
Alternative Assessments and Student Perceptions in the World Language Classroom

Robin Barnard Bachelor
Illinois Central College, East Peoria, IL

Abstract: *World language educators are constantly tasked with evaluating appropriate and beneficial assessments for their students. This study investigated how first-year students in a Midwestern United States high school perceived three different alternative assessments in the world language classroom: Dynamic Assessment, Task-based Assessment, and Formative Assessment using self- and peer-evaluation. The researcher correlated the perceptions to the students' assessment scores. Additionally, the researcher compared final exam scores of the control group with those of the experimental group. The results indicated that the experimental group students favorably perceived the alternative assessments types, chose Formative Assessment as their most preferred and Task-based Assessment as the least desired, and scored equally as well as the control group students on the final exam.*

Keywords: world/foreign languages, student perceptions, alternative assessment, dynamic assessment, task-based assessment, formative assessment, self- and peer-evaluation, world/foreign language policy

Introduction and Rationale

In the ever-evolving world of education, practitioners are faced with the challenge of meeting the needs of all students as well as the demands of state performance evaluations. In an effort to address this quest, educators must research best practices in the areas of both classroom methodology and assessment and must create a perceived value to students. According to Morrow, Shanahan, and Wixson (2012), future state and federal requirements will necessitate appropriate assessment choices which, in turn, will drive curriculum and instruction. They go on and argue that assessment choices are crucial for successful student preparation; the assessments should emphasize critical reading, writing, and higher-order thinking skills. These assessments are not remotely similar to the former, traditional state assessments but will be used to measure teacher success, evaluation, and retention. Teachers now find themselves in the position of designing alternative assessments that measure reading, writing, and higher-order thinking so that students are better prepared for college and the real-world. This task is further complicated for foreign language teachers who are instructed to uphold all of the following standards: National Standards for Foreign Language Learning, existing state standards, and, in some states, the Common Core State Standards (American Council on the Teaching of Foreign Languages [ACTFL], 1998; Illinois State Board of Education, 2012; National Governors Association Center for Best Practices, & Council of Chief State School Officers, 2010). As a result, this researcher developed a study to measure the effectiveness of certain alternative assessments that would meet many of the current demands.

The results of this study may be of interest to world language teachers at the high school level in particular as they restructure their curriculum and assessments to meet current demands and state testing measures. Additionally, while many studies have been completed that show the benefits of certain alternative assessments in the world (foreign) language classroom, very few included high school students as participants, most included university-aged students as subjects. Additionally, finding existing studies centered on a variety of al-

ternative assessments was difficult; the researcher chose to use three different types. The primary reason for providing a variety of alternative assessments was to discern student perceptions for comparison purposes and perceived interest.

Literature Review

Using traditional, summative testing in the classroom as a way for preparing for state testing is no longer sufficient. Poehner and van Compernelle (2011) discussed the need for teaching that promoted development and was a process, not just an end-of-lesson assessment. Sidek (2012) explained that traditional testing, which typically required students to comprehend and process specific data, had to be modified to include assessments with meaningful tasks that were more communicative in nature. Possible assessment types that would facilitate this change include dynamic assessment (Antón, 2009; Poehner & van Compernelle, 2011; Wei, 2011), task-based assessment (Byrnes, 2002; Carless, 2007; Sidek, 2012; Skehan & Foster, 1997), and formative assessment using peer- and self-evaluations (Bryant & Carless, 2009; Colby-Kelly & Turner, 2007; Hanrahan & Isaacs, 2001; Tamjid & Birjandi, 2011).

Dynamic Assessment is based on Vygotsky's (1978) theory that teacher interventions with clear examples and instruction, along with individualized prompts, help students move into self-reliance and mastery. Antón (2009) used a pre/post-test method to go along with Vygotsky's theory. After the initial assessment, the researcher documented specific points of remediation to address during the unit. She conferenced with her students to discuss their individual needs and then retested them. Using a similar approach, Wei (2011), who sought to use Vygotsky's theory, developed a Dynamic Assessment process for her classroom. She discovered her students' needs through an initial analysis followed by goal setting. Wei then planned assessments and made her students comfortable with the testing process. She continuously re-evaluated and revised her plans to meet the needs of her students and gave constant feedback both verbally and through multiple, short assessments during the unit. The format used for the current study was Dynamic Assessment focused on conferencing and group strategy sessions as well as a pre/post-test, similar to the unit designs of Antón and Wei.

Another potential alternative assessment is Task-based Assessment. According to Byrnes (2002), Task-based Assessment focused on language use and meaning that is contextualized in a communicative manner, typically through writing. The purpose of Byrnes' research was to promote a shift from grammar-based instruction through a different assessment approach. Byrnes created rubrics to evaluate students' work to ensure consistency and use as a guideline for students while they worked. Assessments were created with communication in a real-world context, or as close to real-world as possible, at the forefront while still addressing the content needs of the textbook. The assessments were all writing assignments. Ke (2006) also researched Task-based Assessment in order to promote communicative competency. Ke used a combination of reading, writing, listening, and speaking skills with real-world task assessments. Carless (2007) continued the Task-based Assessment research by interviewing secondary teachers who used this assessment in their classrooms. He concluded that some traditional teacher methods incorporating grammar should preclude any task-based assessment. For the purpose of this study, teaching grammar, such as Carless suggested, was incorporated, along with the creation of an assessment combining reading, writing, listening, and speaking skills, as incorporated by Ke.

Formative Assessment using self- and peer-evaluation forms is another viable alternative assessment. Tamjid and Birjandi (2011) realized a need for altering traditional assessments in order to stay current with the movement toward learner-centered classrooms. In their study, all students were assigned writing prompts that the instructors graded; however, the experimental group also completed self-assessment rubrics and were randomly given another student's assignment to assess using the same rubric. The students all revised their assignments before submitting to the instructors. Tamjid and Birjandi found that the experimental group improved their metacognition which led to better thinking and learning skills that could be used on future assignments. Bryant and Carless (2009), also used peer-assessment methods; they hypothesized that self- and peer- assessments would stimulate more learner independence and create an atmosphere in which students desired improvement through reflective thinking. Their students viewed the peer-assessment rubric as a tool to help them earn better grades instead of a waste of time. The current study made use of Tamjid and Birjandi's process of self- and peer-assessment as a reflective measure to improve student writing while incorporating unit grammar and vocabulary into the prompt.

Student-Perceived Value

To continue the thoughts of a learner-centered classroom with student reflective thinking, the critical part of this research was student perceptions. Students must perceive value for any incentive; this is critical to success in the classroom. According to Palloff and Pratt (2007), student-perceived value is essential. The researchers suggested that educators design activities that interest students by relating to their life activities, communicate clear expectations, create positive classroom atmospheres, and make use of alternative assessments.

Research Methods and Procedures

Venue and Subjects for the Research Study

According to the Illinois School Report Cards (2012), School X was a kindergarten through twelfth grade unit district with approximately 920 students, 300 of whom attended the high school. The ethnicity of the district was primarily White, at approximately 79%. Black students encompassed over 8% of the population, and Asian/Pacific Islanders held almost 6%, as did Multi-ethnic students. Less than 1% was Hispanic or Native American. The percentage of low-income students was 11%. The average ACT score of the previous junior class was 24. The graduation rate was just over 95%. School X's location was metropolitan, lying on a major river, and was approximately 165 miles from two megacities in the Midwestern section of the United States.

School X typically offered three sections of Spanish One each fall. Spanish One is a beginner, introductory course for students to start Spanish language study. This is a high school, freshman-level course. Generally, the students in these classes were eighth graders or first-year high school students; however, a few sophomores, juniors, and seniors comprised the overall makeup.

The classroom teacher taught three sections of Spanish One during the fall of 2013. Typically, a section included 25 to 30 students. Most students ranged from 13 to 15 years of age at the onset of the study. To choose the research sample, the researcher included the first section of the day as the control group, 27 students, in which 14 agreed to participate in the study. This group had 9 females and 5 males. The second and third sections were the experimental group, 54 students, in which 34 agreed to participate in the study. This group had 15

females and 19 males. The researcher designated the experimental and control groups before seeing the lists of students in each class to reduce any possibility of bias.

Pre-Data Collection Procedures

The researcher requested and received approval from a School X administrator to complete this study. After meeting the students in August, 2013, the researcher explained the study and asked for the students to sign a research study assent form. The researcher sent home letters to the parent/guardian and acquired consent signatures. The researcher, in cooperation with the classroom teacher from School X, personally controlled all aspects of the study, including alternative assessment design for each unit, classroom procedures during tests, and data collection and storage procedures. The classroom teacher and the researcher also ensured that all students took the departmentally required summative assessment. Student data was accessed through the school secured server and self-reporting on the surveys.

Minimal risks existed for the participants of this study. The students were treated the same as in any other school year or with any other teacher, using similar methodology, using the same text book, and by following all rules and procedures at School X. The experimental group received the same vocabulary and grammar instruction as the control group, along with the required department summative exam. None of the department goals or objectives for Spanish One were altered for either group.

Research Design

The research was completed during the fall of 2013 over a 15-week time frame. The researcher chose a concurrent nested design study that fell within the mixed-method realm (Robson, 2011). The primary method was quantitative through the analysis of the summative test scores of the control and experimental groups as well as data analysis correlating the perceptions of the experimental group with the test scores. Teacher journal observations stood as a secondary, qualitative measure. The teacher recorded naturally-occurring data, notations of attitudes, behaviors, and comments relating to the alternative assessments used with the experimental group during the study in the journal. Additional qualitative measures required the researcher to note the students' overall preferences of alternative assessment type through a set of post-survey open-ended questions by categorizing the responses into themes.

This study was conducted with 48 participants, students in their first year of world language study earning high school credit. The School X teacher assessed the 14 members of the control group using fairly traditional, previously used quizzes and tests throughout the semester. The remaining 34 participants, the experimental group, took alternative assessments instead. The researcher sought the perceptions of these students concerning three different alternative assessments types. This mixed-method study, within the concurrent-nested design (Robson, 2011), assessed students' perceptions.

Data Collection

The researcher began by collecting data from the Educational Planning and Assessment System (EPAS) in School X's confidential computer server space. It would be purposeless to give a pre-test to beginning Spanish learners, so the researcher used the most recent reading score for each student to determine the level of each student in the control and experimental groups. The researcher ran an independent t test on the two groups to show equivalency. The two groups were similar enough to have confidence that any difference in summative assessment data was due to the intervention.

During the 15 weeks of the study, the classroom teacher taught three units to the students while keeping a journal about the process. The teacher used the same teaching methods for covering the required vocabulary and grammar goals for the control and experimental groups. Departmental goals are established at School X and were followed for all classes in the control and experimental groups. The primary difference was in the assessment. The teacher administered the department's traditional, summative final exam to both the control and the experimental groups. In contrast to the control group, the experimental group took a dynamic assessment for unit one, a task-based assessment for unit two, and a formative assessment using peer- and self-evaluations involving writing assignments for unit three. The researcher gave a code word to each alternative assessment type: Blue for dynamic assessment, Red for task-based assessment, and Green for formative assessment using self- and peer-evaluations. At the end of each unit, the students in the experimental group completed a researcher-developed survey with a 4-point Likert scale to measure perceptions.

Research Questions

The following research questions guided the present study:

1. What perceptions do first-year foreign language students have of dynamic assessment, specifically focused on conferencing and group strategy sessions?
2. What perceptions do first-year foreign language students have of task-based assessment, specifically focused on real world communication?
3. What perceptions do first-year foreign language students have of formative assessment, specifically using peer- and self-evaluations?
4. How do the summative assessment scores of the students who received interventions differ from those who did not?
5. How do student perceptions of the alternative assessments correlate with their test scores?

Data Analysis

To answer Research Questions 1-3, Likert-type surveys were given to the experimental group at the end of the three units and were analyzed. For each survey, the researcher categorized the students' responses and used a bar chart to express the experimental group's agreement level, how the particular assessment reflected learning. Since 4-point, Likert-type data are measured on an interval scale, the researcher gave descriptive analysis of the tendencies. A one-way ANOVA measured the associations between the student choices to show statistically significant differences by assessment type (Leedy & Ormrod). To follow up the one-way ANOVA, a Bonferroni *post hoc* determined where the differences existed. The researcher continued by writing descriptive, narrative accounts to relate the teacher journal observations after transcribing the journal into common themes. The themes revealed student perceptions of the alternative assessments from the teacher's perspective.

Research Question 4 required comparative data analysis to be answered. An independent t test was used to "determine whether two groups of scores are significantly different" (Gay, Mills, & Airasian, 2012, p. 351). The means of the control and the experimental groups were compared to show any statistical significance using a box plot.

Finally, Research Question 5 required analysis of the relationship between the experimental students' perceived value for each assessment and their test scores on each assessment

Initially, a spreadsheet was composed to display each student's alternative assessment scores. The researcher continued by categorizing the student perceptions of the three alterna-

tive assessments. Correlational research “involves collecting data to determine whether, and to what degree, a relation exists between two or more quantifiable variables...a decimal number between -1.00 and +1.00” (Gay, Mills, & Airasian, 2012, p. 624). The researcher used a correlation coefficient formula to determine a positive, negative, or nonexistent correlation between the students’ perceptions and their alternative test scores. For each assessment, the researcher correlated the student’s alternative assessment score to his/her combined survey answers and displayed the results with scatter plots. The researcher used the student data to rank the three alternative assessments in order of preference, based on information from the post-study survey and then described these results. Narrative descriptions of the open-ended survey questions completed the analysis. The researcher took the students’ words, found patterns, and developed themes to reveal correlative information.

Findings

Research Questions 1-3

Research Questions 1-3 required descriptive statistics that measured students’ perceptions following the dynamic, task-based, and formative unit assessments. To show the students’ perceptions of each assessment type, the researcher combined the two, 4-point Likert-type question scores, for a maximum of eight points and displayed the data using bar charts (Figures 1, 2, and 3). Figure 1 indicates that 25 of the 34 students in the experimental group agreed that the Dynamic Assessment showed the teacher how much they knew and favored using this type of assessment again in the future. None of the students strongly disagreed that this type of assessment should be continued nor that it showed how much they knew. Out of the 34 students, only 6 had disagreement to the Dynamic Assessment.

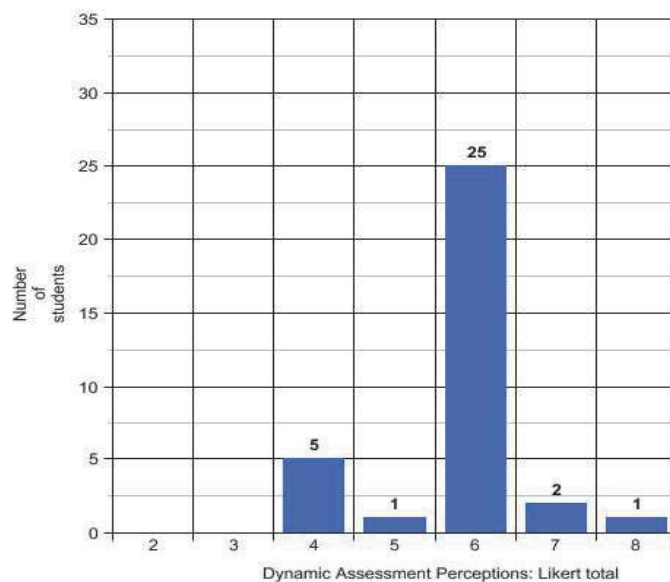


Figure 1. Dynamic Assessment perceptions. This figure illustrates the student’s perceptions of Assessment Blue on a scale of 2 to 8.

Figure 2 shows that five of the 34 students strongly disagreed that the Task-based assessment administered showed what they knew and that the teacher should use this assessment type later during the year. Twelve of the 34 had a mixed opinion of agreement with disagreement on the assessment. Overall, no students strongly agreed concerning Task-based Assessment.

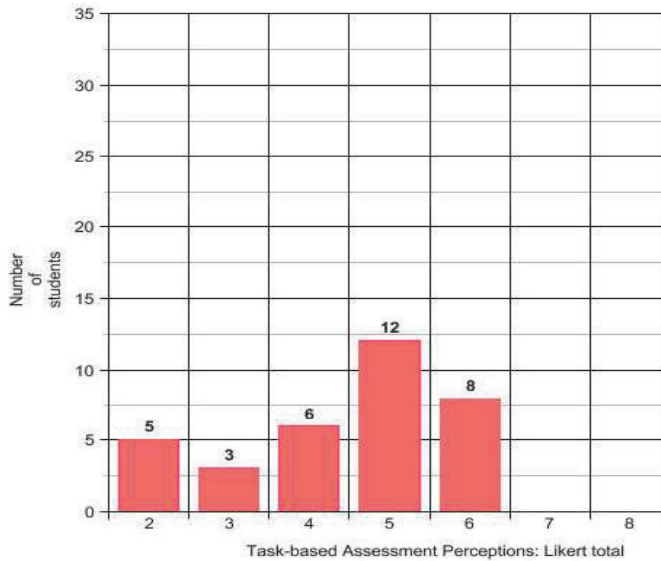


Figure 2. Task-based Assessment perceptions. This figure illustrates the student’s perceptions of Assessment Red on a scale of 2 to 8.

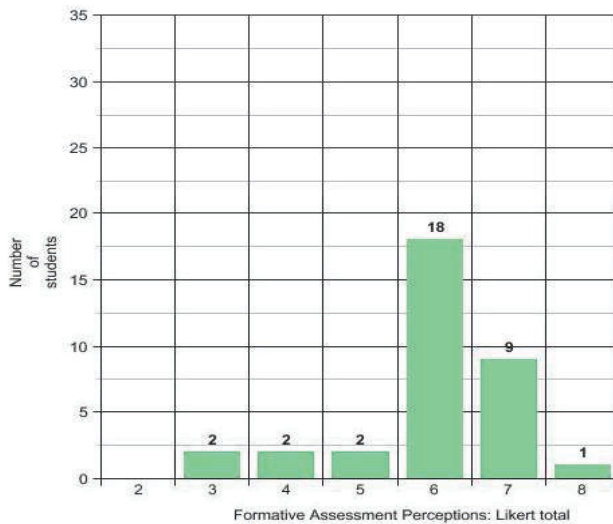


Figure 3. Formative Assessment perceptions. This figure illustrates the student’s perceptions of Assessment Green on a scale of 2 to 8.

To finalize the quantitative analysis for Research Questions One, Two, and Three, the researcher computed a one-way, within subjects ANOVA to compare the effect various alternative assessments had on perceptions, specifically regarding Dynamic, Task-based, and Formative Assessment types. The means of student perceptions were associated with the student perceptions at the three different points after each assessment. The researcher found that at least one of the means for perceptions was different from the others: $F(2, 66.0) = 22.75, p < .001, \text{partial } \eta^2 = .41$. Using Bonferroni’s correction for family-wise errors, the student perceptions of Dynamic Assessment (Blue) did not differ greatly from the perceptions of Formative Assessment (Green): $p = 1.0$, which is not significant. However, when comparing the student perceptions of Dynamic Assessment (Blue) to Task-based Assessment (Red), $p < .001$ which was significant, with Dynamic Assessment being better perceived than Task-

based Assessment. Similarly, Task-based Assessment (Red) perceptions compared to those of Formative Assessment (Green), $p < .001$ which indicated significance, as Formative Assessment was better perceived than Task-based Assessment. The students had a lower perception of Task-based Assessment than they held of either Dynamic or Formative Assessment; while their perceptions of both Dynamic and Formative Assessments were similar. An eta-square of .41 showed a large effect size.

From the post-study survey, the researcher tabulated the number of students citing each assessment type as the best and worst for reflecting knowledge of the material. Twenty-five of the 34 students recorded Formative Assessment, in which they wrote and used self- and peer-evaluation, as the type that best reflected their knowledge (see Figure 4).

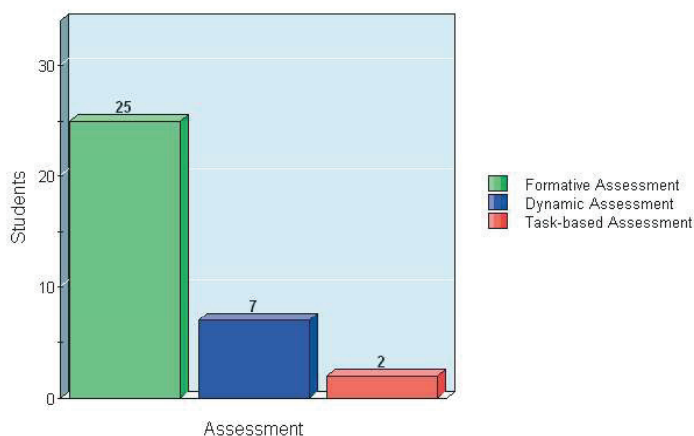


Figure 4. Best-Perceived Assessment type. This figure illustrates the students' perceptions of which assessment was the most favorable.

For the worst-perceived assessment, Task-based Assessment merited 29 of the 34 students' choice as not reflecting their knowledge. See Figure 5.

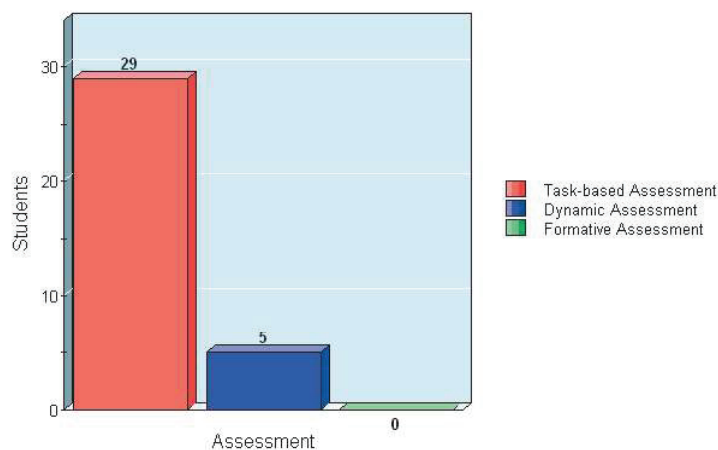


Figure 5. Worst-Perceived Assessment type. This figure illustrates the students' perceptions of which assessment was the least favorable.

Using qualitative procedures, the open-ended responses given by the students on the post-study survey were coded. Then they were assigned themes concerning the students'

perceptions regarding why an assessment type was best or worst for knowledge reflection. Tables 1 and 2 depict those themes with the number of students claiming them. The students remarking that test format was crucial for the best perceived assessment also included specific details, such as, appreciating the revision process on the Formative Assessment, feeling advantaged from experiencing the intervention and taking the pre-test on the Dynamic Assessment, and experiencing better success from the Task-based Assessment's separate task sections. When referring to his preference for Formative Assessment, one student stated, "It let me show that I can communicate." Another student commented, "I could pick what to say based on the words I knew." An additional proponent of Formative Assessment indicated, "I could test my knowledge and have someone revise it." In all, the students clearly favored the Formative Assessment as being most reflective of their knowledge.

Table 1
Best-Perceived Assessment: Themes

Assessment Type	Themes Seemed easy/ Knew material	Freedom	Test Format	Practical/ Communicative	Did Not Respond
Formative Assessment-Green	10	9	2	4	0
Dynamic Assessment-Blue	1	0	5	0	1
Task-based Assessment-Red	0	0	1	0	1

In reference to the worst perceived assessment, students mentioned that tasks were overwhelming to them on the Task-based Assessment and that they were unclear on what exactly was expected on this same test. One student wrote, "It was too much..." and another related, "It was confusing and hard." Two students reported that they did not favor a pre/post-test design, as seen on the Dynamic Assessment. Overall, the students relayed that the Task-based Assessment did not reveal their understanding as fully as did the Formative or Dynamic Assessments.

Table 2
Worst-Perceived Assessment: Themes

Assessment Type	Themes Confusing	Hard/Did not Know or Understand Material	Test Format	Did Not Respond
Task-based Assessment-Red	17	9	5	1
Dynamic Assessment-Blue	0	2	2	1
Formative Assessment-Green	0	0	0	0

The classroom teacher provided limited journal notes concerning students' behaviors and comments at the time of each alternative assessment, but also delivered his perceptions of how each assessment impacted the classroom. He noted that the Dynamic Assessment (Blue) format with a pre- and post-test, as well as an intervention and group strategy session, seemed beneficial to most students. The two test method helped students better anticipate what to expect early in the year. The teacher also stated that the pre-test and intervention guided his instruction to more effectively meet the needs of the students. However, many students complained that Task-based Assessment (Red) was difficult, and some stated it was confusing. The teacher felt that task-based activities may have been better suited as an in-class partner-practice activity or that some sort of task-based activity needed to be used during the unit and not just as the final assessment. Formative Assessment (Green) was well liked by the students. They commented to him that it was beneficial, practical, and allowed them to say whatever they wanted and avoid topics of which they had less comfort. Students appeared motivated to show off their knowledge for this assessment.

Research Question 4

To compare the mean of the final exam scores of the experimental group to that of the control group, an independent samples *t* test was calculated through SPSS. This information was visually displayed as a box plot in *Figure 6*. There was not a significant difference in the scores of the experimental group's final exam ($M= 84.56$, $SD= 13.96$) and from the final exam of the control group ($M= 85.00$, $SD= 15.34$); $t(46)= -.097$, $p = .923$, $d= .03$. These results suggest that the experimental group had comparable Spanish proficiency at the end of the semester to the control group.

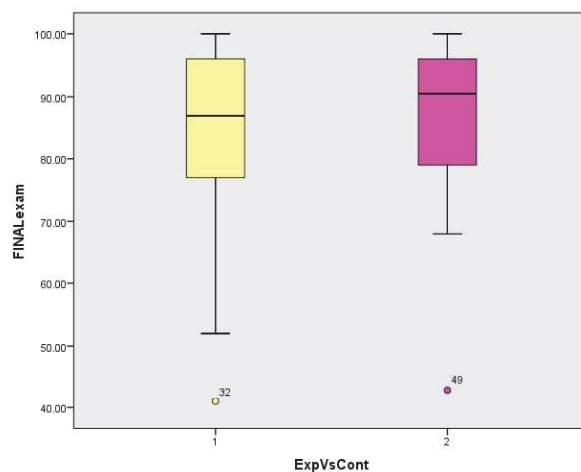


Figure 6. Final Exam scores. This figure illustrates the final exam scores of the experimental group in yellow compared with those of the control group in purple, on a scale of 0 to 100.

Research Question 5

Research Question 5 was concerned with showing relationships and not causes between the students' perceptions of each alternative assessment and each student's score on that assessment. A Spearman's Rho correlation coefficient was computed to assess the possible relationships. *Figures 7, 8, and 9* illustrate that the X axis represented the students' perceptions of an alternative assessment, and the Y represented their test scores on that particular assessment.

When analyzed, the Dynamic Assessment (Blue) showed no correlation between the

students' perceptions and their scores, $r_{s(32)} = .29$, $p = .09$, with a medium effect size of .293. Some students positively perceived the Dynamic Assessment and had high scores, while others viewed it similarly but had average or low scores. *Figure 7* summarizes the results with a scatterplot.

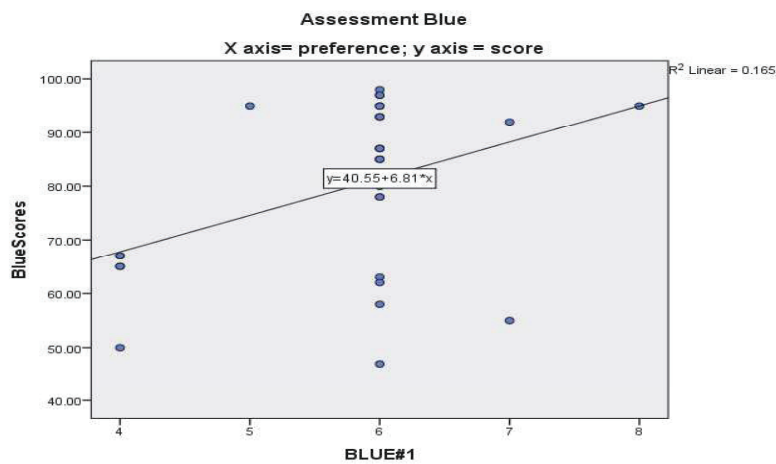


Figure 7. Dynamic Assessment (Blue) Correlation. This figure illustrates the correlation between the test scores (on a scale of 0 to 100) on Assessment Blue and each student's perception of the assessment (on a scale of 2 to 8).

In regard to the Task-based Assessment (Red), there was a significant positive correlation between the students' scores and their perceptions, $r_{s(32)} = .57$, $p < .001$, with a large effect size of .574. As test scores rose, positive perceptions of the assessment also rose; the converse was also true. A scatterplot summarizes the results in *Figure 8*, as well as depicts the positive correlation with the ascending linear regression line.

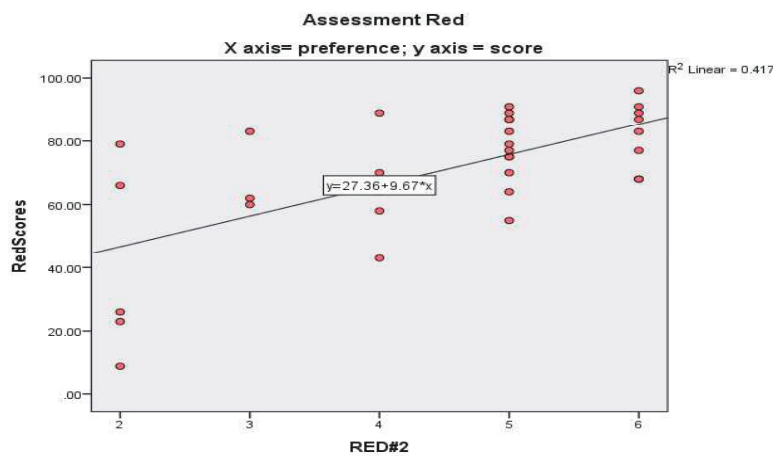


Figure 8. Task-based Assessment (Red) Correlation. This figure illustrates the correlation between the test scores (on a scale of 0 to 100) on Assessment Red and each student's perception of the assessment (on a scale of 2 to 8).

The Formative Assessment (Green) revealed no correlation between the assessment scores and the students' perceptions, $r s(32) = .10, p = .57$, with a small effect size of .101. The flat linear regression line on the scatterplot of *Figure 9* indicates that no meaningful relationship existed between assessment scores and perceptions in relation to this assessment type.

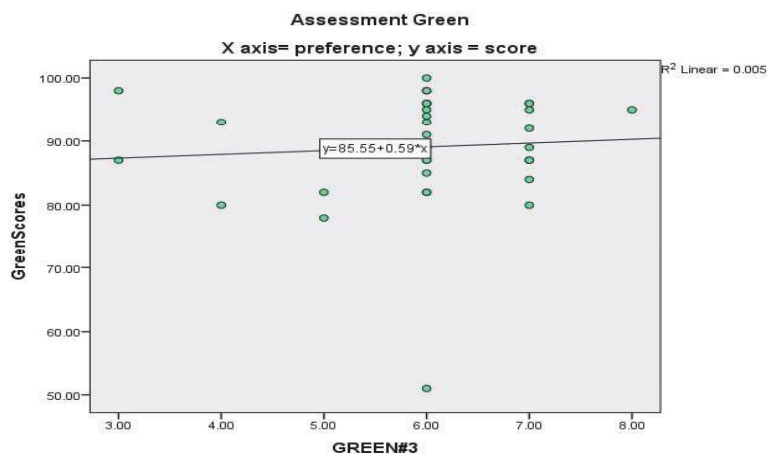


Figure 9. Formative Assessment (Green) Correlation. This figure illustrates the correlation between the test scores (on a scale of 0 to 100) on Assessment Green and each student's perception of the assessment (on a scale of 2 to 8).

Conclusions and Implications

Research Question 1 sought to determine the perceptions of first-year Spanish students about Dynamic Assessment. Twenty-eight of the 34 experimental group students acknowledged Dynamic Assessment as agreeable or strongly agreeable, primarily in the agree category. Further, although Dynamic Assessment was not chosen as the best-perceived assessment type during the study, it was also not chosen as the worst. The most noted positive features were the pre-/post-test design and the benefit of the intervention session. This researcher then concluded that Dynamic Assessment was perceived by the students as a favorable alternative assessment for world language classroom teachers.

In regard to Research Question 2, 20 of the 34 experimental group students noted that Task-based Assessment was agreeable to them, in that it demonstrated their knowledge and would be acceptable to use again. However, eight of the 34 students showed disagreeability or even strongly disagreed that Task-based Assessment was preferable. Task-based Assessment also ranked as the least preferred choice of the three alternatives. This researcher determined that while Task-based Assessment may be an acceptable option, it was not found to be highly preferable.

Formative Assessment perceptions' data needed to answer Research Question 3 revealed that 28 of the 34 students agreed or strongly agreed with this alternative assessment's value. Ten of them showed levels of strong agreement. Additionally, Formative Assessment was established as the most favorable alternative assessment of the three, chosen by 25 of the 34 students; not one student ranked Formative Assessment as the worst choice. This researcher established that Formative Assessment was well-perceived by the students in the experimental group and was the popular choice; thus, making Formative Assessment a valid alternative for L2 educators.

Research Question 4 which compared the means of the control and experimental groups on the final exam, showed no significant difference. This researcher concluded that the end-of-semester knowledge for first-year Spanish students was equal. Students in both groups, as measured by a common assessment, had acquired the same knowledge and skills. This researcher recognized that the use of alternative assessments neither advantaged nor disadvantaged the learners; therefore making alternative assessments viable choices for the foreign language teacher.

When comparing the perceptions of the experimental group students to each one's alternative assessment scores, this researcher was able to draw conclusions to answer Research Question 5. A significant positive relationship existed between the students' perceptions of Task-based Assessment and their assessment scores. This researcher surmised that student perceptions concerning Task-based Assessment did factor into performance. No correlation existed between the student perceptions of Dynamic Assessment and test scores or between Formative Assessment perceptions and test scores. Since these analyses established that no statistical significance existed, this researcher inferred that perceptions did not affect performance, positively or negatively, in relation to Dynamic and Formative Assessments.

This researcher concluded that, while student perceptions are important, as long as the classroom teachers design valid alternative assessments, students are agreeable to a variety of types and perceive them as accurate reflections of their knowledge. It may be noted, though, that some assessment types need to be practiced and modeled throughout a unit to increase their effectiveness and perceived value. For example, task-based assessment was perceived as confusing to students when experienced for the first time on an end-of-unit assessment. Overall, data did not support that student perceptions affected the outcomes required on a unit assessment, with the exception of Task-based Assessment, regardless of the testing format. While student perceptions and choices are important, they are not a reliable indicator on which to base the relationship to students' grades.

Implications

The alternative assessments from this study produce equally positive results to traditional methods; students also have positive perceptions of these assessments. With that said, the new, alternative assessments are superior to traditional assessments, as they produce equally positive results and are also compliant with the demands of state, national, and world language standards.

Teachers will continue to design assessments to meet their programs outcomes but should always consider their students' perceptions, alternatives to traditional design, and the bonus effects of creating assessments that lend themselves to higher-order thinking and real-world application. Dynamic Assessment, Task-based Assessment, and Formative Assessments using self- and peer-evaluation should be strongly considered for the L2 educator.

“Assessment is the key to language learning. Only through the assessment lens can learners find out if they are meeting their goals, to what extent those goals are being met, and what they still need to do. Through the feedback received, our students' motivational fires are fed” (Duncan, 2014, p. 19).

References

- ACTFL. (1998). *Standards for foreign language learning: Preparing for the 21st century*. Retrieved from <http://www.actfl.org/>
- Antón, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals*, 43, 576-599. doi:10.1111/j.1944-9720.2009.01030.x
- Byrnes, H. (2002). The role of task and task-based assessment in a content-oriented collegiate foreign language curriculum. *Language Testing*, 19, 419-437. doi:10.1191/0265532202lt238oa
- Carless, D. (2007). The suitability of task-based approaches for secondary schools: Perspectives from Hong Kong. *System*, 35, 595-608. doi:10.1016/j.system.2007.09.003
- Colby-Kelly, C., & Turner, C. (2007). AFL research in the L2 classroom and evidence of usefulness: Taking formative assessment to the next level. *The Canadian Modern Language Review* 64(1), 9-38. doi:10.3138/cmlr.64.1.009
- Duncan, G. (2014). Embracing the paradigm shift in learning and assessment. *The Language Educator*, 9(2), 18-19.
- Foster, P., & Skehan, P. (1997). Modifying the task: The effects of surprise, time and planning type on task based foreign language instruction. *Thames Valley University working papers in English language teaching*, 4, 86-109.
- Hanrahan, S. J., & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views. *Higher Education Research & Development*, 20(1), 53-70. doi:10.1080/07294360123776
- Illinois State Board of Education. (2012). Realizing Illinois. Retrieved from http://www.isbe.net/common_core/default.htm
- Ke, C. (2006). A model of formative task-based language assessment for Chinese as a foreign language. *Language Assessment Quarterly*, 3, 207-227. doi:10.1207/s15434311laq0302
- Leedy, P. D., & Ormrod, J. E. (2013). *Practical research: Planning and design* (10th ed.). Boston, MA: Pearson.
- Morrow, L. M., Shanahan, T., & Wixson, K. K. (Eds.). (2012). *Teaching with the common core standards for English language arts*. New York, NY: Guilford Press.
- National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). *Common core state standards initiative*. Retrieved from <http://www.corestandards.org/>
- Palloff, R. M., & Pratt, K. (2007). *Building online learning communities: Effective strategies for the virtual classroom*. San Francisco, CA: Jossey-Bass.
- Poehner, M. E., & van Compernelle, R. A. (2011). Frames of interaction in dynamic assessment: Developmental diagnoses of second language learning. *Assessment in Education: Principles, Policy & Practice*, 18, 183-198. doi:10.1080/0969594X.2011.567116
- Sidek, H. M. (2012). EFL reading instruction: Communicative task-based approach. *International Journal of Instruction*, 5(2), 109-128. Retrieved from http://www.e-iji.net/dosyalar/iji_2012_2_7.pdf
- Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1(3), 185-211. doi:10.1177/136216889700100302
- Tamjid, N. H., & Birjandi, P. (2011). Fostering learner autonomy through self- and peer-assessment. *International Journal of Academic Research*, 3, 245-252.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wei, L. (2011). Formative assessment in classrooms: Operational procedures. *Journal of Language Teaching and Research*, 2(1), 99-103. doi:10.4304/jltr.2.1.99-103

Robin Barnard Bachelor, Ed.D., has over 20 years of teaching Spanish, English, and Education at all levels (high school, undergraduate, and graduate). Dr. Bachelor created program and course outcomes for the Spanish program at her high school that complied with national and state standards, created common assessments for all course sections, and served on the Assessment Committee for overall institutional outcomes.

