses a grading contract to determine the credit/no credit grade in the course. Student survey responses suggest that approximately 80% of all English 5B classes in both Spring 2009 and Spring 2010 also used grading contracts.