ESL Essay Placement Assessment Validation, 2017-2018 Study

Prepared by Lan Hoang (research analyst) and Jay Cull (senior IT specialist)*

The California Community College Chancellor's Office (CCCCO) governs matriculation and course placement assessment in California community colleges (CCC) through Title 5 and a set of standards published in the 1990 report (CCCCO 2001). The report requires local colleges to ensure they meet the standards for matriculation and course placement assessment. For test instruments developed by a second party external to the CCC, test publishers are responsible for evaluating the instruments' reliability, validity, and bias while local colleges adopting the instruments are charged with evaluating the uses of the instruments. Specifically, at a minimum, the college is responsible for evaluating the predictive validity of a placement system or set of instruments, the cut scores used for placement, differential impact on various demographic groups, and accommodations for individuals who cannot take tests under standard conditions.

Sacramento City College uses a locally developed, locally managed written essay assessment as part of the multiple measures for ESL Writing placement. In December 2016, the ESL Department implemented a change in prompts and rubrics for the ESL essay assessment. The new prompts were screened for bias and the development of the new scoring rubrics was in line with the ESL Program Review process, which are described in details in the ESL Essay Assessment Validation 2014-2015 Study.

This validation study is conducted with data from the new assessment and includes analyses for reliability, consequential validity, predictive validity, and disproportionate impact.

Interrater and Inter-Prompt Reliability Validation

This validation seeks to answer two key questions about SCC's ESL writing prompts and scoring: first, whether any two essay readers are likely to give the same essay a similar score, and second, whether any given student who enters any given administration and receives any given writing prompt will be likely to receive a similar score. Because it is impractical to randomly assign students to different prompts within a single essay exam administration, rather than being true experimental designs, SCC uses quasi-experimental modified versions of Research Designs 17 and 18 in its essay validation process.[1] Data were analyzed using interrater reliability and analysis of variance (ANOVA) techniques.

---

* Initial data collection and analysis were done by Anne Danenberg, former research analyst at PRIE. Some of the analysis (and text) in this report were done by Anne and are used in this report with her permission. Essay score data were provided by ESL Professor Brett Thomas. However, Lan Hoang takes full responsibility for the accuracy and analyses in this study. Any questions about the study should be directed to Lan Hoang at HoangL@scc.losrios.edu.

[1] See *Assessment Validation Project Local Research Options* at CCC Chancellor's Office

The study includes 530 essay scores from December 2016 to November 2017—including at least 30 essays from all but two writing prompts being used during the time period (See Appendix C for Scoring Rubric). Essay scores are the mode score of the two raters. When the two raters give different scores for a given essay, a third rater is called upon and the score assigned would be the mode score of the three raters. The scores are then used along with standardized test scores (Accuplacer and Compass) and a weighted non-cognitive item (level and type of previous education) to place students using multiple measures.

Interrater Reliability Results

We used the first two raters' scores for the interrater reliability analysis. Four statistical approaches are employed to meet the interrater reliability criterion: (1) Pearson correlation coefficient greater than + .70, (2) at least 90% of the scores within one point of each other, (3) an average difference between readers' scores of less than one point, and (4) Cronbach's *alpha*. Table 1 lists each method for determining agreement along with statistics for each of the methods.

Table 1. Interrater Agreement Statistics

Method	Value
N	530
Mean	3.417
Standard Deviation	1.524
(1) Pearson *r* (sig. at 0.001)	.948
(2) Pairs <= 1pt. (%)	98.868
(3) Avg. Difference	.225
(4) Cronbach's alpha	.973

The "value" column lists each test statistic for this sample. For each of the methods above, the test statistic meets or exceeds the threshold. Pearson's *r* is +0.95 and is highly statistically significant (p<.001), which substantially exceeds the minimum standard of +0.70. The percentage of rating-pairs within a point of each other is almost 99%, solidly meeting the criterion. The average difference between raters' scores is well-below 1 for the sample (.225). Analysis using Cronbach's *alpha* shows a high degree of inter-rater reliability, with the scale reliability coefficient of 0.97.

Each statistic meets the reliability criteria described in *Research Design 18*, demonstrating that there is likely to be a high degree of agreement between pairs of readers for any given essay. Therefore, it is unlikely that reader variation is influencing the student's total score more than the quality of the student's written essay. Furthermore, these statistics suggest that the readers' training and procedures, including use of the scoring rubric, are consistent and effective.

Inter-prompt Reliability Results

The same sample with 530 essay scores from December 2016 to November 2017 is used for this analysis. The essay scores used are the mode scores of individual students' essays by the first two

---

http://extranet.cccco.edu/Portals/1/SSSP/Matriculation/Assessment/AssessmentValidationProjectLocalResearchOptionsFeb.1991.pdf (last retrieved 07/12/2017).

raters, or, in the case when there is disagreement between the first two raters' scores, the mode score of the three raters when a third rater is called upon.

There were twelve prompts in used during the time period of the study (prompts numbered 40 to 51) with scores ranges from 0 to 7. Figure 1 below illustrates the mean essay scores by prompt (Prompt Mean) for the sample, which range from 2.955 (Prompt 41) to 3.735 (Prompt 45), with overall mean for the sample being 3.417. The number shown in each column is the number of essays included in the respective prompt sample. Except for Prompt 46 and 47, each of the prompts has an N larger than 30.

Figure 1. Mean Essay Score by Prompt
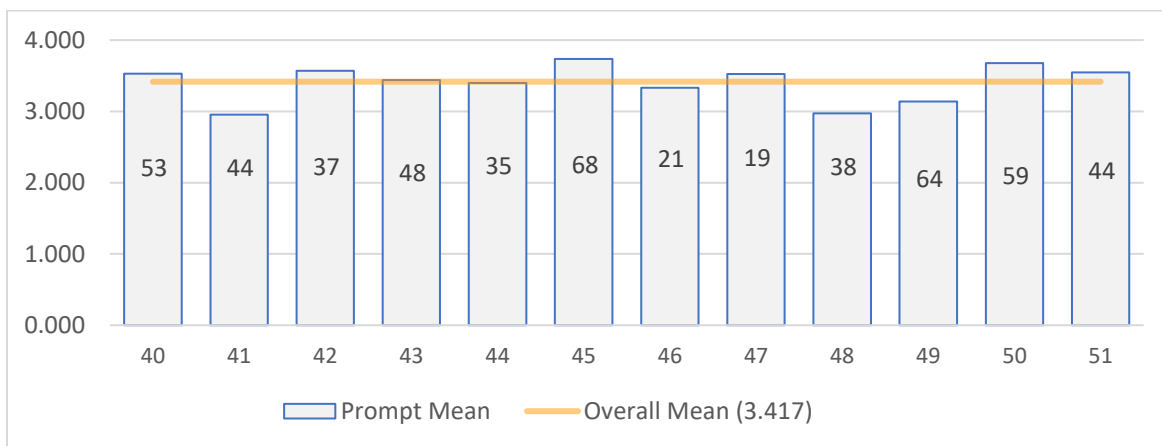
Test statistic (i.e. t-test) shows that none the prompt means are significantly above or below the overall mean across all prompts of 3.417 (p > .05). Analysis of variance (ANOVA) also indicates that the variation in mean essay scores by prompt are not statistically significant (F-ratio = 1.393, p = .172). (Appendix A presents essay score descriptive statistics by prompt, t-test results, ANOVA results, and score distribution by prompt.) Prompt accounts for 2.9 percent of the variation in score but which is not statistically significant (eta-squared = .029) and thus the observed differences in essay scores are unlikely to be attributable to the prompts and might have been due to random factors.

We used Bartlett's statistic for equal variances and Levene's test for equality of variances to examine the essay score variance in greater detail. Bartlett's statistic suggests that the variances for the prompts adhere to ANOVA's assumption of equal variances (Bartlett's K-squared = 11.958, p = 0.367). Levene's robust estimation also suggests that the essay score variances by prompt support the null hypothesis of equal variances (F = .853, p = .587).

These results suggest that, on average, students' performance in one prompt would be expected to be similar to other prompts.

Discussion

Each of the test statistics to evaluate inter-rater reliability meets the standard for reliability. It is unlikely that variations in raters' scores has much effect on students' essay scores. Regardless of which test statistic we calculate, the evidence demonstrates that ESL writing prompts are interchangeable and have little effect on essay scores. It is more likely that student ability is reflected in the essay scores, which, when combined with other measures of ESL ability and

student characteristics that are related to student outcomes (multiple measures) place students in particular course levels. However, it should be noted that we had small N for a couple of prompt samples (prompt numbered 46 and 47). Data collection and analysis should be continued to ensure prompts in used continue to be interchangeable and consistent with the overall score mean across all prompts.

Consequential Validity: Student and Faculty Evaluation of Placement

Research Design 15 was used to examine consequential validity of placement results. Mid-semester student and faculty surveys on students' level of preparedness were administered in the Spring and Fall 2017. All sections of ESLW30, ESLW40, ESLW50, ESLW310, ESLW320, ESLW340, ESL55, ESL315, and ESL325 were surveyed. Note that because the number of students placed into higher course levels is extremely small, data collection is an ongoing process in ESL as of Fall 2017.

Student surveys were conducted in class by teaching faculty and faculty surveys were administered via email in which teaching faculty sent in their ratings of individual students' preparedness in excel file attachments. Student responses were then matched to corresponding faculty ratings. Table 2 lists the placement levels, the course(s) corresponding to a given placement level, and the number and percentage of matched responses for students who indicated that they met the prerequisite through the assessment process. Note that students may elect to enroll in any course lower than their placement as well as the course in which they were placed.

Table 2. Summary of Survey Sample by ESL Writing Placement Level and Course
(Spring & Fall 2017 pooled)

Level	Course	Number of matched responses	Percent
1	ESLW30	71	31.982
2	ESLW40	46	20.721
3	ESLW50, ESL55	65	29.279
4	ESLW310, ESL315	24	10.811
5	ESLW320, ESL325	13	5.856
6	ESLW340	3	1.351
	Total	**222**	**100**

Table 3 below contains overall pooled student and faculty survey results and Table 4a & 4b contain the response distributions by placement level and corresponding courses.
Table 3 suggests that these students are more confident about their preparation levels than are faculty. Overall, faculty think that about 68% of students have adequate preparation while 84% of ESL and ESLW students placed through the assessment process have the same perception about themselves. Note that of the students who answered "adequately prepared" and completed the course in the same semester, about 86% successfully completed the course, i.e. having an official

grade of at least a "C" (155 students out of 181 with an official grade). Nevertheless, ANOVA results show that there is statistically significant variation in student success rates by course level ($F = 2.599$, $p = .026$). It might be that faculty are teaching to the deficiencies in their classes and mitigating those deficiencies by the end of the term, which would result in a higher proportion of passing students.[2]

Table 3. Student and Faculty Survey Results Overall for Students Placed by Assessment
(Spring & Fall 2017 pooled)

Range is 1=under-prepared to 3=over-prepared	N = 222			
Mean estimation	**Mean**	**Std. Err.**	**[95% Conf. Interval]**	
How well-prepared are you? (N = 222)	2.000	0.027	1.947	2.053
How well-prepared is this student? (N = 222)	1.937	0.038	1.863	2.011
Level of work in the class (N = 221*)	1.995	0.024	1.949	2.042

Item 2 and Faculty Rating				
How well-prepared is this student for the work in this class?	**Student Perception**		**Faculty Rating**	
	Number	**Percent**	**Number**	**Percent**
Under-prepared	18	8.108	42	18.919
Adequately-prepared	186	83.784	152	68.468
Over-prepared	18	8.108	28	12.613
Total	**222**	**100**	**222**	**100**

Item 3		
Describe the level of work in this class:	**Student Perception**	
	Number	**Percent**
Too hard	14	6.335
Just right	194	87.783
Too easy	13	5.882
Total	**221**	**100**

* N not equal 222 due to missing data

Table 4a and 4b below provide evidence that there is substantial variation in perceptions of preparedness across course levels.

Table 5 shows respective success rates by course level. Table 4a shows the student and faculty evaluations of placement for the lower levels of the ESL and ESLW courses studied. Because faculty say that only 63% of students placed by assessment are "adequately-prepared," ESLW30 faculty ratings do not meet the 75% threshold. However, 83% of students believe that they are at least adequately prepared for the work they encounter after being placed by the placement process,

---

[2] This is similar to what we saw in the ENGWR for Spanish prerequisite study at SCC 2013 https://goo.gl/VKYDG8.

with over 87 percent indicating that the level of work in the class was "just right" for them. Note that ESLW30 is the lowest level course and has no prerequisite. Furthermore, the success rate of students who said they were placed through the placement process in this level is more than 78%—students who assessed into this level are more likely to succeed than not. Therefore, even though the faculty ratings at this level does not meet the 75% threshold, it is not an indication that lower level course should be added to the curriculum.

Student and faculty evaluations in the second level met the 75% threshold with 80% and 76% respectively, with 93 percent of the students described the level of work being "just right." In the third level, student evaluation met the threshold with 84% but faculty evaluation was short by about 6 percent (69%). However, nearly 91% of the students indicated that the level of work was "just right" and the success rate of students who assessed into this level is over 92%.

<center>Table 4a. Student and Faculty Survey Results for Students Placed by Assessment
by Placement Level (Spring & Fall 2017 pooled)</center>

LEVEL 1	Student Perception		Faculty Rating	
How well-prepared is this student for the work in this class?	Number	Percent	Number	Percent
Under-prepared	8	11.268	17	23.944
Adequately-prepared	59	83.099	45	63.380
Over-prepared	4	5.634	9	12.676
Total	71	100	71	100

LEVEL 1	Student Perception	
Describe the level of work in this class:	Number	Percent
Too hard	4	5.714
Just right	61	87.143
Too easy	5	7.143
Total	70	100

LEVEL 2	Student Perception		Faculty Rating	
How well-prepared is this student for the work in this class?	Number	Percent	Number	Percent
Under-prepared	1	2.174	5	10.870
Adequately-prepared	37	80.435	35	76.087
Over-prepared	8	17.391	6	13.043
Total	46	100	46	100

LEVEL 2	Student Perception	
Describe the level of work in this class:	Number	Percent
Too hard	0	0.000
Just right	43	93.478
Too easy	3	6.522
Total	46	100

LEVEL 3	Student Perception		Faculty Rating	
How well-prepared is this student for the work in this class?	Number	Percent	Number	Percent
Under-prepared	7	10.769	11	16.923
Adequately-prepared	54	83.077	45	69.231
Over-prepared	4	6.154	9	13.846
Total	65	100	65	100

LEVEL 3	Student Perception	
Describe the level of work in this class:	Number	Percent
Too hard	3	4.615
Just right	59	90.769
Too easy	3	4.615
Total	65	100

Table 4b shows the student and faculty evaluations of placement for the higher levels of the ESL and ESLW courses studied. The 75% threshold is met for the fourth level, but not for the two highest levels. Note that there are only thirteen students in Level 5 and three in Level 6 who said they assessed into the course. Twelve out of 13 students in Level 5 (92%) perceived that they were adequately prepared while the faculty gave the same rating for only 8 of them (62%). Student and faculty perception difference notwithstanding, the success rate of students who said they assessed into this level is almost 85%. As for Level 6, all three said they were adequately prepared but the faculty said that two were under-prepared and only one is adequately prepared. Two of the students said the course is "too hard" and only one said that the course is "just right." Only one of them successfully completed the course. The department may want to consider raising the cut-score for the highest level.

Table 4b. Student and Faculty Survey Results for Students Placed by Assessment
by Placement Level (Spring & Fall 2017 pooled)

LEVEL 4	Student Perception		Faculty Rating	
How well-prepared is this student for the work in this class?	Number	Percent	Number	Percent
Under-prepared	2	8.333	4	16.667
Adequately-prepared	21	87.500	18	75.000
Over-prepared	1	4.167	2	8.333
Total	24	100	24	100

LEVEL 4	Student Perception	
Describe the level of work in this class:	Number	Percent
Too hard	4	16.667
Just right	20	83.333
Too easy	-	-
Total	24	100

LEVEL 5	Student Perception		Faculty Rating	
How well-prepared is this student for the work in this class?	Number	Percent	Number	Percent
Under-prepared	-	-	3	23.077
Adequately-prepared	12	92.308	8	61.538
Over-prepared	1	7.692	2	15.385
Total	13	100	13	100

LEVEL 5	Student Perception	
Describe the level of work in this class:	Number	Percent
Too hard	1	7.692
Just right	10	76.923
Too easy	2	15.385
Total	13	100

LEVEL 6	Student Perception		Faculty Rating	
How well-prepared is this student for the work in this class?	Number	Percent	Number	Percent
Under-prepared	-	-	2	66.667
Adequately-prepared	3	100	1	33.333
Over-prepared	-	-	-	-
Total	3	100	3	100

LEVEL 6	Student Perception	
Describe the level of work in this class:	Number	Percent
Too hard	2	66.667
Just right	1	33.333
Too easy	-	-
Total	3	100

Table 5. Success by Course Level
among students placed through placement process (self-reported)

Course level	Number	Count of Success	Success Rate
1	69	54	78.261
2	44	40	90.909
3	63	58	92.063
4	24	20	83.333
5	13	11	84.615
6	3	1	33.333
Total	216*	184	85.185
Note: Total does not sum up to 222 due to missing data			

Predictive Validity

The literature consistently shows that placement tests alone do not predict academic outcomes very well (for example, EdResults Partnership, 2014). Although the ESL Essay alone is not being used to place students and is one of a few multiple measures being used in the placement process at SCC, its predictive ability is explored here.

Transcript data from the study time period were matched to assessment data. Only 127 matches were successfully made using institutional data. When we examine correlations between essay score and subsequent course success, there is variation across courses. However, in some situations we do find that the higher the essay score is, the more likely the success is.

Table 6 presents the correlations between essay score and numeric grade and success. For the two highest levels, there are insufficient matching observations to calculate the correlation.

Table 6. Correlations between EssayScore and NumGrade / Success

	Overall	ESLW30	ESLW40	ESLW50 ESL55	ESLW310 ESL315	ESLW320 ESL325	ESLW340
Correlation	(obs=127)	(obs=34)	(obs=25)	(obs=47)	(obs=15)	Insufficient Obs.	Insufficient Obs.
	EssayScore	EssayScore	EssayScore	EssayScore	EssayScore		
NumGrade	.218*	.540***	.366	.199	.100	—	—
Success	.256**	.408*	.291	.091	.218	—	—
*** p = .001; ** p = .01; * p = .05							

Based on the evidence presented in this table, it appears that essay scores have weak positive linear relationship with numeric grades and success but the correlation is statistically significant. Essay scores have weak to moderate positive relationship with numeric grade and success in ESLW30, ESLW40, and ESLW50, and little relationship with numeric grade and success in ESLW310. Only

correlation coefficients at the lowest level are statistically significant. The numbers of essay scores in the higher course levels are small so generalizability is limited.

Disproportionate Impact on Special Populations

Colleges have a responsibility to monitor any disproportionate impact on student subpopulations and to plan remedies to address any disproportionate impact that is found. This design follows *Research Design 12* and provides evidence to answer the question of whether ESLW essay scores differ significantly for students in specific gender, age, race/ethnicity, language groups for ESL, disability, or level of income groups. Although we examine the full set of essay scores for the ESL essay, in some cases the "cell size" in a particular combination of score and student's characteristic is too small to be of use in drawing conclusions about whether a subpopulation is over- or under-represented and whether there is, indeed a disproportionate impact of essay score on a given subpopulation. Furthermore, disproportionate impact may not be problematic if some groups are disproportionately impacted in a way that we would consider a positive impact rather than a negative impact. For example, scoring disproportionally high would not be considered an adverse impact, while scoring disproportionally low would presumably have an adverse impact on students writing essays. Appendix B contains the full range of scores and comparisons to the overall distribution of scores.

We examine ESL essay scores to see if any groups are receiving low scores at disproportionately high levels or receiving high scores at disproportionately low levels compared to the majority group in the tested sample. Because group sizes would get so small for any meaningful analysis or generalization when broken down by essay score, we group essay scores into two categories: lower score range (0-3) and higher score range (4-7). Table 7 shows student characteristics in lower score range and

Table 8 shows student characteristics in higher score range. In each of the tables, the Disproportionality Index column is calculated by dividing the minority group proportion by the majority group proportion—the farther the resulted number is to 1, the larger the gap is between the minority group proportion and the majority group proportion. The column Minority-Majority % Point Difference is calculated by subtracting the minority group proportion from the majority group proportion. We use a variation of the EEOC (2007) 80-20 rule: percentages are highlighted if the minority group's score proportionality is more than 20% above the percentage of the majority group in the lower score range (the % point difference carrying a positive sign) and if the minority group's score proportionality is more than 20% below the percentage of the majority group in the higher score range (the difference carrying a negative sign).[3] These are noted in *italics* when the cell size is below 30 and ***bold italics*** in grey background when cell size is 30 or more.

---

[3] As elaborated in the introduction paragraph in the section, the rationale is that, in the lower score range, we are interested in identifying groups receiving low scores at a disproportionately higher percentage than the majority group. Similarly, in the higher score range, we are interested in identifying groups receiving high scores at a disproportionately lower percentage than the majority group.

In terms of gender, male is used as the majority group. As for race/ethnicity, although Hispanic/Latino students make up the largest group in the college student population[4] and Asian students the largest in the sample in this study (see Table B.1. in Appendix B), we followed the conventional majority group for English courses being White (CCCCO Assessment Validation Training, 2012). We treated groups by primary language the same way, with English being the majority group. Students in the age group of 18-20 have the highest percentage in the college's student body and are considered the majority group in this study. Regarding income level, middle and above is considered the majority group. We did not include analysis for Disability here because of the small number (only 6 students) but Table B.5. in Appendix B shows essay score by disability (DSPS) status.

Table 7. Student characteristics in lower essay score range (0-3)

Group	Group Proportion	Majority Group Proportion	Minority-Majority % Point Difference	Disproportionality Index
GENDER				
Female	34.7%	49.2%	-14.5%	.7
RACE/ETHNICITY				
African American*	20.0%		-18.8%	.5
Asian	45.7%		6.9%	1.2
Filipino*	25.0%		-13.8%	.6
Hispanic/Latino*	29.1%	38.8%*	-9.7%	.8
Multi-Race*	33.3%		-5.5%	.9
Pacific Islander*	0.0%		-38.8%	-
Unknown*	100.0%		*61.2%*	2.6
AGE				
Under 18*	32.8%		-1.8%	.9
21-24*	31.8%		-2.8%	.9
25-29*	43.7%	34.6%*	9.1%	1.3
30-39*	66.7%		*32.1%*	1.9
40 or over*	100.0%		*65.4%*	2.9
PRIMARY LANGUAGE				
Chinese (Cantonese)*	60.0%		*31.7%*	2.1
Chinese (Mandarin)*	43.8%		15.5%	1.5
Farsi (Persian)*	66.7%		*38.4%*	2.4
Russian*	38.5%	28.3%*	10.2%	1.4
Spanish*	36.6%		8.3%	1.3
Vietnamese*	47.0%		18.7%	1.7
All Other*	44.7%		16.4%	1.6
INCOME LEVEL				
Below Poverty	48.9%	23.7%*	*25.2%*	2.1
Low*	39.3%		15.6%	1.7
* Note: Cell size (number) is smaller than 30.				

---

[4] Fall 2016 data, Hispanic/Latino students account for 32% of the student population https://goo.gl/nZNWD1

Table 8. Student characteristics in higher essay score range (4-7)

Group	Group Proportion	Majority Group Proportion	Minority-Majority % Point Difference	Disproportionality Index
GENDER				
Female	65.4%	49.1%	16.3%	1.3
RACE/ETHNICITY				
African American*	80.0%		18.8%	1.3
Asian	54.2%		-7.0%	.9
Filipino*	75.0%		13.8%	1.2
Hispanic/Latino	70.9%	61.2%*	9.7%	1.2
Multi-Race*	66.6%		5.4%	1.1
Pacific Islander*	100.0%		38.8%	1.6
Unknown*	0.0%		*-61.2%*	-
AGE				
Under 18*	0.0%		*-65.3%*	-
21-24	67.3%		2.0%	1.0
25-29	68.2%	65.3%	2.9%	1.0
30-39	56.3%		.9%	.9
40 or over*	33.4%		*-31.9%*	.5
PRIMARY LANGUAGE				
Chinese (Cantonese)*	40.0%		*-31.8%*	.6
Chinese (Mandarin)*	56.3%		-15.5%	.8
Farsi (Persian)*	33.3%		*-38.5%*	.5
Russian*	61.5%	71.8%	-10.3%	.9
Spanish*	63.3%		-8.5%	.9
Vietnamese*	52.9%		-18.9%	.7
All Other*	55.2%		-16.6%	.8
INCOME LEVEL				
Below Poverty	51.1%	76.3%*	*-25.2%*	.7
Low	60.8%		-15.5%	.8
* Note: Cell size (number) is smaller than 30.				

In the lower essay score range, students with income below poverty are disproportionately overrepresented. Although students of *unknown* race/ethnicity, older students (in 30-39 and 40 or over age groups), students whose primary language is Chinese (Cantonese) and Farsi (Persian) seem to have disproportionately higher proportion compared to the majority groups, the small cell sizes make it difficult to generalize. Similarly, in the higher essay score range, these students groups appear to be disproportionately underrepresented but the cell sizes are smaller than 30. Although Appendix B contains several instances where students have over- or under-representation in an essay score, most of the individual cell sizes are too small to draw strong conclusions about the extent of any disproportionate impact in the general population of students writing an essay for the purpose of course placement recommendations.

Again, although the results in this section are difficult to generalize because some "cell sizes" are quite small for most of the groups, there is evidence that students with income below poverty are

disproportionately overrepresented in the lower essay score range and disproportionately underrepresented in the higher score range. This requires the ESL Department to initiate process to mitigate the impacts, such as working with the college's student services and student support programs to reach out to students with income below poverty and provide additional instructional supports for the group. In addition, the college is responsible for close monitoring of the assessment placement processes and will continue to collect data and conduct future validation studies as part of program review.

Conclusion

The validation study evaluated reliability, consequential validity, predictive validity, and disproportionate impact of the ESL Essay Assessment Placement. The study found empirical evidence that inter-rater and inter-prompt reliability of the ESL writing scoring and prompts meet the standard for reliability while noting that the small N for a couple of prompt samples requires the college to continue data collection and analysis.

Results from consequential validity study suggest that students are more confident about their preparation levels than are faculty. There is substantial variation in perceptions of preparedness across course levels. While students' ratings meet the 75% threshold for adequate preparedness in each of the course levels, faculty ratings do not in Level 1, 3, 5, and 6, noting that the sample sizes for Level 5 and 6 are very small (13 and 3 respectively). Despite the fact that there is difference in student and faculty perception on students' preparedness, with the exception of Level 6, the success rate of students who said they assessed into a certain level exceed 78%. Consistent with the findings in the 2014-2015 ESL Essay Assessment Placement Validation Study, the Department may want to consider raising the cut-score for the highest level.

In terms of predictive validity, while it appears that there is a positive linear relationship between essay scores and grades/success, the relationship is weak and is only statistically significant for the overall sample and at the lowest level. Once again, generalizability is limited because of the small sample sizes of the higher course levels.

Regarding disproportionate impact on special populations, there is variation in the proportionality of some subgroups' representativeness in the lower and higher essay score ranges but which is not generalizable because of the small cell sizes. Nevertheless, the study found empirical evidence that students from households with income below poverty are disproportionately overrepresented in the lower essay score range and disproportionately underrepresented in the higher essay score range. The ESL Department needs to initiate process to mitigate the impacts through working with student services and supports units and providing additional instructional supports to this student group. It is also required that the college continue to monitor the assessment placement process by way of continued data collection and validation studies.

References

Assessment Validation Project Local Research Options at CCC Chancellor's Office
http://extranet.cccco.edu/Portals/1/SSSP/Matriculation/Assessment/AssessmentValidationProjectLocalResearchOptionsFeb.1991.pdf   (last retrieved 02/21/2018)

CCCCO (2001). *Standards, Policies and Procedures for the Evaluation of Assessment Instruments Used in the California Community Colleges*. http://extranet.cccco.edu/Portals/1/SSSP/Matriculation/Assessment/ApprovedGuidelinesMarch2001.pdf (last retrieved 02/21/2018)

CCCCO Assessment Validation Training. (2012). *Studies Monitoring Disproportionate Impact.* *http://extranet.cccco.edu/Portals/1/SSSP/Matriculation/Assessment/DisproportionateImpact.pdf* (last retrieved 02/21/2018)

Educational Results Partnership and the RP Group (2014). *Multiple Measures for Assessment and Placement.* https://www.bakersfieldcollege.edu/sites/bakersfieldcollege.edu/files/MMAP_WhitePaper_Final_9-10-14.pdf   (last retrieved 02/21/2018)

EEOC - Equal Employment Opportunity Commission. (2007). CFR 2007 Title 29, §1607.5. http://www.gpo.gov/fdsys/pkg/CFR-2007-title29-vol4/pdf/CFR-2007-title29-vol4-sec1607-4.pdf (last retrieved 02/21/2018)

APPENDIX A

This appendix presents additional information supporting the inter-prompt reliability analysis (Research Design 18).

Table A.1. Descriptive statistics of scores by prompt

PROMPT	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max	T-TEST (compares prompt mean to overall mean)	
					Lower Bound	Upper Bound			t-stat	Sig.(2-tailed)
40	53	3.528	1.367	0.188	3.151	3.905	0	6	.593	.556
41	44	2.955	1.613	0.243	2.464	3.445	0	6	-1.901	.064
42	37	3.568	1.608	0.264	3.031	4.104	0	7	.570	.572
43	48	3.438	1.623	0.234	2.966	3.909	0	7	.088	.931
44	35	3.400	1.499	0.253	2.885	3.915	1	6	-.067	.947
45	68	3.735	1.442	0.175	3.386	4.084	0	6	1.821	.073
46	21	3.333	2.033	0.444	2.408	4.259	0	7	-.189	.852
47	19	3.526	1.020	0.234	3.035	4.018	2	6	.467	.646
48	38	2.974	1.423	0.231	2.506	3.442	0	6	-1.920	.063
49	64	3.141	1.424	0.178	2.785	3.496	0	6	-1.553	.126
50	59	3.678	1.580	0.206	3.266	4.090	0	7	1.268	.210
51	44	3.545	1.547	0.233	3.075	4.016	0	6	.551	.585
Total	**530**	**3.417**	**1.524**	**0.066**	**3.287**	**3.547**	**0**	**7**		

Table A.2. ANOVA results for scores by prompt

PromptScore					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	35.300	11	3.209	1.393	0.172
Within Groups	1193.548	518	2.304		
Total	1228.847	529			

Figure A.1. Essay Score Distribution by Prompt

Prompt 40

Prompt 41

Prompt 42

Prompt 43

Prompt 44

Prompt 45

Prompt 46

Prompt 47

Prompt 48

Prompt 49

Prompt 50

Prompt 51

Frequency

PromptScore
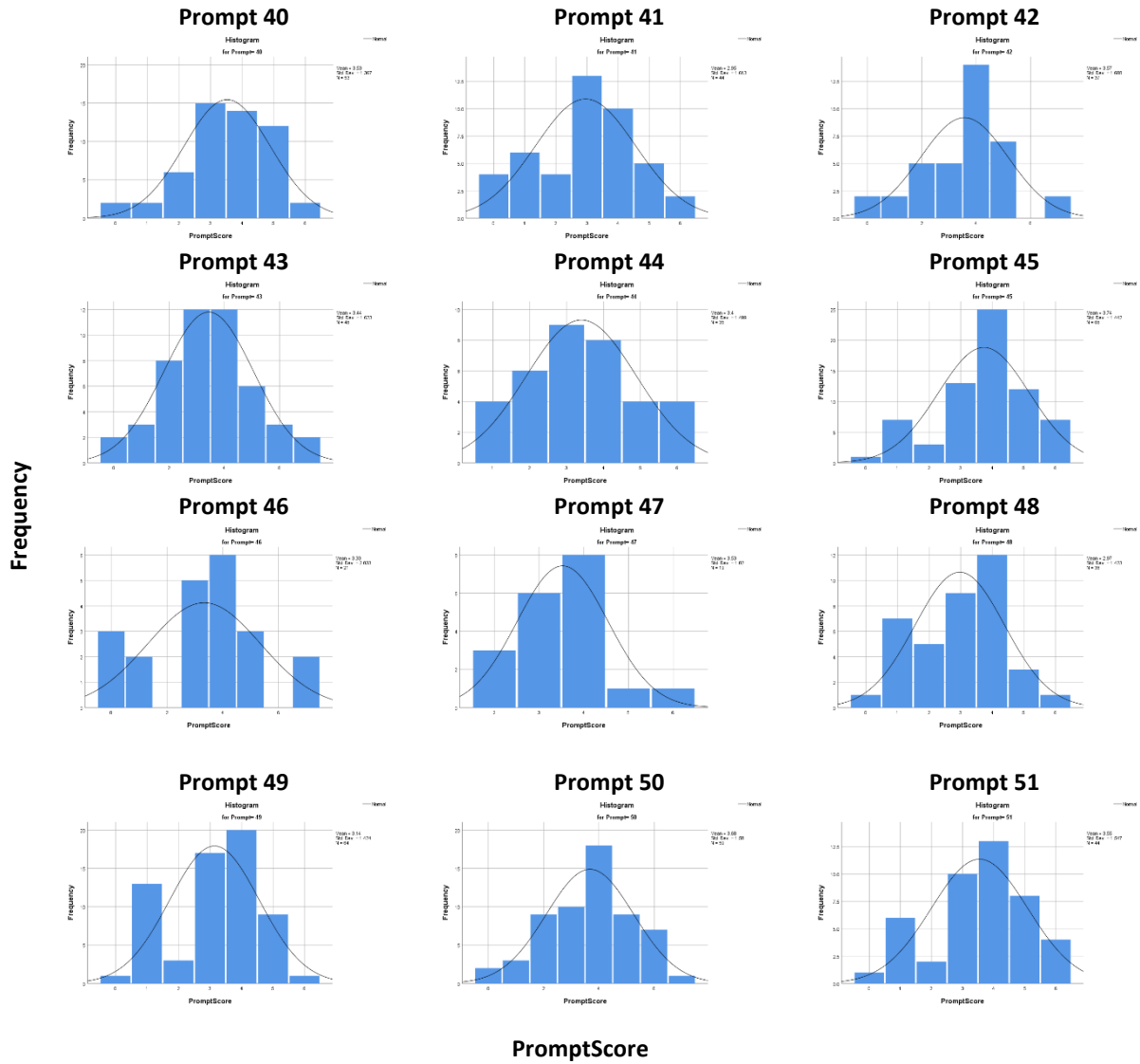
APPENDIX B

Table B.1. Characteristics of Students Tested (with student EMPLID matched to profile data)

GENDER	Frequency	Percent
Female	156	57.1
Male	116	42.5
Unknown	1	0.4
Total	*273*	*100.0*
RACE/ETHNICITY	**Frequency**	**Percent**
African American	5	1.8
Asian	164	60.1
Filipino	4	1.5
Hispanic/Latino	55	20.1
Multi-Race	3	1.1
Pacific Islander	2	0.7
Unknown	4	1.5
White	36	13.2
Total	*273*	*100.0*
AGE	**Frequency**	**Percent**
Under 18	2	0.7
18-20	75	27.5
21-24	55	20.1
25-29	44	16.1
30-39	55	20.1
40 or over	42	15.4
Total	*273*	*100.0*
DISABILITY	**Frequency**	**Percent**
Not disabled	267	97.8
Disabled	6	2.2
Total	*273*	*100.0*

PRIMARY LANGUAGE	Frequency	Percent
Chinese (Cantonese)	30	11.0
Chinese (Mandarin)	32	11.7
English	85	31.1
Farsi (Persian)	15	5.5
Russian	26	9.5
Spanish	30	11.0
Vietnamese	17	6.2
All Other	38	14
Total	*273*	*100.0*
INCOME LEVEL	**Frequency**	**Percent**
Below Poverty	139	50.9
Low	74	27.1
Middle And Above	38	13.9
Unable to Determine	22	8.1
Total	*273*	*100.0*
ESSAY SCORE	**Frequency**	**Percent**
0	5	1.8
1	19	7.0
2	28	10.3
3	60	22.0
4	93	34.1
5	45	16.5
6	21	7.7
7	2	0.7
Total	*273*	*100.0*

Table B.2. Essay Scores by Gender

			GENDER			Total
			Female	Male	Unknown	
EssayScore	0	Count	4	1	0	5
		% within GENDER	2.6%	0.9%	0.0%	1.8%
	1	Count	10	9	0	19
		% within GENDER	6.4%	7.8%	0.0%	7.0%
	2	Count	9	19	0	28
		% within GENDER	5.8%	16.4%	0.0%	10.3%
	3	Count	31	28	1	60
		% within GENDER	19.9%	24.1%	100.0%	22.0%
	4	Count	59	34	0	93
		% within GENDER	37.8%	29.3%	0.0%	34.1%
	5	Count	27	18	0	45
		% within GENDER	17.3%	15.5%	0.0%	16.5%
	6	Count	16	5	0	21
		% within GENDER	10.3%	4.3%	0.0%	7.7%
	7	Count	0	2	0	2
		% within GENDER	0.0%	1.7%	0.0%	0.7%
Total		Count	156	116	1	273
		% within GENDER	100.0%	100.0%	100.0%	100.0%

Table B.3. Essay Scores by Race/Ethnicity

			RACE								Total
			African American	Asian	Filipino	Hispanic /Latino	Multi- Race	Pacific Islander	Un- known	White	
EssayScore	0	Count	0	2	0	3	0	0	0	0	5
		% w/in RACE	0.0%	1.2%	0.0%	5.5%	0.0%	0.0%	0.0%	0.0%	1.8%
	1	Count	0	13	0	2	0	0	0	4	19
		% w/in RACE	0.0%	7.9%	0.0%	3.6%	0.0%	0.0%	0.0%	11.1%	7.0%
	2	Count	1	18	0	3	1	0	2	3	28
		% w/in RACE	20.0%	11.0%	0.0%	5.5%	33.3%	0.0%	50.0%	8.3%	10.3%
	3	Count	0	42	1	8	0	0	2	7	60
		% w/in RACE	0.0%	25.6%	25.0%	14.5%	0.0%	0.0%	50.0%	19.4%	22.0%
	4	Count	4	46	2	20	1	2	0	18	93
		% w/in RACE	80.0%	28.0%	50.0%	36.4%	33.3%	100.0%	0.0%	50.0%	34.1%
	5	Count	0	30	1	11	1	0	0	2	45
		% w/in RACE	0.0%	18.3%	25.0%	20.0%	33.3%	0.0%	0.0%	5.6%	16.5%
	6	Count	0	12	0	8	0	0	0	1	21
		% w/in RACE	0.0%	7.3%	0.0%	14.5%	0.0%	0.0%	0.0%	2.8%	7.7%
	7	Count	0	1	0	0	0	0	0	1	2
		% w/in RACE	0.0%	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%	2.8%	0.7%
Total		Count	5	164	4	55	3	2	4	36	273
		% w/in RACE	100%	100%	100%	100%	100%	100%	100%	100%	100%

Table B.4. Essay Scores by Age Group

			AgeGroup						Total
			18-20	21-24	25-29	30-39	40 or over	Under 18	
Essay Score	0	Count	2	3	0	0	0	0	5
		% within AgeGroup	2.7%	5.5%	0.0%	0.0%	0.0%	0.0%	1.8%
	1	Count	4	3	3	4	4	1	19
		% within AgeGroup	5.3%	5.5%	6.8%	7.3%	9.5%	50.0%	7.0%
	2	Count	4	4	4	5	11	0	28
		% within AgeGroup	5.3%	7.3%	9.1%	9.1%	26.2%	0.0%	10.3%
	3	Count	16	8	7	15	13	1	60
		% within AgeGroup	21.3%	14.5%	15.9%	27.3%	31.0%	50.0%	22.0%
	4	Count	21	21	18	22	11	0	93
		% within AgeGroup	28.0%	38.2%	40.9%	40.0%	26.2%	0.0%	34.1%
	5	Count	17	11	9	6	2	0	45
		% within AgeGroup	22.7%	20.0%	20.5%	10.9%	4.8%	0.0%	16.5%
	6	Count	10	5	3	2	1	0	21
		% within AgeGroup	13.3%	9.1%	6.8%	3.6%	2.4%	0.0%	7.7%
	7	Count	1	0	0	1	0	0	2
		% within AgeGroup	1.3%	0.0%	0.0%	1.8%	0.0%	0.0%	0.7%
Total		Count	75	55	44	55	42	2	273
		% within AgeGroup	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table B.5. Essay Scores by DSPS Status

			DSPS		
			NO	YES	Total
EssayScore	0	Count	5	0	5
		% within DSPS	1.9%	0.0%	1.8%
	1	Count	19	0	19
		% within DSPS	7.1%	0.0%	7.0%
	2	Count	27	1	28
		% within DSPS	10.1%	16.7%	10.3%
	3	Count	60	0	60
		% within DSPS	22.5%	0.0%	22.0%
	4	Count	88	5	93
		% within DSPS	33.0%	83.3%	34.1%
	5	Count	45	0	45
		% within DSPS	16.9%	0.0%	16.5%
	6	Count	21	0	21
		% within DSPS	7.9%	0.0%	7.7%
	7	Count	2	0	2
		% within DSPS	0.7%	0.0%	0.7%
Total		Count	267	6	273
		% within DSPS	100.0%	100.0%	100.0%

Table B.6. Essay Scores by Primary Language

			Chinese (Cantonese)	Chinese (Mandarin)	English	Farsi (Persian)	Russian	Spanish	Vietnam -ese	All Other	Total
EssayScore	0	Count	0	0	3	0	0	2	0	0	5
		% within PRIM_LANG	0.0%	0.0%	3.5%	0.0%	0.0%	6.7%	0.0%	0.0%	1.8%
	1	Count	4	1	5	0	2	1	3	3	19
		% within PRIM_LANG	13.3%	3.1%	5.9%	0.0%	7.7%	3.3%	17.6%	7.9%	7.0%
	2	Count	3	2	6	4	4	1	4	4	28
		% within PRIM_LANG	10.0%	6.3%	7.1%	26.7%	15.4%	3.3%	23.5%	10.5%	10.3%
	3	Count	11	11	10	6	4	7	1	10	60
		% within PRIM_LANG	36.7%	34.4%	11.8%	40.0%	15.4%	23.3%	5.9%	26.3%	22.0%
	4	Count	4	11	32	2	13	13	6	12	93
		% within PRIM_LANG	13.3%	34.4%	37.6%	13.3%	50.0%	43.3%	35.3%	31.6%	34.1%
	5	Count	6	4	19	3	2	4	3	4	45
		% within PRIM_LANG	20.0%	12.5%	22.4%	20.0%	7.7%	13.3%	17.6%	10.5%	16.5%
	6	Count	2	3	9	0	1	2	0	4	21
		% within PRIM_LANG	6.7%	9.4%	10.6%	0.0%	3.8%	6.7%	0.0%	10.5%	7.7%
	7	Count	0	0	1	0	0	0	0	1	2
		% within PRIM_LANG	0.0%	0.0%	1.2%	0.0%	0.0%	0.0%	0.0%	2.6%	0.7%
Total		Count	30	32	85	15	26	30	17	38	273
		% within PRIM_LANG	100%	100%	100%	100%	100%	100%	100%	100%	100%

Table B.5. Essay Scores by Income Level

			INCOME_LEVEL_DESC				Total
			Below Poverty	Low	Middle And Above	Unable to Determine	
EssayScore	0	Count	3	2	0	0	5
		% within INCOME_LEVEL	2.2%	2.7%	0.0%	0.0%	1.8%
	1	Count	18	1	0	0	19
		% within INCOME_LEVEL	12.9%	1.4%	0.0%	0.0%	7.0%
	2	Count	14	9	3	2	28
		% within INCOME_LEVEL	10.1%	12.2%	7.9%	9.1%	10.3%
	3	Count	33	17	6	4	60
		% within INCOME_LEVEL	23.7%	23.0%	15.8%	18.2%	22.0%
	4	Count	44	25	16	8	93
		% within INCOME_LEVEL	31.7%	33.8%	42.1%	36.4%	34.1%
	5	Count	16	12	10	7	45
		% within INCOME_LEVEL	11.5%	16.2%	26.3%	31.8%	16.5%
	6	Count	10	8	2	1	21
		% within INCOME_LEVEL	7.2%	10.8%	5.3%	4.5%	7.7%
	7	Count	1	0	1	0	2
		% within INCOME_LEVEL	0.7%	0.0%	2.6%	0.0%	0.7%
Total		Count	139	74	38	22	273
		% within INCOME_LEVEL	100.0%	100.0%	100.0%	100.0%	100.0%

APPENDIX C

REVISED SCORING RUBRIC FOR ESL COURSES, SACRAMENTO CITY COLLEGE

Revised Dec. 2016/ Valid Dec. 1 2016 - Present

0   RECOMMEND ADULT SCHOOL

A paper in this category will reveal no ability to communicate in written English:

- no response at all
- a non-English response

1   RECOMMEND PLACEMENT IN ESLW 30 (weak writing sample)

A paper in this category will reveal a severely limited ability to communicate in written English. It will have one or more of the following:

- the copied prompt or part of it
- an incomprehensible, illogical, or incoherent response

2   RECOMMEND PLACEMENT IN ESLW 30 (adequate or strong writing sample)

A paper in this category will reveal a severely limited ability to communicate in written English. It will have one or more of the following:

- an extremely brief response
- severe and frequent errors in basic sentence structure and/or basic tenses & forms
- limited vocabulary
- word order patterns frequently do not follow standard English word order
- severe and frequent spelling errors

3   RECOMMEND PLACEMENT IN ESLW 40

A paper in this category communicates but with serious limitations in the use of written English. It will:

- attempt to address the topic but may not develop it sufficiently or may go off topic at some point
- have some appropriate examples and details but may be brief or may be overly repetitive (with deceptive length)
- have simple and compound clause structures but may show frequent sentence boundary errors
- show appropriate use of simple and progressive tenses and verb forms but with regular errors
- have control of basic vocabulary but show inconsistent control of word choice, word form, collocations and/or spelling, which may interfere with meaning
- usually follow standard English word order patterns
- show inconsistent control of punctuation, capitalization and/or formatting (indenting, writing on lines)

4   RECOMMEND PLACEMENT IN ESLW 50

A paper in this category communicates with some limitations in the use of written English. It will:

- logically address the topic with increasing length, but may show simplistic organization and/or uneven development
- state a clear point at either the beginning or end of the writing and support it
- attempt compound and complex sentence structures, often with shaky clause control and some sentence boundary errors

- have general control of basic tenses and forms and appropriate use of simple and progressive, but with inconsistent control of more advanced tenses
- have basic control of topical vocabulary but may include some inappropriate word choice, w ord form, or collocation errors, or some misspellings of high frequency words
- have basic control of punctuation and capitalization
- display a pattern or accumulation of errors which may sometimes interfere with meaning

5   RECOMMEND PLACEMENT IN ESLW 310

A paper in this category will demonstrate developing competence in writing but remain weak on either the rhetorical or syntactic level. It will:

- address the topic with a clearer focus than at lower levels
- be written as one paragraph with several developed supporting points and a starting and finishing idea (paragraphing implied) OR as a minimally developed multi-paragraph composition that may be formulaic, with limited introduction and conclusion sections
- show good control of simple, compound and some complex sentences with logical use of connectors and transitions; good control of sentence boundaries
- demonstrate strong control of simple and progressive tenses and verb forms but may have inconsistent control of perfect tenses
- use generally appropriate word choice with some errors in word form and spelling of academic vocabulary
- begin to rise above a conversational tone to a more academic written register but may still rely more on extended narration or description rather than on analysis

6   RECOMMEND PLACEMENT IN ESLW 320

A paper in this category will show developing competence in academic writing and will:

- address the topic thoughtfully with a clearly stated (or implied) thesis
- show a substantial amount of writing with some explanation or analysis of the ideas and examples
- show a firmer control of logical idea flow , with better parallelism and concrete support for points
- contain sentences that show a variety of clauses and generally correct punctuation
- contain some errors in grammar that may rarely interfere with meaning; word forms and choices will be generally correct
- show emerging variety and sophistication of vocabulary; spelling of academic vocabulary is generally correct

7   RECOMMEND PLACEMENT IN ESLW340

A paper in this category will show basic competence in academic writing and will:

- address the topic in a clear and effective way with a clearly stated (or strongly implied) thesis
- show a substantial amount of writing
- have a fairly consistent analytical tone with evidence of critical thinking skills
- have logical development, smooth flow of ideas , and supportive details and examples which respond to the topic
- include the correct use of more sophisticated complex and compound-complex sentences with few errors in sentence structure, transitions , or parallel structure ; sentences show a high level of fluency
- have few errors in verb tenses, word form, or spelling; these errors will only rarely interfere with meaning
- include a variety of sophisticated vocabulary although collocation and article errors may persist.