Journal of Second Language Teaching and Research. Volume 4 Issue 1

Using Pre/Post-Testing to Evaluate the Effectiveness of Online Language Programs

Angela Felix, Department of Languages, Rio Salado College

Abstract

Learners are often frustrated by their perceived lack of progress in second language courses, language teachers are frequently tasked with the responsibility of providing evidence that learning is really happening, and institutions of higher education in the United States are under increasing pressure to justify the continuation of "foreign" language programs. In answering the call for more increased transparency, this paper presents assessment data for students enrolled in the first four semesters of language study in Spanish, French, German, Japanese, Mandarin Chinese, Arabic, and American Sign Language in a postsecondary institution in the United States. This comprehensive approach to program assessment incorporates online pre- and post-testing as a direct measure for learning that supports accountability at the student, course, program, and college level.

Keywords: pre/post-testing; online learning; assessment; accountability

Introduction

This paper details the results of a multi-year project to assess for learning across seven languages taught online at a public two-year college in the southwestern United States. It is important to note that two-year public colleges in the US typically have no second language placement testing requirements for college admission. Thus, first semester (101-level) courses are often populated by "false beginners" -- students that have prior knowledge of the language. This environment is similar to that of the Open University, which is the largest higher education language provider in the United Kingdom. Coleman and Furnborough's research conducted at the Open University discovered that two-thirds of the beginning Spanish students in their study had previous experience with the language (2010, p. 13). The "prior knowledge" phenomenon is one of the reasons that stand-alone tests of language proficiency and final grades are not direct measures of learning. Heritage language speakers may earn exemplary grades in beginning classes, which may or may not indicate that any learning took place. By pre-testing all of the students, a baseline can be established for comparison with post-test results. Improvements in performance between the pre- and post-test indicate an increase in learning. Admittedly, pre-/post-testing methodology has inherent elements of unreliability beyond just the prior-knowledge phenomenon, including improvements in test scores that are a result of maturation rather than classroom learning and pre-test scores that are so high as to make statements about increases in learning practically impossible, that limit this study. However, other variables typically associated with pre-post testing, such as variations in test administration and higher post-test scores due to lower-performing students that drop out, have been controlled.

Effective assessment methods and strategies are vital components of successful educational programs. In fact, "[t]he apparently ever growing emphasis on evidence of learners' achievements in tests and examinations demonstrates that assessment is seen outside as well as within the educational context as critical in judging both learners' and teachers' performance" (Hunt, Neill, & Barnes, 2007, p. 195). Given the increasing importance placed on transparency in education, language assessment must extend beyond the classroom to provide evidence of instructor, course, and program effectiveness. Pre/post-testing is just one measure that, combined with other assessments such as peer evaluation and program review, can present authentic and holistic data that more accurately reflects the true educational experience.

One approach to this new reality is a focus on assessing *for* learning rather than traditional assessments *of* learning. Assessing *for* learning incorporates emerging trends in language assessment practice including measuring performance against established benchmarks [criterion-referenced testing] instead of student-to-student comparisons [norm-referenced testing], basing assumptions on multiple sources of information over singular measures, and moving from the concept of conclusive assessment to continual assessment (Phillips, 2006).

Building upon the 2002 research conducted by Brooks on the importance of assessing *for* learning, Barnes and Hunt underscore the significance of providing specific feedback and targets for language learners. They advocate the use of self-assessments so language students know what they are supposed to be learning as well as what they need to do to do to meet the objectives: "Such assessment can achieve excellent results in improving pupils' learning, which then feeds through to the summative assessment of which a wider public is aware and by which pupils and teachers tend to be judged" (Hunt et al., 2007, p. 207).

Engaging in assessment *for* learning rather than *of* learning can be an expensive, complex and uncomfortable process. Assessing for learning exposes the strengths and weakness of instructional materials, instructor feedback, assessment methods and program effectiveness.

Assessing for learning cannot be accomplished in a single initiative at a given point in time. It is a process that requires reflection, collaboration and a willingness to engage in ongoing cycles of continuous improvement.

Contemporary research in language learning and technology tends to focus on the integration of multimedia and/or online instructional modules to supplement a hybrid or traditional face-to-face classroom. Bahrani (2011), Swanson and Nolde (2011), and others have explored various technology tools that can be used to create reliable, valid and authentic language activities and assessments. Though benefits and drawbacks of incorporating such tools (cost, accessibility, scalability, etc.) are discussed, such studies do not evaluate the direct impact of technology on the achievement of learning objectives.

Carr, Crocco, Eyring, & Gallego (2011) utilized pre and post surveys of students and instructors/tutors to assess perceptions of technology-enhanced language learning. Their results showed an increase in comfort, enjoyment, and confidence in using technology from the beginning to the end of the course. As the research questions focused on student and instructor/tutor perceptions, direct measures of learning were not included to evaluate the instructional effectiveness of the technology-enhanced curriculum.

Palalas's 2011 study of English language learners combines research on student perception of a hybrid course with pre and post-tests to measure language proficiency. The results indicated a high level of satisfaction with the format combining face to face instruction with online and mobile deliveries. Though the sample was quite small (n=12), all participants improved their listening skills, and all but two of the participants improved their writing scores. The post-tests indicated an overall decrease in the students' reading proficiency. Though reading comprehension was not a learning objective in this particular course, reading skills were included in the pre and post assessments.

A review of current literature addressing online and technology-enhanced language learning sheds light on many innovative pedagogical approaches, but there is little published research that includes empirical evidence of the effectiveness of entirely online programs in terms of student achievement. By implementing systematic, continuous and large-scale pre and post-testing across an online language curriculum, this study provides evidence of assessing *for* learning at the course, language, and program level.

Method

This paper includes two years (six semesters) of data in which 4,061 pre/post-tests were administered to students registered in first and second-year online American Sign Language, Arabic, Chinese, French, German, Japanese and Spanish courses. With the exception of the video components of the American Sign Language assessments, all of the pre- and post-tests included oral and written items designed to assess the stated course competencies, which are based on the proficiency guidelines of the American Council on the Teaching of Foreign Languages (ACTFL, 2012). Item difficulty ranged from Novice Low for students enrolled at the 101-level through Intermediate High for 202-level students, which, according to the work of Goldfield (2010), is roughly equivalent to the A1-B1 levels as measured by the Common European Framework of Reference for Languages (Council of Europe, 2001).

Before engaging in the course content, all online language students were required to submit the corresponding pre-test, which, with the exception of American Sign Language, included 20 multiple-choice items assessing grammar and vocabulary knowledge, 5 questions assessing listening comprehension, and 5 free oral-response questions. The American Sign Language test consisted of 25 questions that assessed knowledge of grammar structure and Deaf culture. Students received credit for completing the pre-test, though the actual score was not recorded in the grade book. Students were not provided with the correct pre-test answers after submittal. However, instructors had access to the student's responses, and could use the information to personalize future communication and feedback. Students took the same assessments as post-tests one week before the final exam. Again, students received credit for completing the post-tests, though the actual scores did not affect their grades. The post-tests were utilized as formative assessments, and the results were provided to students so they could self-assess and focus their studies on concepts they still needed to master before taking the final exam. Instructors could also review the post-test results to provide individualized instructional interventions before the exam. These online pre- and post-tests are examples of “assessment artifacts” that are critical for evaluation, as such artifacts can be archived and referenced later for analyses (Swanson & Nolde, 2011, p. 74).

Though two years of pre/post data are included in this paper, the process of validating the assessment instruments began as early as 2005. Faculty members in each language systematically reviewed and revised test questions based on the results of annual item analysis reports. Once the faculty members were confident that the test items did, in fact, assess the specific skills that they were attempting to measure, the continuous improvement cycle for instructional materials began. All of the online courses included in this study are a "one course, many sections" model, which facilitates the implementation of programmatic interventions and data collection. Once a pre/post-test item was judged to be valid, then an instructional intervention designed to increase learning as measured by that item was deployed across all sections of a course. Post-test results were collected to evaluate the effectiveness of the new module or strategy. Some interventions resulted in an increase of learning and were retained, some actually were deleted due to a decrease in learning, and others had no effect. The interventions that did not produce an effect were modified an incorporated in the next improvement cycle. The initial target for all languages was a 70% post-test average for each course level (101, 102, 201 and 202).

The sections that follow include data and analyses for each of the languages included in this study. Please note that the results do not track longitudinal individual student performance over four semesters of language study. Rather, the data represent a snapshot of anonymous aggregate pre- and post-test results separately for each level throughout the study period. In each case, N represents ending enrollment. N does not include post-test results for students that dropped the course. Numerical test score averages are displayed below the N for each cohort, and they are displayed graphically along the vertical axis.

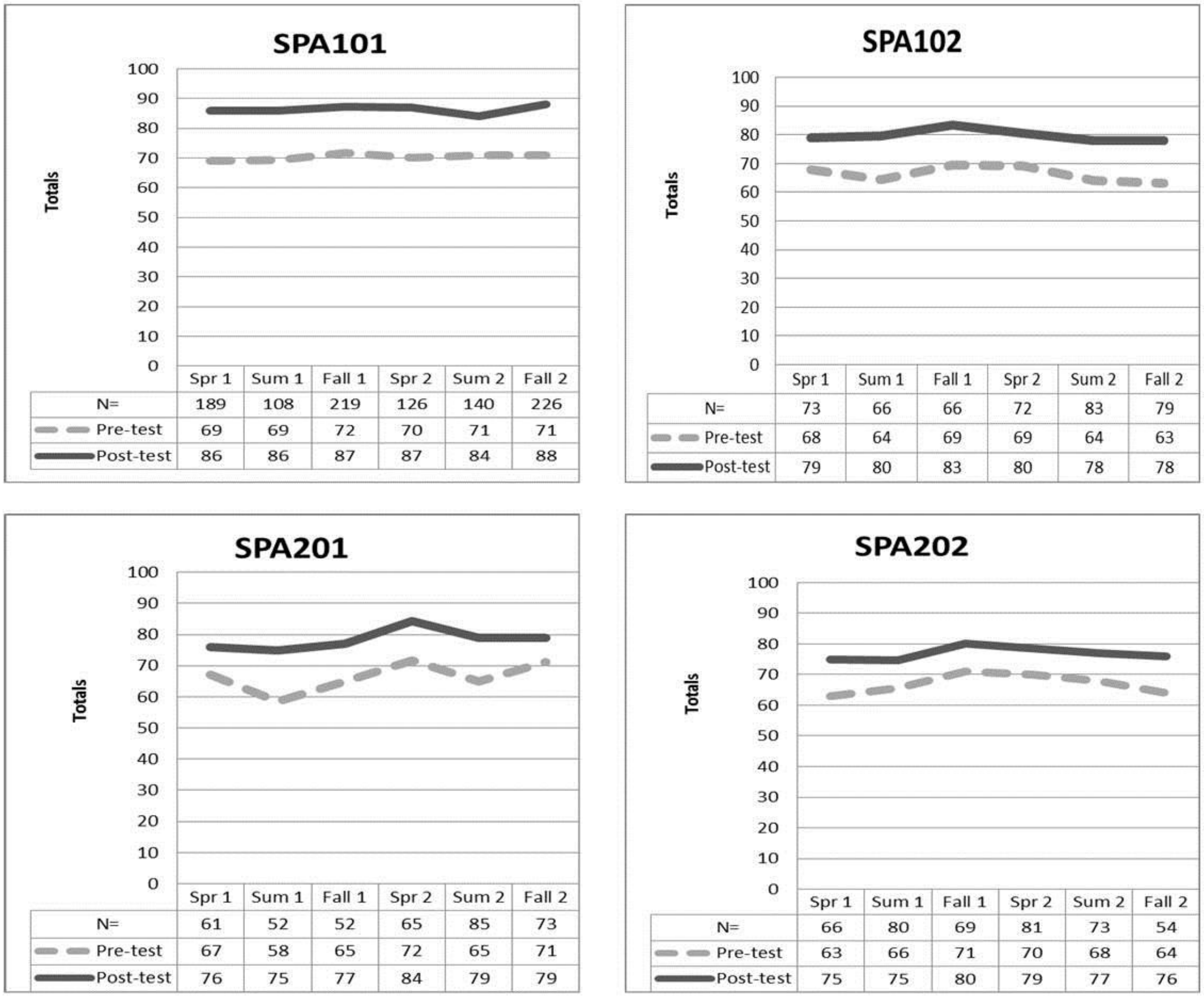
Data and Analysis

Spanish

Spanish (SPA) assessments are over-represented in the study, with 2,335 enrollments in first semester (SPA101), second semester (SPA102) third semester (SPA201) and fourth semester (SPA202) courses. This represents 57% of the total number of pre/post-tests submitted for all languages during the study period. A Spanish curriculum has been available online since 1997, and is the most mature of the language programs offered at the college.

SPA101 assessment data show unusually high average pre-test scores of over 70% (see Figure 1). This can be attributed to the college's location in the southwestern United States. Spanish is not only taught in high schools, but a significant population in the surrounding area consists of heritage Spanish speakers. Even though the data suggests that many learners are entering a Spanish 101 course with prior knowledge of the language, the post-test results do indicate that learning occurred over the semester, as average post-test scores approached 87%. A similar pattern can be observed for the second, third and fourth semester courses (SPA102, 201 and 202), though none produced average pre/post-test results as high as the first semester course.

Figure 1. *Average pre- and post-test scores for Spanish 101-202*



SPA101

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	189	108	219	126	140	226
▬ ▬ Pre-test	69	69	72	70	71	71
▬ Post-test	86	86	87	87	84	88

SPA102

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	73	66	66	72	83	79
▬ ▬ Pre-test	68	64	69	69	64	63
▬ Post-test	79	80	83	80	78	78

SPA201

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	61	52	52	65	85	73
▬ ▬ Pre-test	67	58	65	72	65	71
▬ Post-test	76	75	77	84	79	79

SPA202

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	66	80	69	81	73	54
▬ ▬ Pre-test	63	66	71	70	68	64
▬ Post-test	75	75	80	79	77	76

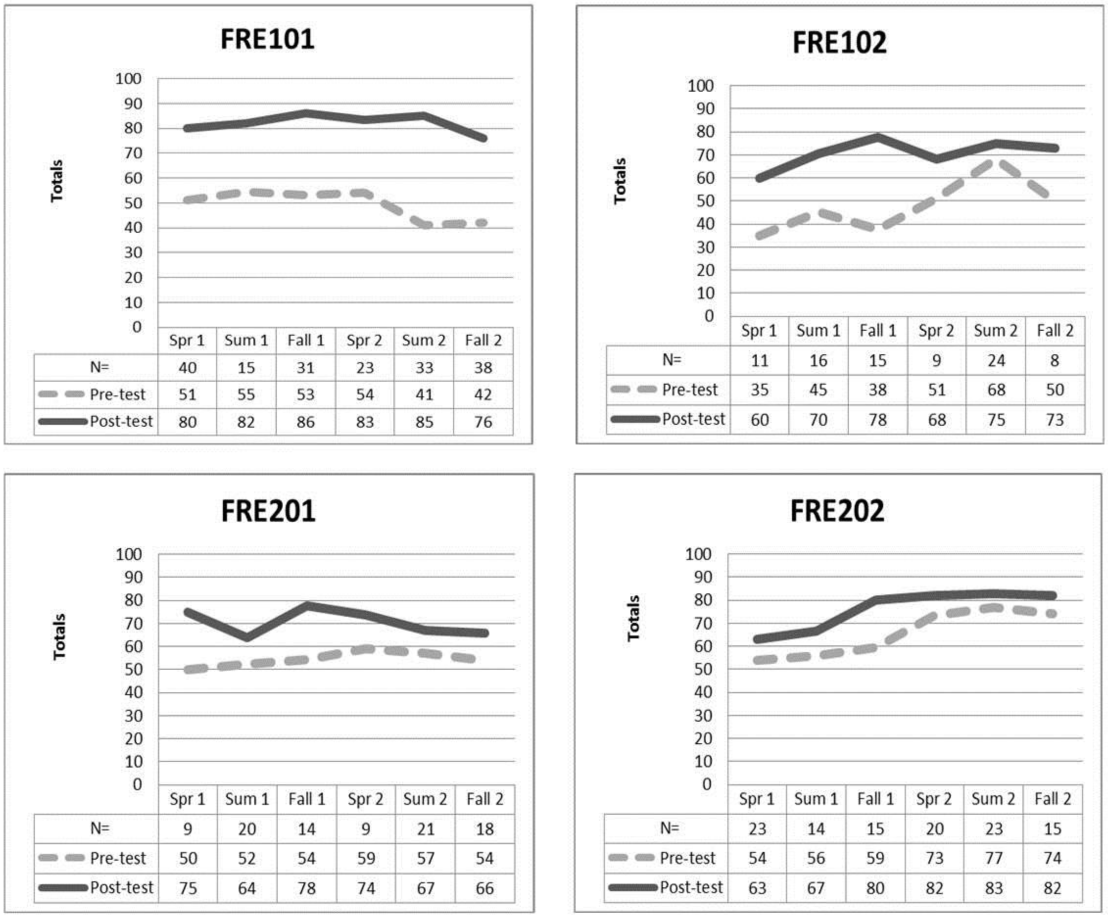
French

French (FRE) courses have been available online since 2005, and a total of 494 French pre- and post-tests were submitted in the two-year period.

As expected, French 101 pre-test scores were significantly lower than those of Spanish 101 (see Figures 1 and 2). Though French is taught in several surrounding high schools, it is not as widespread as Spanish. The average FRE101 pre-test score of approximately 52% likely also reflects a lack of heritage French speakers in the area. The post-test average of almost 82% exceeds the target of 70%, and is especially notable given the apparent lower level of proficiency of incoming beginning French students as compared to their counterparts in beginning Spanish.

Post-test results for FRE102 and 202 also met the target. However, average results for FRE201 fell short (see Figure 1). Instructional interventions incorporated in summer 1did not have a demonstrable effect on student learning in third-semester French, and lesson content is again being modified for a deployment across all sections of FRE201.

Figure 2. *Average pre- and post-test scores for French 101-202*



FRE101

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	40	15	31	23	33	38
Pre-test	51	55	53	54	41	42
Post-test	80	82	86	83	85	76

FRE102

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	11	16	15	9	24	8
Pre-test	35	45	38	51	68	50
Post-test	60	70	78	68	75	73

FRE201

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	9	20	14	9	21	18
Pre-test	50	52	54	59	57	54
Post-test	75	64	78	74	67	66

FRE202

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	23	14	15	20	23	15
Pre-test	54	56	59	73	77	74
Post-test	63	67	80	82	83	82

German

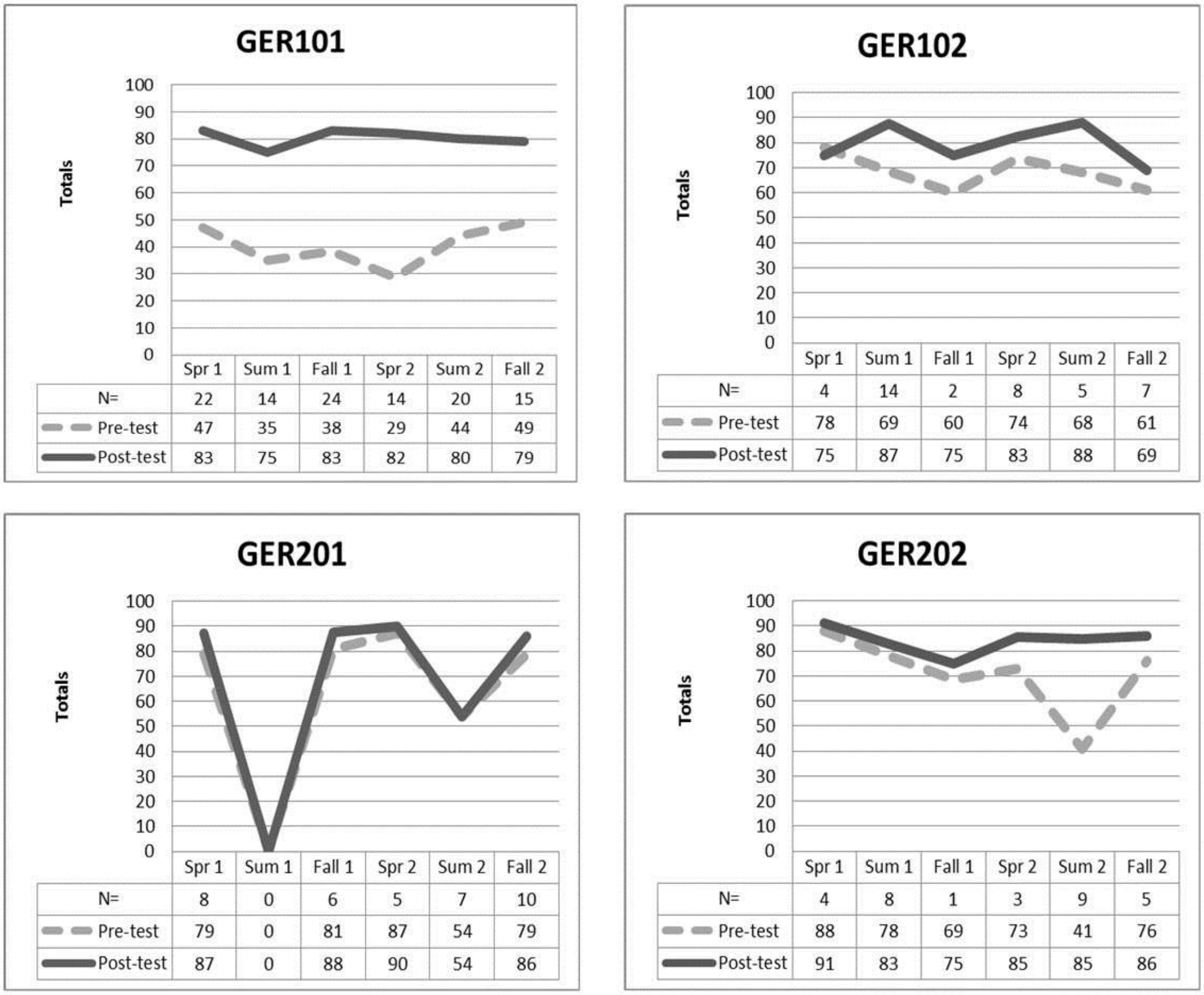
Though German (GER) courses have been available online since 2003, enrollments for levels beyond the first semester are quite low. In total, 234 German pre- and post-tests were completed during the study period.

The average pre-test score for GER101 (39%) was even lower than that of FRE101 (52%). This can be attributed to even fewer students entering GER101 with prior knowledge of the language. Not only is German is consistently offered at just one high school in the area, but there are no sizable heritage German-speaking populations in the surrounding community. Though GER101 pre-test results are significantly lower than those for FRE101, the average post-test results indicate similar levels of achievement (82% for FRE101 and 79% for GER101).

The gap between the average pre-test result and the average post-test results indicate a substantial increase in student learning (see Figure 3).

Though enrollment numbers are low for GER102, 201 and 202, it is interesting to note the high average pre-test score for each level. There is not a wide gap between those scores and the post-test results. With pre-test averages so high, there is not much room to show significant improvement. All of the average post-test scores meet the target of 70%. In fact, with the exception of GER101, the target was met with the pre-test results as well.

Figure 3. *Average pre- and post-test scores for German 101-202*



GER101

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	22	14	24	14	20	15
Pre-test	47	35	38	29	44	49
Post-test	83	75	83	82	80	79

GER102

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	4	14	2	8	5	7
Pre-test	78	69	60	74	68	61
Post-test	75	87	75	83	88	69

GER201

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	8	0	6	5	7	10
Pre-test	79	0	81	87	54	79
Post-test	87	0	88	90	54	86

GER202

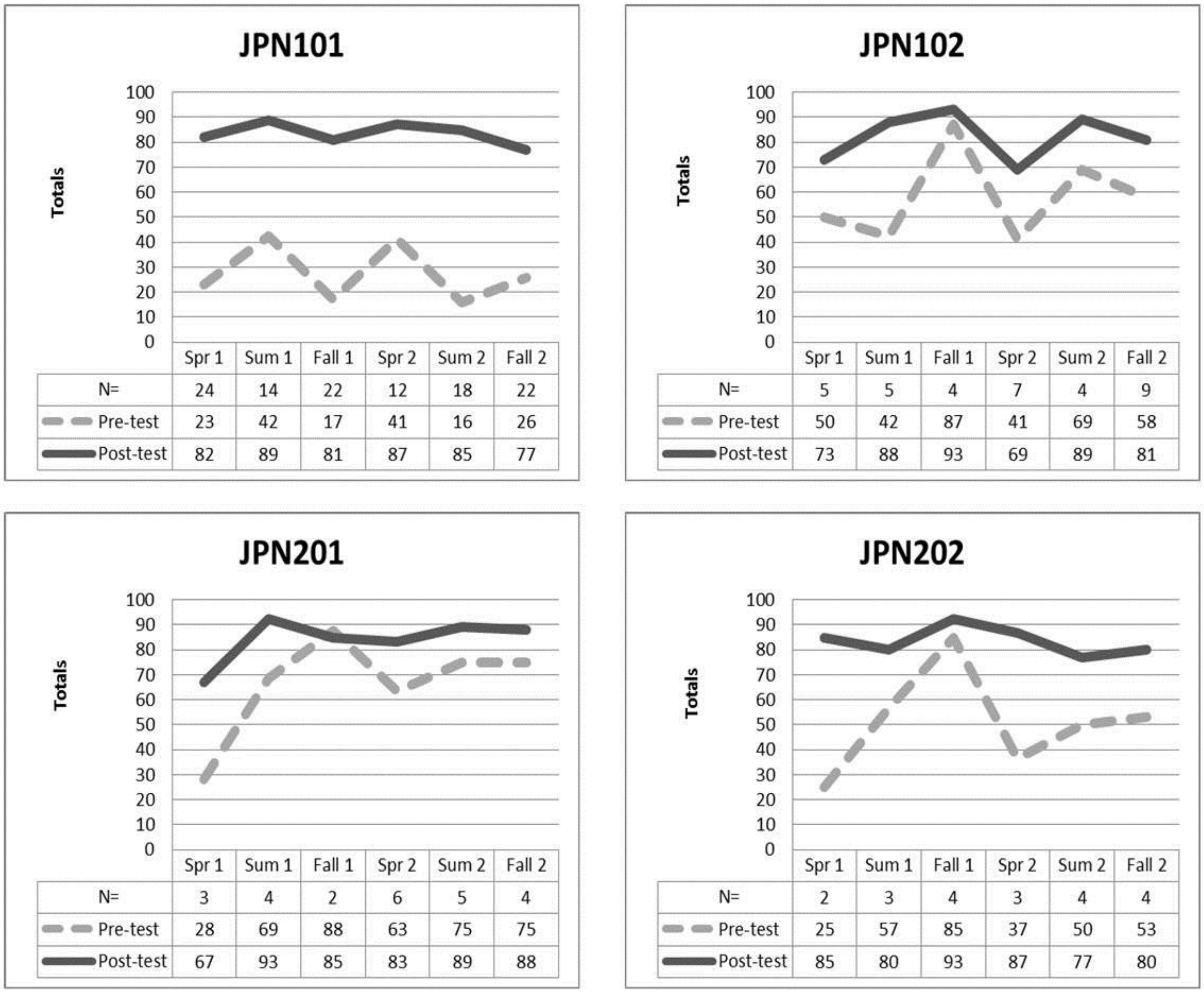
	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	4	8	1	3	9	5
Pre-test	88	78	69	73	41	76
Post-test	91	83	75	85	85	86

Japanese

Japanese (JPN) courses have been available since 2007, and 196 students took the pre- and post-test during the two-year study period.

Average pre-test scores for JPN101 were quite low (just under 28%). However, the post-test scores were remarkably high (over 85%), indicating a dramatic increase in learning over the length of the course (see Figure 4). As was the case with German, enrollments in JPN102, 201 and 202 are much lower than in the beginning-level course. Though average post-test scores consistently exceeded the target, there are dramatic fluctuations in the pre-test data points for the higher levels due to the low number of enrollees in those courses.

Figure 4. *Average pre- and post-test scores for Japanese 101-202*



JPN101

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	24	14	22	12	18	22
Pre-test	23	42	17	41	16	26
Post-test	82	89	81	87	85	77

JPN102

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	5	5	4	7	4	9
Pre-test	50	42	87	41	69	58
Post-test	73	88	93	69	89	81

JPN201

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	3	4	2	6	5	4
Pre-test	28	69	88	63	75	75
Post-test	67	93	85	83	89	88

JPN202

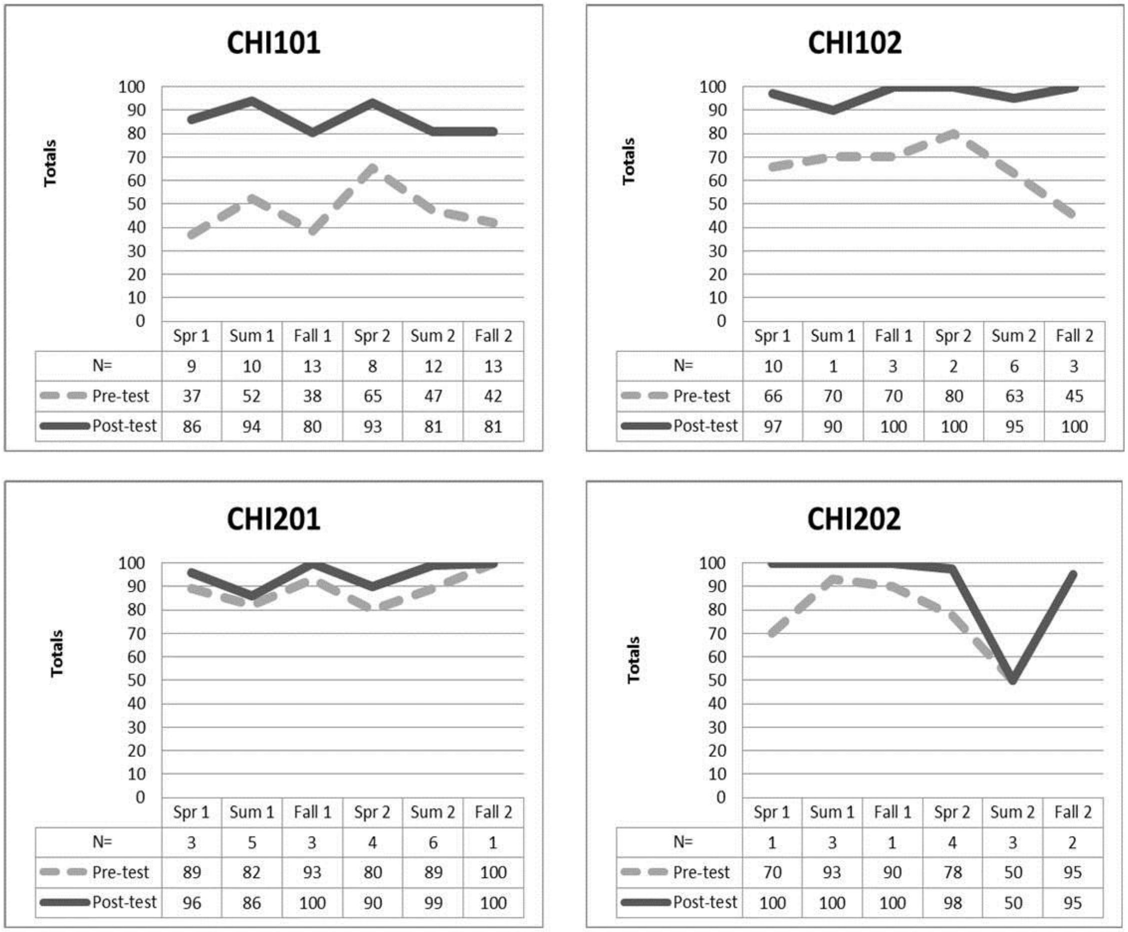
	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	2	3	4	3	4	4
Pre-test	25	57	85	37	50	53
Post-test	85	80	93	87	77	80

Mandarin Chinese

A Mandarin Chinese (CHI) curriculum has been offered online since 2007, and 128 pre- and post-tests were submitted over the duration of the study. Enrollments in all of the Mandarin Chinese courses are the lowest of the languages presented so far, with only twelve CHI202 post-tests being completed during this period.

As expected, the largest gap between average pre- and post-test scores was at the 101 level, which indicates the greatest increase in student learning (see Figure 5). Though the gap narrows for the subsequent levels, average post-test scores for all levels are the highest of the seven languages included in this study.

Figure 5. *Average pre- and post-test scores for Mandarin Chinese 101-202*



CHI101

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	9	10	13	8	12	13
Pre-test	37	52	38	65	47	42
Post-test	86	94	80	93	81	81

CHI102

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	10	1	3	2	6	3
Pre-test	66	70	70	80	63	45
Post-test	97	90	100	100	95	100

CHI201

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	3	5	3	4	6	1
Pre-test	89	82	93	80	89	100
Post-test	96	86	100	90	99	100

CHI202

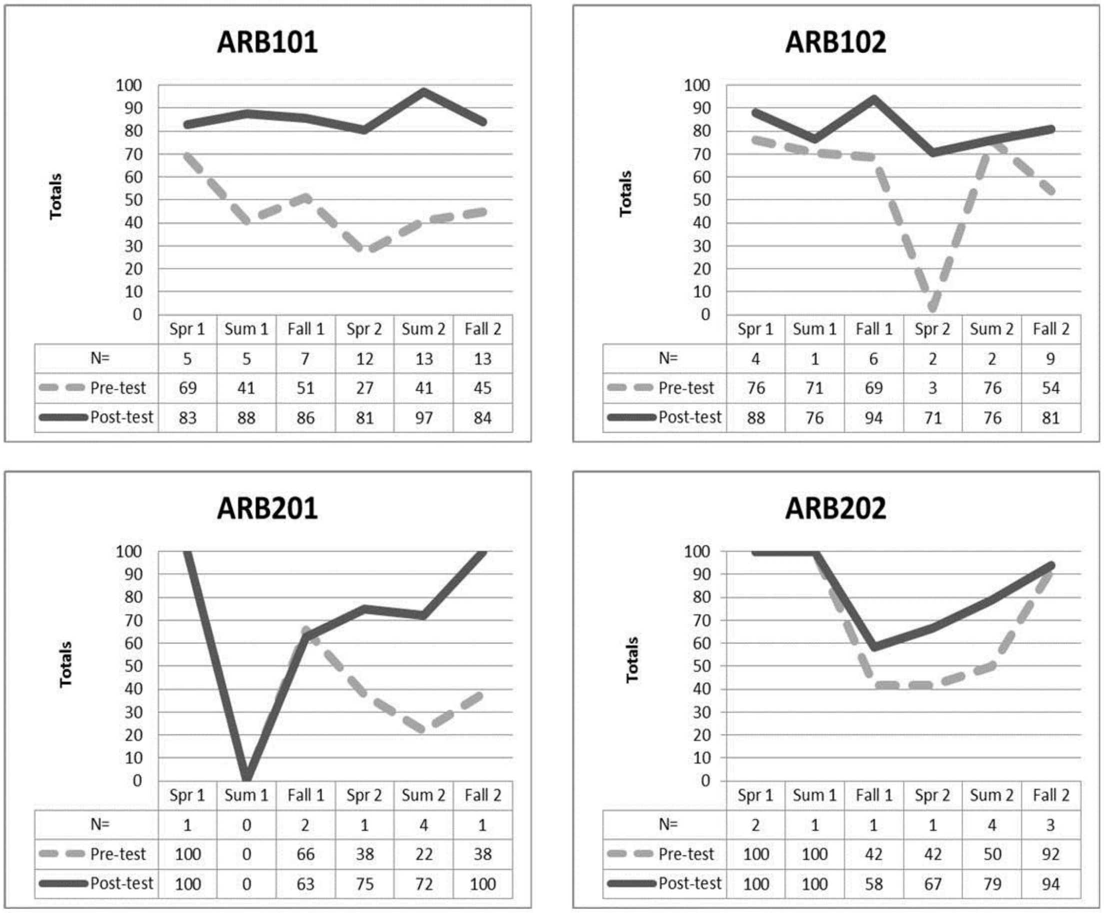
	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	1	3	1	4	3	2
Pre-test	70	93	90	78	50	95
Post-test	100	100	100	98	50	95

Arabic

Arabic courses have been available online since 2007, and 103 pre- and post-tests were submitted during the study period. The largest consistent gap between average pre- and post-test scores was at the 101 level, which indicates the greatest increase in student learning (see Figure 6). This is a shared characteristic among the German, Japanese, Chinese and Arabic results.

Enrollments in all Arabic courses are quite low ─ Just 10 pre/post-tests were submitted for ARB201 over the two-year study period. The semesters in which only 1 or 2 students were enrolled caused dramatic fluctuations in the reported results. Though the data is not significant enough to draw meaningful conclusions, the average post-test scores do exceed the 70% target for all four levels.

Figure 6. *Average pre- and post-test scores for Arabic 101-202*



ARB101

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	5	5	7	12	13	13
Pre-test	69	41	51	27	41	45
Post-test	83	88	86	81	97	84

ARB102

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	4	1	6	2	2	9
Pre-test	76	71	69	3	76	54
Post-test	88	76	94	71	76	81

ARB201

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	1	0	2	1	4	1
Pre-test	100	0	66	38	22	38
Post-test	100	0	63	75	72	100

ARB202

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	2	1	1	1	4	3
Pre-test	100	100	42	42	50	92
Post-test	100	100	58	67	79	94

American Sign Language

Courses in American Sign Language (SLG) have only been offered since 2008, and 571 pre- and post-tests were completed during the study period. Even though it is the youngest curriculum, enrollments for SLG courses are greater than those of all the other languages except Spanish.

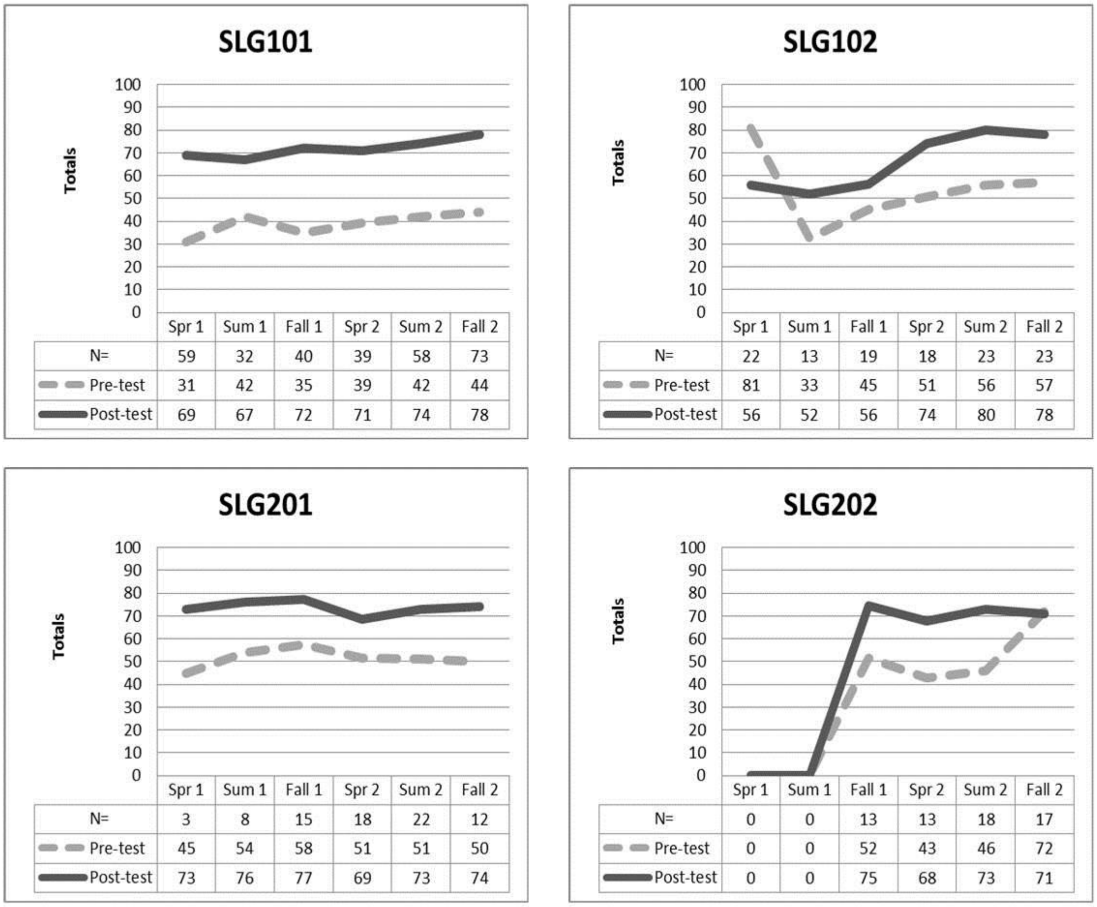
SLG101 is the most mature of the American Sign Language courses, and the predictable average pre- and post-test scores are the result of several improvement cycles to validate the assessment items and implement effective instructional interventions. Data indicate a significant increase in learning during the study period (see Figure 7).

SLG102 was offered for the first time in 2009, and the unusual inversion of pre- and post-test results emerged during the validation of assessment items. More than a year was spent on this work, and test validity was achieved by the beginning of the fall 2010 semester.

Average pre/post-test results for SLG201 are similar to those in FRE201, and the enrollments for those courses are comparable as well.

Fourth-semester American Sign Language (SLG202) was offered for the first time in the summer of 2010 with only one enrollee. Validation of test items is not yet complete, so no instructional interventions have been incorporated to date. Though average post-test results for SLG202 met the target, additional research is needed to validate the instrument in light of the fall 2 pre/post-test results.

Figure 7. *Average pre- and post-test scores for American Sign Language 101-202*



SLG101

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	59	32	40	39	58	73
Pre-test	31	42	35	39	42	44
Post-test	69	67	72	71	74	78

SLG102

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	22	13	19	18	23	23
Pre-test	81	33	45	51	56	57
Post-test	56	52	56	74	80	78

SLG201

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	3	8	15	18	22	12
Pre-test	45	54	58	51	51	50
Post-test	73	76	77	69	73	74

SLG202

	Spr 1	Sum 1	Fall 1	Spr 2	Sum 2	Fall 2
N=	0	0	13	13	18	17
Pre-test	0	0	52	43	46	72
Post-test	0	0	75	68	73	71

Key Findings

All Languages

- Approximately 4059 student scores (pre- and post-test) were included in the study.
- Students scored an average of 57.75% on the pre-test and 79.67% on the post-test for all courses.
- Results show that student scores increased an average of 21.92% (from pre- to post-test) for **all** courses.
- Languages that met the 70% post-test target standard include Spanish, French, German, Japanese, Mandarin Chinese, and Arabic.
- American Sign Language did not meet the 70% post-test target standard with an average post-test score of 67.42%.

Spanish

- Approximately 2332 student scores (pre- and post-test) were included in the study.
- Students scored an average of 67.83% on the pre-test and 80.21% on the post-test for all SPA courses.
- Results show that student scores increased an average of 12.38% (from pre- to post-test) for all SPA courses.
- Spanish had the highest enrollment and second highest pre-test scores. Also, this course had the lowest increase in post-test scores.

French

- Approximately 494 student scores (pre- and post-test) were included in the study.
- Students scored an average of 54.21% on the pre-test and 74.92% on the post-test for all FRE courses.
- Results show that student scores increased an average of 20.71% (from pre- to post-test) for all FRE courses.

German

- Approximately 234 student scores (pre- and post-test) were included in the study.
- Students scored an average of 56.99% on the pre-test and 78.82% on the post-test for all GER courses.
- Results show that student scores increased an average of 21.83% (from pre- to post-test) for all GER courses.

Japanese

- Approximately 196 student scores (pre- and post-test) were included in the study.
- Students scored an average of 51.50% on the pre-test and 83.38% on the post-test for all JPN courses.
- Results show that student scores increased an average of 31.88% (from pre- to post-test) for all JPN courses. *This is the highest increase of all language courses.*

Mandarin Chinese

- Approximately 128 student scores (pre- and post-test) were included in the study.

- Students scored an average of 70.21% on the pre-test and 91.79% on the post-test for all CHI courses.
- Results show that student scores increased an average of 21.58% (from pre- to post-test) for all CHI courses.

Arabic

- Approximately 104 student scores (pre- and post-test) were included in the study.
- Students scored an average of 56.42% on the pre-test and 81.17% on the post-test for all ARB courses.
- Results show that student scores increased an average of 24.75% (from pre- to post-test) for all ARB courses.

American Sign Language

- Approximately 571 student scores (pre- and post-test) were included in the study.
- Students scored an average of 47.12% on the pre-test and 67.42% on the post-test for all SLG courses.
- Results show that student scores increased an average of 20.30% (from pre- to post-test) for all SLG courses.

Discussion

A unique feature of this study is that faculty across all the languages were involved in the creation and implementation of the pre- and post-tests, and they continue to contribute to the ongoing improvement cycles that are the hallmark of this approach. Pre-post data is shared at department meetings, and faculty are tasked with designing interventions to address problem areas. Sometimes there are immediate gains (see American Sign Language), but other times initial interventions did not correlate with increased post-test scores (see French). In the case of French 201 specifically, data from another improvement cycle will be gathered and analyzed before embarking on further curricular modifications. In addition to engaging in intentional work to mitigate deficiencies, faculty are able to take pride in the results of previous efforts that have correlated with increased post-test scores. Thus, encouraging transparency in this environment is as much about celebrating successes as it is about solving problems.

Conclusions

A comprehensive study of this type is one answer to the call for increased accountability at all levels in higher education. The availability of online grade books to facilitate student access to their own pre/post results provides evidence of their individual learning, the sharing of cumulative results with faculty informs course-level improvement work, and the posting data for college-level review provides a direct measure of learning that can be used to evaluate the effectiveness of the language programs. Assessing *for* learning at all levels is a critical component of this approach. Providing students with tools to self-assess supports learner autonomy, which, according to Murphy (2008), is a critical component of successful distance language programs.

The ongoing continuous improvement cycles required to validate the assessments and design instructional interventions is arduous, time-consuming work. Using a "one course, many sections" model, multiple faculty members teaching the same courses must collaborate and make data-driven decisions designed to increase student learning. They must examine their own motives and practices to fight against the inertia that often plagues online courses. There is no true "end product" for this approach. Once a target has been reached, the bar is set higher, and the work continues.

Incorporating an all-inclusive pre/post-test methodology goes beyond a reliance on indirect measures such as grade distributions and subjective student/peer evaluations to evaluate the effectiveness of language instructors, courses and programs. Providing direct evidence of student learning at the college level encourages the use of objective metrics for program assessment that promote transparency and accountability at all levels.

Biodata

Angela Felix is the Faculty Chair of Languages at Rio Salado College. Her main research interests include adult heritage Spanish speakers in the United States and transparency and accountability in postsecondary education.

References

ACTFL Proficiency Guidelines. (2012). Retrieved from http://actflproficiencyguidelines2012.org/

Bahrani, T. (2011). Technology as an assessment tool in language learning. *International Journal of English Linguistics, (1)*2, 295-298.

Brooks, V. (2002). *Assessment in secondary schools: the new teacher's guide to monitoring, assessment, reporting, recording and accountability*. Buckingham: Open University Press.

Carr, N., Crocco, K., Eyring, J., & Gallego, J. (2011). Perceived benefits of technology enhanced language learning in beginning language classes. *The IALLT Journal, 41*(1), 1-32.

Coleman, J., & Furnborough, C. (2010). Learner characteristics and learning outcomes on a distance Spanish course for beginners. *System*, *38*(1), 14-29.

Council of Europe. (2001). *The Common European Framework of Reference for Languages.* Cambridge: Cambridge University Press.

Goldfield, J. (2010). Comparison of the ACTFL Proficiency Guidelines and the Common European Framework of Reference (CEFR). Retrieved from http://www.faculty.fairfield.edu/jgoldfield/ACTFL-CEFRcomparisons09-10.pdf

Hunt, M., Neill, S., & Barnes, A. (2007). The use of ICT in the assessment of modern languages: The English context and European viewpoints. *Educational Review*, *59*(2), 195-213.

Murphy, L. (2008). Supporting learner autonomy: Developing practice through the production of courses for distance learners of French, German and Spanish. *Language Teaching Research*, *12*(1), 83-102.

Palalas, A. (2011). ESP for busy college students: is the blend of in-class, online & mobile learning the answer? *The IALLT Journal, 41*(1), 108-136.

Phillips, June K. (2006). Assessment now and into the future. In *ACTFL 2005-2015: Realizing Our Vision of Languages for All*. ACTFL Foreign Language Education Series. (pp 75-103). New Jersey: Pearson Prentice Hall.

Swanson, P. & Nolde, P. (2011). Assessing student oral language proficiency: cost-conscious tools, practices & outcomes. *The IALLT Journal, 41*(2), 72-88.