Peralta Community College District
Berkeley City College, College of Alameda, Laney College, Merritt College

ESL Writing Assessment Validation Report

Submitted to

Patty Falero
Administrative Support
California Community College Chancellor's Office

Dr. Tzu-Yun Chin
Buros Center for Testing
University of Nebraska-Lincoln

Prepared by

Dr. Eun Rhee
Research Specialist

Nathan Pellegrin
Director of Institutional Research

Peralta Community College District
333 East 8th Street
Oakland, CA 94606
(510) 466-7210
npellergin@peralta.edu

November 15, 2014

Peralta Community College District
Berkeley City College, College of Alameda, Laney College, Merritt College

ESL Writing Assessment Validation Report

Introduction

The four colleges in the Peralta Community College District (PCCD) utilize an approved locally managed ESL writing assessment instrument to place students into ESL writing courses. For placement into ESL listening and speaking and grammar courses, the colleges use ACT COMPASS ESL tests, which are approved second-party assessment instruments. In summer 2014, the Peralta ESOL Advisory Council (PEAC) modified the placement rubric of the ESL writing assessment instrument to be aligned with a newly adopted ESL curriculum sequence. At the same time, the prompts used in the writing assessment were also revised based on input from test proctors, readers and faculty, with the goal that all prompts be of similar complexity, free of bias, and to involve comparisons. This process was led by faculty, with input and facilitation from the Office of Institutional Research.

This report describes the studies that were conducted in summer and fall 2014 to assess the reliability and validity of the new writing prompts and placement rubric in accordance with the procedures for local districts that manage a direct performance assessment instrument. Specifically, this report provides evidence addressing prompt bias, interprompt agreement, content-validity of the placement rubric, interscorer reliability of the scoring rubric, and consequential validity of the placement process. This report also identifies points where further work is needed, and plans for addressing those needs.

Background on PCCD ESL Curriculum

The new, accelerated ESL core curriculum is a four level program (High Beginning, Intermediate, High Intermediate, and Advanced) for the subject areas of reading and writing, listening and speaking, and grammar. The goals of the new curriculum are 1) to support students to improve their ability to speak and understand both oral and written English, and 2) to develop reading, writing, and critical thinking skills. The courses for reading and writing are listed in Table 1.

Table 1: PCCD ESL Reading and Writing Courses

Level	Course
High Beginning	285A/B
Intermediate	222A/B
High Intermediate	223A/B
Advanced	52A/B

Evaluation of Writing Prompts for Bias

In summer 2014, the PEAC created new prompts for the ESL writing assessment test as well as a new format. The writing assessment test is a timed 30 minutes test in which students choose one of two questions to answer in an essay. They are asked to give specific reasons and examples to explain their opinion (See Appendix A for the test format).

A committee composed of ESL faculty, Disabled Students Programs and Services (DSPS) faculty, and others described below reviewed the new writing prompts for possible cultural or linguistic bias, insensitivity, or offensiveness (see Appendix B for a list of the prompts). The committee composition was diverse in terms of ethnicity and gender, reflecting the student population at PCCD. Two committee members were also selected from DSPS for expertise in student disabilities. Table 2 describes the composition of the nine reviewers.

Table 2: Reviewers of Prompts for Bias

Ethnicity	Gender	Expertise
African American	Female	DSPS instructor (A)
African American	Female	ESL student (B)
Asian	Female	Researcher (C)
Asian	Male	Director of Asian & Pacific American Student Success Program (D)
Hispanic	Male	DSPS instructor (E)
White	Female	3 ESL instructor (F, G, H)
White	Male	ESL instructor (I)

Each reviewer individually evaluated the 11 prompts for evidence of any kind of bias, insensitivity, or offensiveness. They were asked to comment on any prompts that may pose a problem. Table 3 describes the comments by the reviewers on the 11 writing prompts. After a discussion about the comments for each prompt, the committee decided that the 5 prompts (Prompts 7 to 11) that used too high a level of grammatical sophistication, or were deemed potentially insensitive, should be excluded. Although comments concerning Prompts 3, 5, and 6 involved possible lack of experience or clarity of concept for some students, they were retained because they were not perceived as biased, and students would have a choice of two prompts. That is, if a student does not have experience with or opinion about one prompt, then he or she can choose the other prompt to write about (see bottom of Appendix A for an example). As suggested by the Chancellor's office, giving students a choice in prompts was considered a good resolve to issues of possible bias in the prompts. Thus, 6 prompts (Prompts 1 to 6) that were judged to be generally free of cultural or linguistic bias were retained.

Table 3: Reviewer Comments

Prompt	Comments
1	
2	
3	D: "Some students may not have experienced life in a small town and may experience a disadvantage compared to those who have both experiences."
4	
5	D: "Some students may not have experienced life in another country to create a reasonable comparison response."
6	B: "It is not clear what my family wants? They want from me or what they want in general."
7	F: "How much can anyone say about eating in a restaurant vs. eating at home?" G: "It's not a prompt that I see much to write about."
8	F: "Requires a lot of vocabulary and fairly difficult grammar."
9	F: "Students may not have opinions about a day or a night job."
10	G: "Assumes students have experience with computers";" a great digital divide affect many of our students, particularly older students" H: "Computer question might be challenging for older students."
11	G: "Potentially insensitive for students who have lost their parents."

Placement Rubric and Content Validity

A representative sample of students from all levels of ESL classes was chosen during the summer 2014 term to write sample essays using the new prompts. Each student chose one prompt from a set of two prompts. The six prompts were arranged into 12 orderings such that each prompt was included twice, once in the first position and once in the second position (see bottom of Appendix A for the format), as a means to check for preference associated with the order of presentation. These sample essays were used to establish content validity, interscorer reliability, and interprompt reliability.

These data informed the modification of the writing placement rubric which was revised by the PEAC to accord with the new ESL curriculum sequence. The content-related validity of the placement rubric was conducted by a representative committee of 5 ESL instructors. Content validity was established by ensuring a direct relationship between the writing placement guide and the district-wide agreed upon objectives, course prerequisite skills, and writing skill expectations for the reading and writing courses (see Appendix C for the placement rubric). The placement rubric consists of a 1 to 6 scale, corresponding to the curriculum as shown in the table 4.

Table 4: Placement Rubric for ESL Reading and Writing Courses

Score	Course	Description
1	Refer to Adult School	Refer to Adult School
2	ESL 285A	High Beginning Reading and Writing
3	ESL 222A	Intermediate Reading and Writing
4	ESL 223A	High Intermediate Reading and Writing
5	ESL 52A	Advanced Reading and Writing
6	Refer to English	Refer to English Department for Placement

A norming session was conducted with 5 ESL instructors, who rated 20 sample essays to evaluate the descriptors of the placement rubric and then discussed, revised, and refined the levels until they were completely satisfied. Scoring of the writing samples was based upon a holistic reading; thus, descriptors of each placement level reflect the dominant characteristics of the writing sample at each of these levels. Specifically, the committee focused on the overall content and four specific areas: addressing prompt, organization, development of ideas, and clarity. They also identified anchor papers for each level.

Interscorer Reliability

To assess interscorer reliability, a 3-reader system was used. Specifically, after the three faculty readers were trained on the new placement rubric, two readers independently rated a set of 50 randomly selected essays (see below for the description of the training session). If the two readers did not agree on a score for a sample essay, then the third reader resolved the inconsistency between the first two readers. For example, if two scores out of three were the same, a valid score had been reached for the sample. If three different scores were given by three readers, then the three readers had to conference and reach a consensus on a score. Rater agreement was calculated by correlating the ratings from the two independent readers. The interscorer reliability coefficient was .89.

After establishing interscorer reliability with the new placement rubric, a training session was conducted with 20 potential new readers. During the session, the placement rubric was reviewed and discussed. The two trainers, who participated in the previous norming session, verbally went over the characteristics for placement at each level and discussed the 5 anchor papers (Adult School, High Beginning, Intermediate, High Intermediate, and Advanced). The participants examined labeled writing examples at each level and then discussed how each example related to the placement rubric. They identified the characteristics of that example that prevented it being scored at the next higher level or the next lower one. Each participant then received identical unlabeled examples at each level to score. Results were charted for all participants to examine and discrepancies were discussed in the whole group. Then, groups of three participants did simulated placement sessions using the three-reader system, and continued to work until they reached agreement 90% of the time.

Interprompt Agreement

Because multiple prompts will be used in the ESL writing assessment, we examined interprompt agreement. We were unable to conduct double testing of students to assess equivalent-form reliability of the prompts, thus we conducted a preliminary interprompt agreement with the ratings of the 50 essay samples. These essays were not ideal for assessing interprompt agreement because the students were not randomly assigned a prompt and we did not have an equal number of students for each prompt. Table 5 describes the percentage of students who chose each prompt. Of the 50 essays, 3 students did not choose a prompt to address because of low English proficiency.

Table 5: Students' Choice of Prompt

Prompt	Percent Chosen
1	8%
2	20%
3	14%
4	16%
5	20%
6	16%
None*	6%
	100%

Note: * Low English proficiency students who did not choose a prompt. The readers had assigned them to adult school.

Ten students chose Prompt 2 and another ten chose Prompt 5. We attempted to assess interprompt agreement with these two prompts because they had the most number of students. However, as can be seen in Table 6, although the prompt distributions show some overlap, there are too few students to obtain a reliable result. Thus, we are conducting a follow-up study to establish equivalency of prompts in the spring term.

Table 6: Interprompt Agreement

Prompt	n	Rating					
		1	**2**	**3**	**4**	**5**	**6**
2	10	20%	20%	30%	20%	0%	10%
5	10	10%	30%	20%	30%	10%	0%

Consequential Validity

In fall 2014, a pilot consequential validity study was conducted to evaluate the placement accuracy of the ESL writing assessment process. During the sixth week of classes, a survey was administered to students and instructors in the five reading and writing ESL courses at Berkeley City College (see Appendices D and E for the student and instructor surveys). Students were asked about their satisfaction with their placement into their ESL reading and writing courses by indicating whether their placement was the right level, too difficult, or too easy. Instructors were asked to assess whether each first time student was appropriately placed in the course. Specifically, they were asked, "Please indicate the appropriateness of each student's placement by completing the following statement. This student is:
1. Very overprepared, definitely should be in the next level.
2. Somewhat overprepared, perhaps should be in the next level.
3. Well prepared, should pass with reasonable effort.
4. Somewhat underprepared, perhaps should be in previous level.
5. Very underprepared, definitely should be in previous level."

For the instructor ratings, ratings of 2, 3, and 4 were combined as "adequately prepared" (appropriate placement), 1 as "overprepared", and 5 as "underprepared".

Only the students who took the writing assessment test at Berkeley City College were included in evaluating the appropriateness of the placement. Enrollment records were matched with assessment data. Because Berkeley City College typically has high level ESL students, only the two higher level reading and writing ESL courses were offered in fall 2014. Approximately 131 students were enrolled in 5 reading and writing ESL courses (see Table 7). Of these, 79 first time students took the writing assessment test. Most of the other students in the courses were continuing in the ESL sequence and did not take the placement test or they took the placement test outside Peralta colleges.

Instructors were asked to rate all first time students in their classes. First time students were chosen because they were most likely to have recently taken the ESL writing assessment test. Student self-ratings were obtained from the students in attendance on the day of the survey. The instructor and student surveys were conducted in class using paper surveys.

Table 7: Description of Participating ESL Courses

Courses	Sections Surveyed	Enrolled	Instructor Ratings	Student Ratings
High Intermediate Reading and Writing (ESL 223A)	2	55	33	25
Advanced Reading and Writing (ESL 52A)	3	76	46	29
Total	5	131	79	54

Students were generally very satisfied with their placement (See Table 8). Both levels of ESL classes had a higher than 75% agreement rate that the course was the right level, 84% for ESL 223A and 93% for ESL 52A. Overall, only 4% of the students felt they were not qualified for the course (too difficult), while 7% felt they were overqualified (too easy). Slightly more students felt they were overqualified for ESL 223A (12%).

Table 8: Student Rating of Placement Accuracy

Courses		Not Qualified	Qualified	Overqualified	Total
High Intermediate Reading and Writing (ESL 223A)	n	1	21	3	25
	%	4.0	84.0	12.0	100
Advanced Reading and Writing (ESL 52A)	n	1	27	1	29
	%	3.4	93.1	3.4	100
Total	n	2	47	4	54
	%	3.7	88.9	7.4	100

Note: For the two ESL 223A classes, the qualified ratings were 81.8% and 85.7%. For the three ESL 52A classes, the qualified ratings ranged from 87.9% to 100%.

As can be seen in Table 9, instructors were generally very favorable in the placement of the students. Both levels of ESL classes had rates higher than 75% agreement that the students were appropriately placed, 94% for ESL 223A and 96% for ESL 52A. Overall, only 5% of the students were thought to be underprepared and no student was considered overprepared.

Table 9: Instructor Rating of Placement Adequacy

Courses		Underprepared	Adequately Prepared	Overprepared	Total
High Intermediate Reading and Writing (ESL 223A)	n	2	31	0	33
	%	6.1	93.9	0	100
Advanced Reading and Writing (ESL 52A)	n	2	44	0	46
	%	4.3	95.7	0	100
Total	n	4	75	0	79
	%	5.1	94.9	0	100

Note: For the two ESL 223A classes, the adequately prepared ratings were 89.5% and 100%. For the three ESL 52A classes, the adequately prepared ratings ranged from 90% to 100%.

Both instructors and students evidence high levels of satisfaction with the placement process. However, because these sample sizes are small and only the two higher level courses were evaluated, we will conduct follow-up studies with larger sample of students and with all four levels of ESL classes in the spring term.

Disproportionate Impact

Disproportionate impact could not be assessed because of the small sample size, restricted range of classes, and only one college was included in the pilot study. In order to build a sample sufficiently large enough to support basic statistical inference for each subgroup, the Office of Institutional Research will be conducting consequential validity surveys in ESL writing courses each term until an adequate sample size is reached. Once we obtain a large enough sample we will apply the "80% rule" on data disaggregated by gender, age, disability status, primary ethnicity, and possibly other significant subgroups within the ESL population who may be adversely affected by the placement process (such as East African immigrants).

Conclusion

The findings from the studies suggest that 1) the new prompts are generally free of bias and the choice format may be a good resolve to issues of possible bias in the prompts, and 2) the revised placement rubric appears effective in placing students in the appropriate level of ESL reading and writing courses.

The results also indicate a need for further studies to assess equivalent-form reliability of the prompts, consequential validity with all four levels of ESL reading and writing courses, and disproportionate impact of the placement process. We are implementing plans to address these key issues. To establish equivalent-form reliability additional tests will be proctored and scored

according to the rubric drafted by the PEAC. The PEAC will be presented with the findings and, if necessary, the district will convene a team to address deficiencies that are found. Also, beginning in spring 2015 we will conduct a consequential validity survey of ESL students across all levels of ESL and in all colleges and combine results from fall 2014. If the resulting sample size is still not large enough to warrant conclusions in a study examining disproportionate impact, then additional surveys will be conducted in fall 2015.

Finally, we look forward to receiving feedback from the assessment working group regarding the information provided in this report, and shaping our work plan for the coming year accordingly.

Name

| Last (family) | First | Middle |

Student ID #

Date   /   /

☐ REFER to ENGL

☐ ESL 52A (ADV)

☐ ESL 223A (HIGH INT)

☐ ESL 222A (INT)

☐ ESL 285A (HIGH BEG)

☐ ADULT SCHOOL

Peralta Community College District

English for Speakers of Other Languages

Writing Placement Test

Directions:

Please do not turn this page until the test-giver tells you what to do.

On the next page, you will see two questions. Choose ONE of the two questions and answer it in an essay. In your essay, give specific reasons and examples to explain your opinion.

You have 30 minutes.

Please do not talk or use a dictionary. You may use a pencil or pen.

Do your best and write as much as you can.

Good luck!

Please answer ONE of these questions in an essay.   #1

| Which do you think is more important, love or money? | **OR** | Which do you think is better, life now or life 100 years ago? |

Appendix B: Writing Prompts

Prompt 1: Who do you think has a harder life, men or women?

Prompt 2: Which do you think is better, life now or life 100 years ago?

Prompt 3: Which do you think is a better place to live, a big city or a small town?

Prompt 4: Which do you think is more important, love or money?

Prompt 5: Which do you think is easier, life in your country or life in the United States?

Prompt 6: Which do you think is more important, doing what you want or doing what your family wants?

Prompt 7: Which do you think is better, eating in a restaurant or eating at home?

Prompt 8: Which do you think is better, a friend who is different from you or a friend who is the same as you?

Prompt 9: Which do you think is better, a day job or a night job?

Prompt 10: Which do you think is better, life with computers or life without computers?

Prompt 11: Who do you think has a better life, you or your parents?

Appendix C: Writing Placement Rubric

Peralta Community College District ESOL Writing Placement Rubric
Developed June 2014 by Peralta ESOL Advisory Council

	Refer to English Category 2	Advanced Reading/Writing ESL 52A	High Intermediate Reading/Writing ESL 223A	Intermediate Reading/Writing ESL 222A	High Beginning Reading/Writing ESL 285A	Refer to Adult School
Addressing Prompt	• may address prompt with a sense of purpose and audience • goes beyond the concrete • may address other perspectives/ make concessions		understands and provides an intelligible, extended response to prompt		understands and provides an intelligible response to prompt	• doesn't address prompt • seems not to understand task
Organization	• main points clear and distinct • paragraphs have clear internal organization • uses transitional techniques effectively		• main points mostly clear • paragraphs have some internal organization		• main points unclear • may be "stream-of-consciousness"	none
Development	supports points with clear, specific, and sufficient examples and details			• attempts to support points • support may lack clarity or relevance	points minimally or not supported	
Clarity	• strong control of structures and specific vocabulary • errors do not interfere with meaning • does not require rereading	• sentence structure and vocabulary sufficient to express meaning • some sentence variety • may require occasional rereading		• some effective sentence structure and vocabulary • may require some rereading	• limited sentence structure and vocabulary • requires rereading	not comprehensible

Refer to English Category 1: regardless of essay writing skill, writer has idiomatic, fluent English; if errors, similar to fluent English speaker errors.

Appendix D: Consequential Validity Survey – Student Survey

PCCD **ESL** Placement Validation Student Survey - Fall 2014
To be completed by the student

Please take a few minutes to provide feedback on this course. Your judgment will help
Peralta CCD to improve course placement processes.

Date _____   Birthdate _____

Name _____

Course Section (circle one):   50A   52A   215A   216A   217A   222A   223A   232A   233A
                               283A   284A   285A   Other _____

1. Which **ONE** of the following statements is most true about your placement in this
   course?
   _____ This course is the right level for me.
   _____ This course is too difficult for me.

   _____ This course is too easy for me.

2. How did you get placed into this course?
   _____ Placement test at one of the Peralta colleges (Alameda, Berkeley, Laney, Merritt)

   _____ Placement test outside Peralta colleges

   _____ Completed prerequisite course

   _____ Successfully challenged the prerequisite

   _____ Other _____

3. If you took the placement test at one of the Peralta colleges, did you prepare for the test
   by studying?
   _____ Yes   _____ No

4. To what extent are personal reasons (illness, job/family responsibilities, personal
   problems, etc.) interfering with your performance in this course?
   _____ Not at all

   _____ Somewhat

   _____ Very much

5. Do you have enough time to meet the studying/homework demands of this course?
   _____ Yes   _____ No

Appendix E: Consequential Validity Survey – Instructor Survey

PCCD Placement Validation Survey - Fall 2014

To be completed by the instructor

Date _____ Class section _____

Instructor's name _____

As part of the revalidation process for Peralta Community Colleges' assessment placement tool, we need data from you about the students in your class.

Please indicate the appropriateness of each student's placement by completing the following statement. This student is:

    1 = Very overprepared, definitely should be in the next level.

    2 = Somewhat overprepared, perhaps should be in the next level.

    3 = Well prepared, should pass with reasonable effort.

    4 = Somewhat underprepared, perhaps should be in previous level.

    5 = Very underprepared, definitely should be in previous level.

Student Name	SID	Rating