

ESL University Students' Testing Preferences

ANNELA TEEMANT

Indiana University-Purdue University Indianapolis

This descriptive study examined university ESL students' testing preferences in content-area courses. Thirty-eight ESL students participated by completing a survey. They used a semantic differential scale of ten adjective pairs to rate their preferences regarding five test formats: true/false, short answer/completion, multiple choice, restricted response, and extended essay. Statistical analyses revealed statistically significant differences in testing preferences by test format, productive vs. recognition items, gender, and language groups. Furthermore, ESL students indicate that their testing preferences are largely consistent for their first and second languages. Practical implications for content-area faculty and future directions for research are provided.

There has been a dramatic increase in the number of English-as-a-second-language (ESL) students entering American colleges and universities (Erisman & Looney, 2007; Institute of International Education, 2009; Thomas, 2008). By 2050, Passel and Cohn (2008) estimate that 82% of U.S. population growth will be new immigrants and their descendants. With the passage of Dream Act legislation, such as the California Dream Act (2011), more culturally and linguistically diverse students will flow into post-secondary institutions.

Unfortunately, preparing teachers at all levels to work successfully with diverse students remains a national challenge (e.g., Sleeter, 2008; Steward, 1991; Suarez-Balcazar, Orellana-Damacela, Portillo, Rowan, & Andrews-Guillen, 2003). Assessment practices are of particular concern (e.g., Ewell, 2004; Solano-Flores & Trumbull, 2003). On the one hand, universities (Shavelson, 2009) and public schools (Nichols & Berliner, 2007) are under increasing pressure to provide evidence of student learning. On the other hand, the fairness and value of assessment practices and policies are being questioned (Schaeffer, 2013).

Because much of the literature in educational measurement and research addresses the reliability and validity of various examinations in differentiating, placing, or predicting success among students (e.g., Pitoniak, Yong, Martiniello, King, & Ginsburgh, 2009), the opinions, concerns, and preferences of test-takers themselves have often been ignored (Gellman & Berkowitz, 1993; Nield & Wintre, 1986; Zeidner, 1987). This avenue of research holds some promise in elucidating ESL student testing preferences for university faculty.

RELATED LITERATURE

Every person approaches a testing situation with personal assumptions, concerns, and expectations (Chapell et al., 2005; Pitoniak et al., 2009; Salamonson, Everett, Koch, Andrew, & Davidson, 2008). In this section, I briefly highlight research addressing native and non-native English speakers' perceptions related to test preferences. For clarity in this review of research, I use Gronlund's (1993) definitions of true/false, short answer/completion, multiple choice, restricted response, and extended essays. First, true/false test items are comprised of a declarative statement that a student must judge as true or false. Second, short-answer or completion test items require a student to provide only a word, number, or symbol that answers a question or completes a statement. Third, multiple-choice items have a stem, and more than two alternatives for students to choose among as answers. Fourth, restricted-response test items provide boundaries and limits to how a student answers a question; for example, a student will be asked to list, define, or give reasons. Finally, extended-essay test items allow students the most freedom in determining the structure and scope of their written responses to questions.

Native English Speaker Research

Several studies have focused on native English speakers' testing preferences. Nield and Wintre (1986) found university students most preferred taking tests with restricted-response questions, followed in order of preference by essay, multiple-choice, completion, and true/false questions. Zeidner's (1987) study with junior high school students revealed that "essay exams were perceived to be more fair than multiple-choice exams" because essay exams allowed students

“the opportunity of accurately and optimally expressing their knowledge and ideas in writing” (p. 355). Studies by Gellman and Berkowitz (1993) and Nield and Wintre (1986) also confirmed student perceptions that essay questions are fair and valid.

These studies also suggested that students preferred objective items (e.g., multiple choice) to subjective items (e.g., essay) for specific reasons. Junior high school students believed that they had “a better chance of succeeding on multiple-choice. . . exams” (Zeidner, 1987, p. 355). University students found objective items advantageous because “answers were provided (which allowed them to guess) . . . and required little mental or physical effort” (Nield & Wintre, 1986, p. 197). Gellman and Berkowitz (1993) found that students “do not necessarily prefer the measure that is perceived as most likely to demonstrate the extent of their knowledge, rather they prefer the type of test on which they believe it is easier to do well” (p. 18). Foos’ (1992) study with university students ($N = 84$) found that when students were told to expect a difficult and/or essay test, they performed better than when they were told to expect easy and/or a multiple-choice test.

Finally, gender is an important variable in understanding test-format preferences (Anderson, 1989). Gellman and Berkowitz (1993) found women strongly preferred essay items over multiple-choice items. A study by Bridgeman and Lewis (1994), using Advanced Placement Examinations, found that while essay scores were nearly equivalent for females and males, males significantly out scored females on multiple-choice sections.

Non-Native English Speaker Research

Several studies have focused on university ESL students’ test performance (e.g., Harklau, 1994; Horowitz, 1986; In’nami & Koizumi, 2009; Knoch & Elder, 2010; Leki, 1995; Leki & Carson, 1994). Certain test formats, such as essay or restricted-response questions, pose special problems for ESL students (Christe & O’Shea, 1988; Horowitz, 1986; Kinsella, 1992; Leki, 1995). For example, students must “find, organize, and present data according to fairly explicit instruction” (Horowitz, 1986, p. 455). In answering essay questions, Hayward (1990, p. 754) underscored the importance of both language (“vocabulary, syntax, organization, and rhetoric”) and content (“a grasp of the material under discussion”).

Objective test items, such as multiple-choice items, also pose difficulties for ESL students. Leki and Carson (1994) stated that ESL students require a “disproportionate amount of time” (p. 94) for just reading and selecting correct answers on multiple-choice items. Harklau (1994), on the other hand, found ESL students “proficient in bluffing their way through such mechanical exercises without a clear idea of what they were talking about” (p. 254).

More research is justified given the increasing concern over ESL student success in university content-area courses. From a test-taker perspective, perceived ease and difficulty, perceived potential for success, language group, and gender are all possible variables influencing test format preferences for ESL students. Three research questions guide this study:

1. What are ESL students’ preferences related to various test question formats (i.e., true/false, short answer/completion, multiple choice, restricted response, and extended essay)?
2. Is there a difference in how ESL students rate various test question formats?
3. Is there a relationship between language group, gender, and testing preferences?

METHOD

Using a nonexperimental approach, I collected quantitative data to describe university ESL students’ preferences in testing. Nonexperimental designs lack both treatment manipulation and randomization, seeking to explain “what is going on” rather than “what caused this,” (Hatch & Lazaraton, 1991, p. 100).

Subjects

Thirty-eight ESL students (females = 20; males = 18) were purposefully selected based on enrollment in two multicultural sections of a history course to provide quantitative data regarding their general preferences in testing. Subjects were divided into either romance ($n = 21$) or non-romance ($n = 17$) language groups. Spanish, French, Portuguese, and Italian comprised the romance language group; Arabic, Chinese, Japanese, Korean, Cambodian, Russian, Belorussian, Polish,

Czech, Dutch, and Norwegian were in the non-romance language group. The average length of residency was 3.56 years ($SD = 3.15$), and subjects had an average U.S. university experience of 1.64 years ($SD = 1.34$).

Testing Preference Questionnaire

I adapted Zeidner's (1987) semantic differential approach to rating test formats for this study. Subjects rated Gronlund's (1993) five test formats along a seven-point continuum anchored by the following 10 adjective pairs: difficult/easy, complicated/simple, unclear/clear, boring/interesting, tricky/straightforward, unfair/fair, worthless/valuable, low expectancy of success/high expectancy of success, high anxiety producing/low anxiety producing, and feeling uncomfortable with exam/feeling comfortable with exam. Pedhazur and Schmelkin (1991) noted that this type of scale allows investigation of directionality (difficult/easy) and intensity (how difficult/how easy) as well as relations among adjective pairs within and across test formats. The higher the subjects' ratings, the more favorable the test format is perceived. Subjects were also asked if their attitudes toward each of the five test formats were the same in English and their native language. If their opinion of a test format differed between their two languages, subjects were asked to briefly explain why they believed this difference existed in first and second language testing situations.

Data Analysis

The survey data were analyzed descriptively for frequencies, means (M), standard deviations (SD), zero-order correlations ($PPMr$), and plotted on scattergrams to check for linear relationships and influential outliers (e.g., using Cook's D). Multiple regression analyses were used to investigate the relationships among gender and language group (transformed through dummy-coding) and testing preferences. Because this is a non-experimental and exploratory study, regression coefficients (i.e., unstandardized coefficients) were evaluated using two criteria of significance. A strong correlation is defined at the $p < .05$ level. A weak, but potentially important, correlation is defined at $p < .10$ but $> .05$.

FINDINGS

ESL students were asked to assess five test formats using a seven-point semantic differential scale comprised of 10 adjective pairs (e.g., difficult/easy, unclear/clear, etc.). These data were analyzed using overall composite scores and individual scores by adjective pairs. Table 1 presents the frequency and percent of ESL students' format preferences when asked which question format allowed them to demonstrate content knowledge best. Among the five test formats considered, approximately one-third of the ESL students felt that multiple-choice tests were the best for demonstrating content knowledge; however, short-answer/ completion and extended-essay test formats were also more frequently preferred formats. ESL students preferred test formats requiring them to write/produce answers (production formats) over recognition formats, such as true/false and multiple choice, for demonstrating their content knowledge.

Table 1

Question Formats Student Feel Best Allow Them to Demonstrate Knowledge

Formats	Frequency	Percent
Multiple Choice	12	31.6
Short Answer/Completion	11	28.9
Extended Essay	9	23.7
Restricted Response	5	13.2
True/False	1	2.6
<hr/>		
Production Formats	25	65.8
Recognition Formats	13	34.2

For comparison, Table 2 presents means and standard deviations for composite scores derived from student ratings of each format based on the synthesis of the ten adjective pairs. The higher the mean score, the more favorable ESL students' disposition is toward the question format. Survey responses showed that ESL students were most favorably disposed to short- answer/completion items, followed by restricted-response and multiple-choice formats.

Table 2
Mean Composite Ratings of Question Formats on a Semantic Differential Scale

Format	<i>M</i>	<i>SD</i>
Short Answer/Completion	45.07	9.06
Restricted Response	43.38	9.56
Multiple Choice	42.11	12.48
True/False	41.55	8.61
Extended Essay	40.68	9.83
Production	43.04	7.43
Recognition	41.77	7.61

ESL students were least favorably disposed toward extended essays. This ranking of test formats differs from those presented in Table 1, when students merely selected categorically among the five formats. Nevertheless, question formats requiring students to write/produce answers were generally rated slightly more favorably by ESL students than recognition question formats, which is consistent with findings presented in Table 1.

An analysis of variance of the results presented in Table 2 was carried out, using a repeated-measures approach, to determine if students' mean score ratings of the five test formats differed significantly from each other. The omnibus *F*-test was significant, $F(41, 147) = 1.92, p < .003$, suggesting statistically significant mean differences in preferences among the five test formats. Nevertheless,

the Scheffé post hoc analysis was not significant: There was not sufficient statistical power (i.e., enough subjects) to pinpoint where the mean differences lie among these five test formats. An analysis of variance considering production and recognition format mean differences was not significant.

Table 3 summarizes ESL students' rating of each test format by adjective pair. Test formats are listed in order from least to most favorable. ESL students rated true/false and extended-essay formats more often at the extreme ends of the adjective scale. For example, ESL students rated true/false formats most negatively. Although they were comfortable with and considered true/false questions easy, they rated them as the most unclear, unfair, boring, worthless, and tricky format with a low expectancy of success.

The extended-essay format was also rated as an extreme for five of the adjective pairs. ESL students were most uncomfortable with the extended-essay format, rating it as the most difficult, complicated, and high-anxiety producing format. Nevertheless, ESL students also rated essay tests as the most valuable format alternative.

Short-answer/completion, restricted-response, and multiple-choice formats were rated more moderately. Nevertheless, short-answer/completion questions were preferred for being simple, clear, and providing a high expectancy of success. Multiple-choice questions were considered interesting and low-anxiety producing for ESL students. ESL students preferred restricted-response questions for being the most straightforward and fair test format.

Table 3*Mean Score Ratings of Test Formats by Adjective Pairs*

A. Difficult to Easy	<i>M</i>	<i>SD</i>
Extended Essay	3.27	1.47
Restricted Response	3.73	1.54
Short Answer/Completion	4.38	1.22
Multiple Choice	4.40	1.42
True/False	4.88	1.20
B. Complicated to Simple		
Extended Essay	3.37	1.43
Restricted Response	3.73	1.42
Multiple Choice	4.20	1.39
True/False	4.27	1.38
Short Answer/Completion	4.54	1.21
C. Unclear to Clear		
True/False	4.02	1.28
Multiple Choice	4.18	1.57
Extended Essay	4.49	1.24
Restricted Response	4.58	1.39
Short Answer/Completion	4.67	1.38
D. Boring to Interesting		
True/False	4.28	1.35
Extended Essay	4.42	1.51
Restricted Response	4.45	1.26
Short Answer/Completion	4.50	1.23
Multiple Choice	4.60	1.29
E. Tricky to Straightforward		
True/False	3.04	1.41
Multiple Choice	3.26	1.64
Extended Essay	4.53	1.21
Short Answer/Completion	4.71	1.50
Restricted Response	4.74	1.45

(Table 3 continued)

	<i>M</i>	<i>SD</i>
F. Unfair to Fair		
True/False	4.08	1.25
Multiple Choice	4.12	1.57
Extended Essay	4.30	1.42
Short Answer/Completion	4.56	1.59
Restricted Response	4.78	1.15
G. Worthless to Valuable		
True/False	4.23	1.34
Multiple Choice	4.44	1.50
Short Answer/Completion	4.76	1.22
Restricted Response	5.08	1.09
Extended Essay	5.15	1.19
H. Low to High Expectancy of Success		
True/False	4.29	1.22
Multiple Choice	4.40	1.59
Restricted Response	4.54	1.21
Extended Essay	4.54	1.45
Short Answer/Completion	4.65	1.07
I. High to Low Anxiety Producing		
Extended Essay	3.32	1.41
Restricted Response	3.76	1.36
True/False	4.14	1.39
Short Answer/Completion	4.15	1.17
Multiple Choice	4.17	1.56
J. Feeling Un/Comfortable with Exam		
Extended Essay	3.51	1.58
Restricted Response	3.98	1.38
Multiple Choice	4.28	1.73
Short Answer/Completion	4.29	1.30
True/False	4.38	1.42

Table 4 lists the means and standard deviations for production and recognition test formats by adjective pair. Overall, ESL students rated production test formats (i.e., short answer/completion, restricted response, and extended essays) more favorably than recognition formats (i.e., true/false and multiple choice). In particular, ESL students rated production items as being slightly more clear, interesting, straightforward, fair, valuable, and as holding higher expectancy of success than recognition formats. On the other hand, ESL students also rated recognition formats as easier, simpler, less anxiety producing, and as producing more comfortable feelings than production formats on exams

Table 4
Mean Scores for Production and Recognition Formats by Adjective Pairs

Adjective Pairs	Production		Recognition	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
A. Difficult/Easy	3.50	1.34	4.66	0.98
B. Complicated/Simple	3.88	0.96	4.23	1.05
C. Unclear/Clear	4.58	1.05	4.10	0.99
D. Boring/Interesting	4.46	1.04	4.42	1.05
E. Tricky/Straightforward	4.66	1.10	3.15	1.15
F. Unfair/Fair	4.54	1.02	4.12	1.13
G. Worthless/Valuable	5.00	0.83	4.33	1.23
H. Low/High Expectancy of Success	4.58	0.96	4.31	1.11
I. High/Low Anxiety Producing	3.74	0.97	4.14	1.09
J. Feeling Un/comfortable	3.93	1.13	4.30	1.14

Regarding testing preferences, students were asked to indicate whether their preferences in testing differed in their first and second languages. For true/false (60.5%), multiple choice (68.4%), short answer/completion (71.1%), restricted response (65.8%), and extended essay, a clear majority of ESL students felt their attitudes toward formats were the same in both languages. Only 36% indicated that their preferences would differ by format.

When students indicated a first/second language difference, they were asked to explain or give a reason for the difference. Overall, students most commonly identified vocabulary—wording, tricky words, or a limited vocabulary—as the factor that made their attitude toward a particular test format different in their first and second language. A student observed, “When American teachers try to be tricky with words, I get confused. No matter how good I am in English, I won’t be as good as native speakers.” One student shared: “Many times you’re asked for specific words. It may happen that you can’t remember that one word in English, but you do remember it in your native tongue.” Another student wrote, “I understand usually the concepts, but for this kind of question, everything is centered in a single word that you can easily forget.”

ESL students mentioned time limits, forgetting, memorizing, writing, grammar, spelling, appropriate style, and not being able to use dictionaries as some of the concerns they have with test formats that require writing or production. One student wrote that without a dictionary, “I cannot express myself clearly enough. . . . I may know the answer, but I just don’t know how to say/write that.” Another student observed, “I always worry about grammar and vocabulary when I have an essay test. Sometimes I have hard time expressing what I really want to say.”

Table 5 presents the relevant statistics for the regression of testing preferences on language group and gender. Together language group and gender account for 17% of the variance in students’ testing preferences, and were significantly associated to testing preferences.

Table 5
Multiple Regression Analysis of Testing Preferences on Language Group and Gender

Independent Variable	<i>B</i>	β eta	<i>t</i>
Language Group	-2.94	-0.28	-1.75*
Gender	3.30	0.32	2.00
$R^2 = .17, F(2,33) = 3.33^{**}$			
* $p < .10$, ** $p < .05$			

Using the $p < .10$ criterion, language group proved to be a weak regression coefficient in association to students' testing preferences in the present study, $r(35) = -.26, p = .12, r^2 = .07$. Non-romance language group students ($M = 43.73, SD = 4.32$) were more positive toward all testing formats on average than romance language group students ($M = 40.99, SD = 5.69$). In analyses of each of the five test formats, only short-answer/completion questions differed by language group, $r(37) = -.29, p < .05, r^2 = .08$. Non-romance language students viewed short-answer/completion questions more favorably ($M = 48.21, SD = 8.17$) than romance language students ($M = 42.54, SD = 9.13$).

As the beta coefficients indicate, gender proved to be more important than language group in understanding testing preferences, $r(35) = .30, p = .07, r^2 = .09$. Gender met the criterion for being a strong regression coefficient in this analysis. Females ($M = 43.70, SD = 5.39$) had a more favorable attitude toward all testing formats on average than males ($M = 40.56, SD = 4.80$). Analyses of format types also revealed that short-answer/completion formats were viewed much more favorably by females ($M = 49.02, SD = 7.35$) than males ($M = 41.53, SD = 9.14$), $r(37) = -.29, p < .05, r^2 = .08$.

Production and recognition test format preferences were analyzed by language group and gender using composite scores. Students' preferences for recognition test formats did not correlate significantly with language groups or gender; however, preference for production test formats did. Table 6 presents the regression statistics. Together language group and gender account for 20% of the variance in preference for production formats. Females ($M = 45.37, SD = 7.06$) were more favorable toward write/produce answers than males ($M = 40.95, SD = 7.30$), $r(37) = .30, p < .05, r^2 = .09$. ESL students in the non-romance language group ($M = 45.41, SD = 7.12$) preferred production test formats more than students in the romance language group ($M = 41.13, SD = 7.28$), $r(37) = -.28, p < .10, r^2 = .08$.

Table 6*Multiple Regression Analysis of Production Format Preferences on Language and Gender*

Independent Variable	<i>B</i>	β eta	<i>t</i>
Language Group	-4.83	-0.33	-2.15*
Gender	4.95	0.34	2.21*
$R^2 = .20, F(2, 35) = 4.28^*$			
* $p < .05$			

DISCUSSION AND IMPLICATIONS

The survey results of this descriptive study document university level ESL students' preferences in testing. I summarize four main findings in light of the literature and present implications for practice and research.

First, in demonstrating their content knowledge, ESL students generally believe that they can demonstrate their content-knowledge on tests better with test formats that require answer production over answer recognition. Native speakers also prefer restricted-response and essay items over multiple-choice formats on the criteria of "value, fairness, and validity in assessing content knowledge" (Zeidner, 1987, p. 357). Nevertheless, ESL students consider recognition formats easier, simpler, more comfortable, and the least anxiety producing. These findings support native English speaker research finding it is easier to do better--score higher--on recognition formats (i.e., Foos, 1992; Gellman & Berkowitz, 1993; Nield & Wintre, 1986). As Foos (1992, p. 209) summarized, students believe that "any test, even an essay test, could be easy but that a multiple-choice test can never be very difficult." ESL studies by Leki and Carson (1994), Horowitz (1986), Kinsella (1992), and Leki (1995) described writing as an obstacle for production items and time limits as a problem for recognition items, shedding light on how ESL students can hold preferences for both multiple-choice and production items simultaneously.

University faculty members often consider test-taking strategies an a priori skill set possessed by students, ignoring important cultural, linguistic, and experiential considerations in testing for diverse

students. Several implications for improving assessment practices for ESL students in content-area classes should be considered. As Kinsella (1992) argued, ESL students need strategies for coping with both production and recognition test questions. Developing format-specific test-taking strategies should be a priority, not an elective, for university-bound ESL students. Therefore, university ESL faculty should consider if they go far enough in acclimating ESL students to the demands of university testing. Content-area university faculty should intentionally prepare tests with a combination of production and recognition test items to allow students alternatives for demonstrating the extent of their content knowledge. Careful consideration should be given to response time allowed, the balance of items, and the weighting of responses in grading for production and recognition test items. In this way, university faculty can begin to acknowledge that different test formats require different language skills and time for cognitive processing without disadvantaging their ESL students (Pitoniak et al., 2009).

Second, this study's use of a semantic differential scale to rate Gronlund's five test item formats was a new contribution. Composite and individual ratings of the five test formats show ESL students as most positively disposed toward short-answer/ completion formats, followed in order of preference by restricted-response, multiple-choice, true/false, and extended-essay formats. Although more research is needed, it appears that native speakers a) do not rate short-answer/ completion items as positively, or b) dislike essay formats as much as ESL students (e.g., Nield & Winter, 1986). Perhaps, these ESL testing preferences reflect how students are used to being tested in English; namely, short answer/completion items are typical in many English language textbooks, as well as on teacher-constructed and standardized language tests.

Overall, the differences in test format preferences were statistically significant; however, in future research more subjects—and statistical power—will be necessary to delineate exactly where those preference differences lie among test formats.

For many ESL students, their native linguistic, cultural, and educational knowledge about testing does not match the assessment practices they face in American university settings (Teemant, 2010). If and to what degree this experiential disparity actually impacts the ability of ESL students to accurately represent their content-area

learning on tests has yet to be quantified. Pitoniak et al., (2009) argue ESL students:

should have not only multiple opportunities, but also multiple ways to show what they know, and that assessment specifications should include a variety of item and response types that may lead to assessments on which ELLs are more likely to be able to show their strengths. (p. 10)

Third, results show gender differences in ESL students' test format preferences. Among these ESL students, females preferred production test formats more and were more positive about all test forms than males. Gender and testing research with native speakers show similar gender differences (e.g., Anderson, 1989; Chapell et al., 2005; Gellman & Berkowitz, 1993). Future studies could explore whether ESL students match native speaker patterns with males outscoring females on multiple-choice (Bridgeman & Lewis, 1994) or females outscoring males on essay exams (Gellman & Berkowitz, 1993).

Fourth, findings also uniquely suggest a slight language group difference in testing preferences, with non-romance students being more positive toward all test formats than romance language students. Further research is needed to clarify the nature of language group differences. Nevertheless, ESL students indicated that their attitudes toward test formats were the same in their first and second languages. Of those students who felt their attitudes differed across languages, the most common reason given by students for that difference was a limited vocabulary. Time limits, writing expectations, and dictionary use were other reasons for viewing various test formats differently in their L1 and L2. These findings are consistent with the ESL literature suggesting that problems in testing arise from a lack of appropriate accommodation by teachers (e.g., Kinsella, 1992; Pitoniak et al., 2009). The findings suggest testing preferences are influenced by language proficiency as well as judgments regarding how test formats allow individual students to demonstrate their content knowledge. University faculty would benefit from learning about how and when to make appropriate language accommodations in content-area testing to promote "equity and validity in assessment" (Pitoniak et al., 2009, p. 22).

CONCLUSION

Understanding how university ESL students make sense of accurately and consistently representing their content knowledge on tests is a second language issue worthy of systematic investigation. The present study contributed to understanding testing from the ESL perspective by providing descriptive evidence of ESL student testing preferences by test format, gender, and language group. With changing student demographics, university faculty will increasingly be charged with understanding content-area testing from a second language perspective.

REFERENCES

- Anderson, J. (1989). Sex-related differences on objective tests. *Educational Studies in Mathematics*, 20(2), 165 - 177.
- Banerji, S. (2007, April 12). Give legal immigrants same opportunities in higher education, says report. Retrieved from <http://diverseeducation.com/article/7226/>.
- Bridgeman, B., & Lewis, C. (1994). The relationship of essay and multiple-choice scores with grades in college courses. *Journal of Educational Measurement*, 31(1), 37-50.
- California Dream Act of 2011, AB 130, Section 68130.7 and AB 231, Section 66021.7 to the Education Code.
- Chapell, M. S., Blanding, Z. B., Silverstein, M. E., Takahasi, M., Newman, B., Gubi, A., &
- McCann, N. (2005). Test and anxiety and academic performance in undergraduate and graduate students. *Journal of Educational Psychology*, 97(2), 268-274.
- Chiste, K. B., & O'Shea, J. (1988). Patterns of question selection and writing performance of ESL students. *TESOL Quarterly*, 22(4), 681-684.
- Erisman, W., & Looney, S. (2007, April). Opening the door to the American dream: Increasing higher education access and success for immigrants. Retrieved from <http://www.ihep.org/assets/files/publications/m-r/OpeningTheDoor.pdf>.
- Ewell, P. (2004). General education and the assessment reform agenda. Washington, DC: Association of American Colleges and Universities.

- Foos, P. W. (1992). Test performance as a function of expected form and difficulty. *Journal of Experimental Education*, 60(3), 205-211.
- Gellman, E. S., & Berkowitz, M. (1993). Test-Item type: What students prefer and why. *College Student Journal*, 27(1), 17-26.
- Gronlund, N. E. (1003). *How to make achievement tests and assessments* (5th ed.). Boston: Allyn and Bacon.
- Harklau, L. (1994). ESL verses mainstream classes: Contrasting L2 learning environments. *TESOL Quarterly*, 28(2), 241-272.
- Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston, MA: Heile & Heinle Publishers
- Hayward, M. (1990). Evaluations of essay prompts by nonnative speakers of English. *TESOL Quarterly*, 24(4), 753-760.
- Horowitz, D. M. (1986). What professors actually require: Academic tasks for the ESL classroom. *TESOL Quarterly*, 20(3), 445-462.
- In'nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing*, 26(2), 219-244.
- Institute of International Education. (2009, November 16). Record number of international students in U.S. higher education. Retrieved from <http://opendoors.iienetwork.org/?p=150649>
- Kinsella, K. (1992). How can we encourage active learning strategies in content-based second language instruction? *CATESOL Journal*, 5(1), 127-138.
- Knoch, U., & Elder, C. (2010). Validity and fairness implications of varying time conditions on a diagnostic test of academic English writing proficiency. *Systems*, 38(1), 63-74.
- Leki, I., & Carson, J. G. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly*, 28(1), 81-101.
- Leki, I. (1995). Coping strategies of ESL students in writing tasks across the curriculum. *TESOL Quarterly*, 29(2), 235-260.
- Nield, A. F., & Wintre, M. G. (1986). Multiple choice questions with an option to comment. *Teaching of Psychology*, 13(4), 196-199.
- Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.

- Passel, J., & Cohn, D. (2008). U.S. population projections: 2005–2050. Washington, DC: Pew Research Center. Retrieved February 22, 2008, from <http://pewhispanic.org/files/reports/85.pdf>
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Pitoniak, M. j., Yong, J. W., Martiniello, T. C., King, A. B., & Ginsburgh, M. (2009). *Guidelines for the assessment of English language learners*. Retrieved from www.ets.org/ellguidelines.
- Salamonson, Y., Everett, B., Koch, J., Andrew, S., and Davidson, P. M. (2008). English-language acculturation predicts academic performance in nursing students who speak English as a second language. *Research in Nursing and Health*, 31(1), 86-94.
- Schaeffer, B. (2013, August 21). New polls show “Americans fed up with school-test overkill.” The National Center for Fair and Open Testing. Retrieved from <http://www.fairtest.org>.
- Shavelson, R. J. (2009). *A brief history of student learning assessment*. Washington, DC: Association of American Colleges and Universities.
- Sleeter, C. E. (2008). Preparing White teachers for diverse students. In M. Cochran-Smith, S.
- Feiman-Nemser, D. J. McIntyre, & K. E. Demers (Eds.), *Handbook of research on teacher education: Enduring questions in changing contexts* (3rd ed.) (pp. 559-582). New York: Routledge, Taylor & Francis Group and the Association of Teacher Education.
- Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher*, 32(2), 3-13.
- Steward, D. W. (1991). Immigration and higher education: The crisis and the opportunities. *Educational Record*, 72(1), 20-26.
- Suarez-Balcazar, Y., Orellana-Damacela, L., Portillo, N., Rowan, J.M., Andrews-Guillen, C. (2003). Experiences of Differential Treatment among College Students of Color. *Journal of Higher Education* 74(4), 428-444.
- Teemant, A. (2010, November). ESL student perspectives on university classroom testing practices. *Journal of the Scholarship of Teaching and Learning*, 10(3), 89-105.

- Thomas, J. (2008). International student enrollment at U.S. colleges breaks records. Retrieved from www.america.gov/st/.../200811171600491CJsamohT0.646908.html
- Zeidner, M. (1987). Essay versus multiple-choice type classroom exams: The student's perspective. *Journal of Educational Research*, 80(6), 352-358.

ABOUT THE AUTHOR

Annala Teemant is an Associate Professor of Second Language Education (Ph.D., Ohio State University, 1997) at Indiana University-Purdue University Indianapolis. She is actively engaged in developing, implementing, and researching quality teacher preparation for mainstream teachers of English Language Learners. She has been awarded four U.S. Department of Education grants focused on teacher quality and ESL teacher preparation. Inquiries should be directed to ateemant@iupui.edu.