

Integrated Performance Assessments: A Review of the Literature and Steps to Move Forward

Stephanie Madison Schenck

Clemson University

Abstract: Integrated Performance Assessments (IPAs) have been heralded as a more authentic manner in which to evaluate language learners. Based on the principles of Dynamic Assessment (DA) and Performance-Based Assessment (PBA), IPAs continue to be used in both K-12 and university language classrooms. However, some teachers may be hesitant to implement IPAs due to concerns with logistics, including time constraints and giving feedback to every student in a large class. This review of the literature explores the theoretical underpinnings to the IPA, as well as practical considerations for teachers.

Keywords: Integrated Performance Assessment, Dynamic Assessment, Performance-Based Assessment, proficiency, feedback

As language teaching and learning continue to evolve in foreign language classrooms across the United States, so, too, are the methods of assessing students' performance (Malone and Sandrock 2016). The American Council on the Teaching of Foreign Languages (ACTFL) first proposed the idea of a task-based assessment that would demonstrate what foreign language students could do in the target language in 1997. ACTFL, sponsored by the U.S. Department of Education International Research and Studies, participated in a three-year program to develop the Integrated Performance Assessment (IPA) which evaluates students' abilities in reading, writing, speaking, and listening in overlapping and integrated ways (Adair-Hauck, Glisan, Koda, Swender, and Sandrock 2006). This assessment model has its roots in Dynamic Assessment (DA), which focuses as much on instruction as the assessment (Davin 2013). The DA model was paired with Performance-Based Assessment (PBA), in which students address a task that reflects real-world situations in a holistic way that integrates linguistic skills, lexical knowledge, and cultural knowledge (Byrnes 2002). The result is what is now known as the IPA. The purpose of this literature review is to explore the rationale and benefits for using the IPA, as well as the shortcomings of such an assessment. I conclude by presenting possible solutions to address the reasons that classroom teachers are hesitant to implement the model.

Theoretical Framework

ACTFL developed the IPA in order to shift the focus of assessment from strict memorization to a performance and communicative model, which includes interaction with the teacher or with peers. This approach takes on a sociocultural perspective as opposed to a traditional cognitive perspective (Donato 2000).

Drawing on the work of Vygotsky, the zone of proximal development (ZPD) is “the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers” (1978: 86). Providing feedback and interaction throughout the assessment process is a concrete example of the ZPD in action.

Method

Research pertaining to IPAs was reviewed in order to examine the relationship between IPAs and participation by classroom teachers. Specifically, the purpose of this review is to identify the pros and cons to IPAs, and possible strategies to facilitate successful, or at least modified, implementation. In summary, this review explores the reasons IPAs are a beneficial assessment model for language learners, the implications for classroom instruction, and strategies to encourage IPA adoption for reluctant teachers.

A literature search was conducted for articles focusing on assessment models in foreign language classrooms. The following search terms were used in various combinations: assessment, integrated, performance, dynamic, task-based, acquisition, development, feedback, proficiency, L2, and foreign language. Academic Search Complete, Education Full Text, Education Research Complete, EBSCO Host, ERIC, PsychINFO, and Teacher Reference Center were searched for articles. I also searched with Google Scholar and used the “cited by” feature to see who referenced these articles in other papers.

To be included, an article had to meet the following criteria: (a) printed in a peer-reviewed journal published between 1997–2019; (b) written in English; (c) must address foreign language assessment; (d) focused on learners in K-12 and postsecondary settings; and (e) empirical study or scholarly article. The year 1997 was chosen as the start date because that is the year that IPAs were beginning to be developed. Book chapters and seminal pieces were also included for analysis, as well as information from ACTFL’s website and publications. Essays, commentaries, government reports, book reviews, and magazine and/or newspaper articles were excluded.

Findings

This review will outline the principals of DA and PBA to demonstrate their influence on the IPA as we know today. All three models focus on the importance of feedback throughout the learning and assessment process. However, there are several factors and challenges that may hinder the implementation of the IPA in classrooms.

Dynamic Assessment

DA is how students are able to achieve greater success in their individual ZPD in the same way they achieve success with guidance from the teacher in daily lessons and interactions (Vygotsky 1978). The assessment is not only designed to show what students know and can do, but what they can do with the support of the teacher. When applied to advanced learners of a second language, Antón (2009) found that DAs show more deeply and more clearly the extent of the student's emerging language abilities, allowing for more individualized guidance from the teacher. On the other hand, one of the complaints against DA is the lack of validity or reliability in the final scores, and it may appear some students unfairly receive more support from teachers than others (Lantolf and Poehner 2007; 2013). If the student's grade on the assessment is what they can do with the support of the teacher, then it is not going to demonstrate what students are capable of producing on their own. However, this view is problematic when considering DA through the lens of the ZPD, as DA intentionally provides varying degrees of support for each student according to what they need (Kozulin and Garb 2002).

DA frequently follows a three-step process, in which students are given an assessment, the teacher provides feedback, and then students are given an additional assessment, with the goal that the feedback will be sufficient to help students grow in the target language (Antón, 2009). The process focuses as much on instruction as the assessment, as the two are integrated and intertwined throughout the course (Davin 2013; Leung 2007). The individualized guidance is a logical consequence of DA, as the teacher is constantly interacting with each student throughout the feedback and assessment cycle. However, this constant interaction can be difficult, because it requires a great deal of in-the-moment decision making from teachers as they decide how to best respond to each student (Davin, Herazo, and Sagre 2017; Davin and Troyan 2015). In foreign language classrooms, the next step in the assessment process is to apply this model specifically in performance-based tasks.

Performance-Based Assessment

The goal of PBA is for students to address a task that reflects real-world scenarios in a holistic way that integrates skills, lexical knowledge, and cultural knowledge (Adair-Hauck et al. 2006). Byrnes (2002) said a task-based or performance-based approach to instruction and assessment is important, because students need to see the connection between what they are doing in class and what could potentially be done in the real world. The learning that is taking place is not simply for the sake of learning, but for the sake of applying it to interactions with others outside of the classroom.

However, even though tasks in PBAs are meant to mimic real-life situations in the target language, and therefore predict how an L2 learner would actually perform in such a situation, the unpredictable nature of language interactions make evidence for this assumption difficult to prove (Bachman 2002). The teacher can only provide an approximation of what a real-world interaction might look like as factors such as planning time and task structure can influence student performance (Tavakoli and Skehan 2005). In the classroom, these interactions are then judged holistically, not with specific point values deducted for each error or added for each correct answer. Rather, the teacher judges the answer provided as a whole before assigning a grade. Research suggests even a minimal amount of training in judging student performance can help with consistency in this type of evaluation (Norris, Brown, Hudson, and Bonk 2002).

Integrated Performance Assessment

The IPA model of assessment requires three separate tasks under the umbrella of a single theme or context. The student must exhibit presentational (speeches, essays), interpretive (reading or listening to an authentic resource in the target language), and interpersonal (exchange of information, conversation) skills that are all related to one another (Adair-Hauck et al. 2006). All of these tasks are based on an authentic context or topic that one would likely encounter in the target language in a similar way to PBA.

IPAs judge how well students can use what they know in different situations and how they apply one task to the performance of another, such as using the reading assignment to inform their writing or the listening section to inform their speaking (Frost, Elder, and Wigglesworth 2012; Plakans and Gebril 2013). As in DA, IPAs also include feedback so students can see how they are progressing in various areas of language learning. When assigning a grade for an IPA, carefully designed rubrics tied to authentic tasks are the recommended form of evaluation (Montgomery 2002). Even though this model is quite different from traditional language testing, research suggests students' perception of the holistic and integrated approach of IPAs is generally positive (Altstaedter and Krosl 2018; Martel 2018).

Feedback

Whether one is discussing DA, PBA, or IPAs as strategies to encourage students to progress in the ZPD, feedback is an essential component. The feedback needs to come throughout the learning process, not merely at the end with the assessment. Formative assessments, not only summative, should be used regularly to help students gauge their own progress (Adair-Hauck and Troyan 2013; Hattie and Timperley 2007; Warin 2008).

According to Hattie and Timperley (2007), “To be effective, feedback needs to be clear, purposeful, meaningful, and compatible with students’ prior knowledge and to provide logical connections” (102). Waring (2008) took this assumption a step further and said that by solely providing unspecific praise, the teacher essentially “shuts down” any additional attempts the student may have made at communication and stifles the learning process. Clearly, this is not the goal in any DA model. Lantolf and Poehner (2011) pointed to the struggle the learner must go through in order to develop in the target language, which comes from feedback, reflection, and ultimately growth. In short, it is the quality of feedback that matters.

One of the challenging aspects to feedback is that it can be difficult to provide the best type in every situation that classroom teachers are likely to encounter (Lyster and Ranta 1997; Lyster, Saito, and Sato 2013; Rassaei 2014). For example, the great “implicit versus explicit” feedback and error correction debate continues as research has not conclusively shown one to be more universally advantageous than the other in all situations. Explicit feedback may be better if the goal is to increase student language production and pronunciation development (Ellis et al. 2009; Saito 2013). However, implicit feedback in the form of “recasts,” or restating the incorrect sentence to the student to draw attention to the error, has been shown to have advantages with long-term L2 acquisition (Ellis 2008; Li 2010). In addition, recasts that cause students to notice their own errors could help move them along their individual ZPD, as the teacher feedback served the goal of helping students discover something about the language on their own (Lantolf and Poehner 2011). Because feedback is such an essential part of the DA model, it is important for teachers to keep feedback types and strategies in mind.

Weaknesses

Even though IPAs tend to reward the student for what they know instead of penalizing them for what they don’t, there are still some challenges to be considered (Barkaoui, Brooks, Swain, and Lapkin 2013). For example, if a student does not fully understand the reading or listening section, which generally comes first in the IPA process, then the subsequent sections may show lower scores since the tasks build upon prior sections (Jamieson, Eignor, Grabe, and Kunnan 2011). On the other hand, university language instructors reported concerns about grade inflation when awarding points for completing tasks instead of docking points for grammatical inaccuracy (Martel and Bailey 2016).

Integrated tasks may also measure not only language production skills, but cognitive components as well, such as the ability to organize, remember, and appropriately use information from one task to the next (Frost et al. 2012; Skehan 2009). This is especially true when the IPAs take many days to complete.

Students who struggle with grasping material and fully understanding each task may have a difficult time holding a specific concept in their mind for days at a time. The reading or listening activity, which generally comes first, may not actually be helpful in the “integrated” part of this assessment if the student cannot remember it.

The need for vertical alignment becomes clear when one examines the use of IPAs in elementary school all the way up to college classes. IPAs used for elementary school L2 learners can be valuable to help students to progress in the target language (Davin, Troyan, Donato, and Hellman 2011). Although many elementary language programs focus on oral proficiency, which is reflected in student performance on IPAs, this type of assessment serves an important purpose in helping teachers know where to focus their efforts in the future. In addition, there should be a seamless connection between what is done in elementary, middle, and high schools, and beyond that, what is done in college language classrooms. Research suggests, however, this may not always be the case (Glisan, Uribe, and Adair-Hauck 2007; Ricardo-Osorio 2008; Zapata 2016).

Communicative Mode Imbalance

With the current trend of IPAs in the L2 classroom, it appears each section of the assessment is not always given equal attention. The emphasis on “communicative tasks” such as writing a thank-you note or completing an application comes at the cost of teaching students to produce extended discourse in academic writing (Bernhardt, Molitoris, Romeo, Lin, and Valderrama 2015). Studies show a positive correlation between writing proficiency and oral proficiency in the target language (Hubert 2013; Rifkin 2005), which suggests the need for the writing portion of the IPA to receive as much attention as the speaking portion. In addition, research suggests middle and high school teachers tend to focus more on interpretive reading and presentational writing than on speaking and listening (Kaplan 2016; Kissau and Adams 2016), while elementary programs tend to focus on oral proficiency (Davin et al. 2011), further illustrating this imbalance.

One problem with judging oral proficiency of students is a lack of training for the teachers doing the assessing (Glisan and Foltz 1998; Malone 2013; Malone and Sandrock 2016). The ACTFL guidelines do, to be fair, include descriptions of what students should be able to do at each level of language development. For example, speakers at the Novice High sublevel “are able to handle a variety of tasks pertaining to the Intermediate level, but are unable to sustain performance at that level. They are able to manage successfully a number of uncomplicated communicative tasks in straightforward social situations” (ACTFL 2012: 8). Perhaps, though, the definition of “uncomplicated communicative tasks” might vary from teacher to teacher, and perhaps the threshold for how long one must “sustain performance at that level” is easier or harder for

different teachers. Teachers may also have a difficult time encouraging students to participate in non-rehearsed conversations, as factors such as whether or not it is a teacher interview or peer conversation can influence students' language production (Tsou 2005; Sandlund, Sundqvist, and Nyroos 2016).

Teacher Resistance

Even with all of the benefits to DA, PBA, and IPAs in the language classroom, some teachers may be reluctant to abandon the traditional way of testing. One reason teachers are hesitant to embrace IPAs is the perception that these assessments require too much class time (Adair-Hauck et al. 2006; Glisan et al. 2007; Kaplan 2016). Even though DAs help students to progress through the ZPD with feedback and guidance from their teacher, this is generally carried out through one-on-one interactions with each student, which is time-consuming and may not be practical in large classes (Antón 2009). In fact, while most of the research and implementation of IPAs is focused on middle and high schools, post-secondary classrooms may also be slow to embrace the IPA model for these same reasons. The problem isn't that teachers are not in favor of this type of assessment; even teachers who believe in a communicative approach to teaching and assessment still face challenges such as limited resources and large class sizes that can make implementing this approach difficult (Bell 2005; Borg 2003; Kissau, Algozzine, and Yon 2012). The problem is in the logistics.

Another challenge with the IPA model is the degree to which the assessment is mimicking authentic interactions and tasks. For the interpretive task, it can be time-consuming for teachers to find appropriate authentic materials, particularly with novice-level students (Gilmore 2007; Simonsen 2019). In addition, even though performance assessments are meant to be as authentic as possible with a real-world context, students still employ various strategies to take these assessments and navigate the language. They do not simply pretend they are actually in a real-world situation, but rather, they use cognitive, metacognitive, and affective strategies in addressing the tasks before them (Huang 2010; Nasab and Motlagh 2015; Seker 2016). Perhaps, even though the goal is to promote communication in the target language by mirroring what is done in real life, the IPA is not quite as "authentic" as we may have believed (Spence-Brown 2001).

Finally, if the IPA is meant to be used as an assessment, teachers want to see how their students are performing in various aspects of the language. On more traditional tests, there is a standardization against which all students are measured, which is important for teachers giving common assessments (Martel and Bailey 2016). In DA models, the student's abilities fluctuate and, through the ZPD, increase with the support of the interaction with the teacher or peers. In fact, according to Lantolf and Poehner (2007), "In what to some is no doubt a curious turn, the more reliable the procedure, the less effective it is in promot-

ing individual development” (67). Standardized tests may give more accurate scores, but do not address students’ ZPD. DA does address the ZPD, certainly, but may not always be an accurate reflection of what the student can do in the target language.

Discussion

The IPA is a recommended method of assessment by ACTFL and has many benefits in helping students progress in the target language. It relies on the DA model of feedback, the PBA principles of real-world contexts, and addresses students’ ZPD. Even with the challenges associated with implementing IPAs in the classroom, there are ways teachers can adapt and modify this type of assessment to meet the needs of their students.

Time Constraints

One of the main complaints of IPA implementation is the time-intensive nature of conducting each section. Even teachers who agree with a proficiency-based approach to language teaching choose assessments such as true/false or multiple choice in order to manage the burden of grading (Huhn 2013; Kaplan 2016). There must be a trade-off somewhere. If teachers, for example, give all three sections of the IPA in the same class period, they lose the feedback loop that comes with grading and returning each section separately. Yet if teachers administer the IPA over multiple days, they feel overwhelmed with the time commitment.

I propose separating the formative DA classroom activities from the summative IPA. Constant feedback on formative assessments is one way teachers can help their students increase their language proficiency and interact in authentic contexts with the important caveat that it can be done within the class period (Davin et al. 2017). These assessments could be conversations the teacher has with an individual student or a group of students. Assessments could also be written quizzes or in-class assignments where the teacher provides feedback to which a student must respond. Finally, formative assessments such as the aforementioned examples could be taken for a grade or they could be part of daily class instruction. Teachers should not feel as if DA is only possible with lengthy, summative assessments.

The body of research is clear on why IPAs are important, yet in-depth exploration of solutions to the logistical challenges is glaringly absent in the academic literature. However, practicing teachers have filled in the gaps by sharing their own coping strategies. For example, professional conferences have long had a great number of sessions geared towards this very topic. One of these sessions recommended teachers develop a “flow” to administer the IPA within a single class period (Reschly, Schenck, Carlson, and Howard 2016). Under this model,

each piece of the IPA is given at once. Students are given the reading portion with comprehension questions, along with a writing prompt that aligns with the topic of the authentic resource. Students can choose which section to do first and use the reading to inform their writing. In the meantime, students participate in oral interviews at the teachers' desk while everyone else works on the other two parts. While one student (or pair of students) is being interviewed and the teacher completes the evaluation rubric, the following student waits in the wings, immediately ready to be interviewed the moment the previous student finishes. In this manner, it is possible for 30 students to complete all three portions of an IPA within a 90-minute block.

Technology Solutions

Authenticity can be challenging in a classroom setting, but that doesn't mean teachers should abandon the premise of setting a context and establishing the reason for using the language in a particular scenario. One way to help with the daunting task of finding and implementing authentic resources is for teachers to use banks of resources available online that have already been curated and organized for language learners. Kissau and Adams (2016) mentioned the Center for Open Educational Resources and Language Learning at the University of Texas at Austin with free authentic listening resources (laits.utexas.edu) as well as yabla.com and fluentu.com. In addition, websites such as pinterest.com are easily searchable and teachers from all over the world contribute to building banks of resources for thematic units. In the Spanish classroom, zachary-jones.com/zambombazo and speakinglatino.com are but two examples of extensive collections of authentic resources with accompanying activities appropriate for students of all levels. Teachers can integrate these resources daily as part of their classroom activities in addition to selecting resources specifically for an IPA.

Technology can also be leveraged to address the challenges of providing sufficient feedback to individual students (Chun 2016; Blake 2016). Even in large classrooms with too many students for the teacher to give sufficient feedback in speaking interactions, the teacher can still engage and evaluate students' oral proficiency through videoconferences with programs such as Skype, or asynchronous conversations with programs such as VoiceThread. If a teacher struggles to provide feedback on written assignments, she can allow students to write collaboratively in programs such as Google Docs and address multiple students on one written product. These types of solutions can be applied to both formative assessments conducted throughout the unit as well as summative IPAs at the end. By design, the IPA allows for freedom in development and implementation as teaching contexts and challenges can vary widely.

Conclusion

This review explored the theoretical underpinnings of the IPA to highlight the principles of DA and PBA. Research demonstrates the benefits of such an approach, while also recognizing the challenges faced by teachers. Future areas of study should look for meaningful, sustainable ways to address the logistical challenges and time constraints that make IPA implementation difficult. In addition, as teachers are already adapting and modifying the IPA to meet their needs as well as curating and utilizing authentic resources, researchers should investigate how the IPA plays out in real life. A balance must be struck between what scholars have determined to be best practices and what classroom teachers are actually able to accomplish.

Works Cited

ACTFL. (2012). *Proficiency Guidelines*. Alexandria, VA: American Council on the Teaching of Foreign Languages.

Adair-Hauck, B., Glisan, E. W., Koda, K., Swender, E. B., and Sandrock, P. (2006). “The Integrated Performance Assessment (IPA): Connecting Assessment to Instruction and Learning.” *Foreign Language Annals* 39.3: 359-382.

Adair-Hauck, B., and Troyan, F.J. (2013). “A Descriptive and Co-constructive Approach to Integrated Performance Assessment Feedback.” *Foreign Language Annals* 46.1: 23-44.

Altstaedter, L. L., and Krosl, M. (2018). “Perceptions of Integrated Performance Assessments Among Beginning Spanish College Students: A Preliminary study.” *Journal of Foreign Language Teaching and Applied Linguistics* 5.2: 99-114.

Antón, M. (2009). “Dynamic Assessment of Advanced Second Language Learners.” *Foreign Language Annals* 42.3: 576-598.

Bachman, L. F. (2002). “Some Reflections on Task-based Language Performance Assessment.” *Language Testing* 19.4: 453-476.

Barkaoui, K., Brooks, L., Swain, M., and Lapkin, S. (2013). “Test-takers’ Strategic Behaviors in Independent and Integrated Speaking Tasks.” *Applied Linguistics* 34.3: 304-324.

Bell, T. R. (2005). “Behaviors and Attitudes of Effective Foreign Language Teachers: Results of a Questionnaire Study.” *Foreign Language Annals* 38.2: 259-270.

Bernhardt, E., Molitoris, J., Romeo, K., Lin, N., and Valderrama, P. (2015). “Designing and Sustaining a Foreign Language Writing Proficiency Assessment Program at the Postsecondary Level.” *Foreign Language Annals* 48.3: 329-349.

Blake, R. (2016). “Technology and the Four Skills.” *Language Learning & Technology* 20.2: 129-142.

Borg, S. (2003). “Teacher Cognition in Language Teaching: A Review of Research on What Language Teachers Think, Know, Believe, and Do.” *Language Teaching* 36.2: 81-109.

Byrnes, H. (2002). “The Role of Task and Task-based Assessment in a Content-oriented Collegiate Foreign Language Curriculum.” *Language Testing* 19: 419-437.

Chun, D. M. (2016). "The Role of Technology in SLA Research." *Language Learning & Technology* 20.2: 98-115.

Davin, K. J. (2013). "Integration of Dynamic Assessment and Instructional Conversations to Promote Development and Improve Assessment in the Language Classroom." *Language Teaching Research* 17.3: 303-322.

Davin, K. J., Herazo, J. D., and Sagre, A. (2017). "Learning to Mediate: Teacher Appropriation of Dynamic Assessment." *Language Teaching Research* 21.5: 632-651.

Davin, K. J., and Troyan, F. J. (2015). "The Implementation of High-leverage Teaching Practices: From the University Classroom to the Field Site." *Foreign Language Annals* 48.1: 124-142.

Davin, K., Troyan, F. J., Donato, R., and Hellman, A. (2011). "Research on the Integrated Performance Assessment in an Early Foreign Language Learning Program." *Foreign Language Annals* 44.4: 605-625.

Donato, R. (2000). "Sociocultural Contributions to Understanding the Foreign and Second Language Classroom." *Sociocultural Theory and Second Language Learning*. Ed. J. P. Lantolf. New York: Oxford University Press. 27-50.

Ellis, R. (2008). *The Study of Second Language Acquisition*. 2nd edition. Oxford: Oxford University Press.

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philip, J., and Reinders, H. (2009). *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*. Bristol: Multilingual Matters.

Frost, K., Elder, C., and Wigglesworth, G. (2012). "Investigating the Validity of an Integrated Listening-Speaking Task: A Discourse-based Analysis of Test Takers' Oral Performances." *Language Testing* 29.3: 345-369.

Gilmore, A. (2007). "Authentic Materials and Authenticity in Foreign Language Learning." *Language Teaching* 40.2: 97-118.

Glisan, E. W., and Foltz, D. A. (1998). "Assessing Students' Oral Proficiency in an Outcome-based Curriculum: Student performance and teacher intuitions." *Modern Language Journal* 82.1: 1-18.

Glisan, E. W., Uribe, D., and Adair-Hauck, B. (2007). "Research on Integrated Performance Assessment at the Post-secondary Level: Student Performance across the Modes of Communication." *Canadian Modern Language Review* 64.1: 39-67.

Hattie, J., and Timperley, H. (2007). "The Power of Feedback." *Review of Educational Research* 77.1: 81-112.

Huang, L. S. (2010). "Do Different Modalities of Reflection Matter? An Exploration of Adult Second-language Learners' Reported Strategy Use and Oral Language Production." *System* 38.2: 245-261.

Hubert, M. D. (2013). "The Development of Speaking and Writing Proficiencies in the Spanish Language Classroom: A Case Study." *Foreign Language Annals* 46.1: 88-95.

Huhn, C. (2013). "Evaluating Effective Teaching in the 21st Century World Language Classroom." *Dimension* 104-119.

Jamieson, J. M., Eignor, D., Grabe, W., and Kunnan, A. J. (2011). "Frameworks for a New TOEFL." *Building a Validity Argument for the Test of English as a Foreign Language*. Ed. C. A. Chapelle, M. K. Enright, and J. M. Jamieson. New York: Routledge. 55-96.

Kaplan, C. S. (2016). "Alignment of World Language Standards and Assessments: A Multiple Case Study." *Foreign Language Annals* 49.3: 502-529.

Kissau, S., Algozzine, B., and Yon, M. (2012). "Similar but Different: The Beliefs of Foreign Language Teachers." *Foreign Language Annals* 45.4: 580-598.

Kozulin, A., and Garb, E. (2002). "Dynamic Assessment of EFL Text Comprehension." *School Psychology International* 23.1: 112-127.

Lantolf, J. P., and Poehner, M. E. (2007). "Dynamic Assessment of L2 Development: Bringing the Past into the Future." *Journal of Applied Linguistics and Professional Practice* 1.1: 49-72.

Lantolf, J. P., and Poehner, M. E. (2011). "Dynamic Assessment in the Classroom: Vygotskian Praxis for Second Language Development." *Language Teaching Research* 15.1: 11-33.

Lantolf, J. P., and Poehner, M. E. (2013). "The Unfairness of Equal Treatment: Objectivity in L2 Testing and Dynamic Assessment." *Educational Research and Evaluation* 19.2-3: 141-157.

Leung, C. (2007). "Dynamic Assessment: Assessment for and as Teaching?" *Language Assessment Quarterly* 4.3: 257-278.

Li, S. (2010). "The Effectiveness of Corrective Feedback in SLA: A Meta-analysis." *Language Learning* 60.2: 309-365.

Lyster, R., and Ranta, L. (1997). "Corrective Feedback and Learner Uptake: Negotiation of Form in Communicative Classrooms." *Studies in Second Language Acquisition* 19.1: 37-66.

Lyster, R., Saito, K., and Sato, M. (2013). "Oral Corrective Feedback in Second Language Classrooms." *Language Teaching* 46.1: 1-40.

Malone, M. E. (2013). "The Essentials of Assessment Literacy: Contrasts between Testers and Users." *Language Testing* 30.3: 329-344.

Malone, M., and Sandrock, P. (2016). "Assessment Issues in World Languages." *Meeting the Challenge to Measurement in an Era of Accountability*. Ed. H. Braun. New York: Routledge. 217-235.

Martel, J. (2018). "Postsecondary Students' and Instructors' Evaluative Comments about ACTFL's Integrated Performance Assessment." *Applied Language Learning* 28.1: 1-18.

Martel, J., and Bailey, K. M. (2016). "Exploring the Trajectory of an Educational Innovation: Instructors' Attitudes toward IPA Implementation in a Postsecondary Intensive Summer Language Program." *Foreign Language Annals* 49.3: 530-543.

Montgomery, K. (2002). "Authentic Tasks and Rubrics: Going Beyond Traditional Assessments in College Teaching." *College Teaching* 50.1: 34-40.

Nasab, M. S., and Motlagh, S. F. (2015). "A Complete Review for Metacognitive, Cognitive, and Social/Affective Strategies as Essential Components of Learning Strategies and their Relationships with EFL Learners' Reading Comprehension Promotion." *Advances in Language and Literary Studies* 6.3: 166-184.

Norris, J. M., Brown, J. D., Hudson, T. D., and Bonk, W. (2002). "Examinee Abilities and Task Difficulty in Task-based Second Language Performance Assessment." *Language Testing* 19.4: 395-418.

Plakans, L., and Gebril, A. (2013). "Using Multiple Texts in an Integrated Writing Assessment: Source Text Use as a Predictor of Score." *Journal of Second Language Writing* 22.3: 217-230.

Rassaei, E. (2014). "Scaffolded Feedback, Recasts, and L2 Development: A Sociocultural Perspective." *The Modern Language Journal* 98.1: 417-431.

Reschly, J., Schenck, S. M., Carlson, E., and Howard, C. (2016, November). "Kick Back with a Cool IPA." Session presented at the Annual Convention and World Languages Expo of the American Council on the Teaching of Foreign Languages, Boston, MA.

Ricardo-Osorio, J. G. (2008). "A Study of Foreign Language Learning Outcomes Assessment in US Undergraduate Education." *Foreign Language Annals* 41.4: 590-610.

Rifkin, B. (2005). "A Ceiling Effect in Traditional Classroom Foreign Language Instruction: Data from Russian." *The Modern Language Journal* 89.1: 3-18.

Saito, K. (2013). "Reexamining Effects of Form-focused Instruction on L2 Pronunciation Development: The Role of Explicit Phonetic Information." *Studies in Second Language Acquisition* 35.1: 1-29.

Sandlund, E., Sundqvist, P., and Nyroos, L. (2016). "Testing L2 Talk: A Review of Empirical Studies on Second-Language Oral Proficiency Testing." *Language and Linguistics Compass* 10.1: 14-29.

Seker, M. (2016). "The Use of Self-regulation Strategies by Foreign Language Learners and its Role in Language Achievement." *Language Teaching Research* 20.5: 600-618.

Simonsen, R. (2019). "An Analysis of the Problematic Discourse Surrounding 'Authentic Texts.'" *Hispania* 102.2: 245-258.

Skehan, P. (2009). "Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis." *Applied Linguistics* 30.4: 510-532.

Spence-Brown, R. (2001). "The Eye of the Beholder: Authenticity in an Embedded Assessment Task." *Language Testing* 18.4: 463-481.

Tavakoli, P., and Skehan, P. (2005). "Strategic Planning, Task Structure, and Performance Testing." *Planning and Task Performance in a Second Language*. Ed. R. Ellis. Philadelphia: John Benjamins Publishing Company. 193-216.

Tsou, W. (2005). "Improving Speaking Skills through Instruction in Oral Classroom Participation." *Foreign Language Annals* 38.1: 46-55.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.

Waring, H. Z. (2008). "Using Explicit Positive Assessment in the Language Classroom: IRF, Feedback, and Learning Opportunities." *The Modern Language Journal* 92.4: 577-594.

Zapata, G. C. (2016). "University Students' Perceptions of Integrated Performance Assessment and the Connection between Classroom Learning and Assessment." *Foreign Language Annals* 49.1: 93-104..