



Castledown

 OPEN ACCESS

# Language Education & Assessment

ISSN 2209-3591

<https://www.castledown.com/journals/lea/>

*Language Education & Assessment*, 2 (2), 57–69 (2019)

<https://doi.org/10.29140/lea.v2n2.105>

## Washback of ACTFL's Integrated Performance Assessment in an Intensive Summer Language Program at the Tertiary Level



JASON MARTEL <sup>a</sup>

<sup>a</sup> *Middlebury Institute of International Studies at Monterey, USA*

Email: [jmartel@middlebury.edu](mailto:jmartel@middlebury.edu)

### Abstract

Washback, that is, the effect of tests on teaching and learning, has captured the interest of scholars in applied linguistics for almost 30 years. However, most research has been conducted on large-scale tests, rather than classroom-based language assessments like ACTFL's Integrated Performance Assessment (IPA). The present study explored washback of the IPA on teachers of Arabic, Chinese, and Spanish in a higher education intensive summer language program. Through a series of interviews, participants explained ways in which the IPA influenced not only their practices, but also their thinking about language teaching. Directions for future research and considerations for changes to IPA implementation are shared based on these findings.

**Keywords:** classroom-based assessment, washback, Integrated Performance Assessment

### Introduction

In the 1990s, the American Council on the Teaching of Foreign Languages (ACTFL) spearheaded the development of a set of national standards for foreign language learning in the United States (National Standards Collaborative Board, 2015; henceforth, the Standards). Recently refreshed, the Standards are comprised of five main goal areas, known as the “Five Cs”: communication, cultures, connections, comparisons, and communities. These standards are meant to shape foreign language curricula; for example, the first culture standard asks teachers to design activities in which learners have the opportunity to use language “to investigate, explain, and reflect on the relationship between the practices and perspectives of the cultures studied” (National Standards Collaborative Board, 2015, Relating Cultural Practices to Perspectives, para. 1). A concise description of the Standards can be found on ACTFL's website at <https://www.actfl.org/sites/default/files/publications/standards/World-ReadinessStandardsforLearningLanguages.pdf>.

**Copyright:** © 2019 Jason Martel. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within this paper.

Soon after the Standards were released, scholars with ACTFL developed the Integrated Performance Assessment (IPA) as a way of tracking learner progress in relation to the Standards, notably in the communications goal area. The IPA is a robust framework that includes specifications for a summative assessment of language ability, as well as recommendations for how to create corresponding thematic curricular units through a process of backward design (see Wiggins & McTighe, 2005). The IPA consists of three tasks, in alignment with the three communication standards: (a) an interpretive task, which assesses learners' ability to comprehend authentic texts both literally and figuratively; (b) an interpersonal task, which assesses learners' ability to engage in a spontaneous, negotiated conversation with an interlocutor involving an information gap; and (c) a presentational task, which assesses learners' ability to deliver a message to an authentic audience in writing or by speaking (Adair-Hauck, Glisan, & Troyan, 2013; National Standards Collaborative Board, 2015). The IPA allows for connections with other standards beyond communication; for instance, learners may be asked to deliver an oral presentational task to users of the target language in their town or city, fulfilling the communities standard. Teachers are the principal designers of the IPAs they use in their classrooms.

As a classroom-based, standards-based assessment, the IPA and its underlying constructs (i.e., interpretive, interpersonal, and presentational communication) are in tune with current research in second language acquisition. However, if adopted in a language program, the IPA may not align with teachers' curricular practices, which may engender a variety of effects, such as teachers changing their practices to be more IPA-like or conversely even resisting the IPA. From an administrative perspective, the former may be conceived of as a positive process, and the latter, a negative one (although note that teachers may not see these processes as such). Language program administrators might thus choose to use the IPA as a tool to shape the teaching in their programs to be more standards-based. According to Andrews (2004), "attention has increasingly been paid to the possibility of turning the apparently powerful effect of tests to advantage and using it to exert a positive influence in support of curriculum innovation" (p. 39). It should be noted that change is needed in several aspects of foreign language education in light of (a) data showing that teachers have not yet fully implemented the Standards as they were conceptualized (ACTFL, 2011), and (b) generally lackluster student outcomes (Center for Applied Second Language Studies, 2010). In particular, oral interpersonal communication, one of the IPA's three principal constructs, "has not been integrated effectively into language classrooms, largely because teachers do not understand the features of interpersonal communication or how it develops in learners" (Glisan, 2012, p. 516).

It would therefore be helpful to understand the effect that the IPA, principally as a framework for summative testing, has on teachers and their teaching, a process known in the language assessment literature as *washback* (Alderson & Wall, 1993; Bailey, 1996; Bailey & Curtis, 2015; Cheng, Watanabe, & Curtis, 2004; Cheng, Sun, & Ma, 2015). Such knowledge could verify whether the IPA can indeed serve as a mechanism for helping teachers to better understand and more faithfully implement the Standards (Davin, Troyan, Donato, & Hellman, 2011). Some scholars have claimed that "research on the Integrated Performance Assessment has revealed a positive and constructive 'washback effect' on instruction; i.e., it can inform and improve the curriculum, teaching and learning practices beyond the test" (Adair-Hauck *et al.*, 2013, p. 103). However, as only one study has reported empirical data on the topic – Adair-Hauck, Glisan, Koda, Swender, and Sandrock, 2006 – this claim remains somewhat premature.<sup>1</sup> Furthermore, Adair-Hauck *et al.*'s (2006) study was conducted in primary and secondary schools, meaning that additional data are needed to determine whether the IPA is an effective change mechanism in post-secondary contexts as well. The research questions for this study are as follows:

---

<sup>1</sup> Despite claiming positive washback, Davin *et al.*'s (2011) study did not provide any supporting data.

- Was there washback of the IPA on new instructors in the summer language program studied?
- If so, how was it manifested?

## Literature Review

### Washback in Language Assessment

Washback is defined as “the effect a test has on teaching and learning” (Bailey & Curtis, 2015, p. 3). Scholars in applied linguistics became interested in washback in the late 1980s, in conjunction with a set of hypotheses articulated by Alderson and Wall (1993), who posited that:

1. A test will influence teaching.
2. A test will influence learning.
3. A test will influence what teachers teach; and
4. A test will influence how teachers teach; and by extension from (2) above,
5. A test will influence what learners learn; and
6. A test will influence how learners learn.
7. A test will influence the rate and sequence of teaching; and
8. A test will influence the rate and sequence of learning.
9. A test will influence the degree and depth of teaching; and
10. A test will influence the degree and depth of learning.
11. A test will influence attitudes to the content, method, etc., of teaching and learning.
12. Tests that have important consequences will have washback; and conversely
13. Tests that do not have important consequences will have no washback.
14. Tests will have washback on all learners and teachers.
15. Tests will have washback effects for some learners and some teachers, but not for others. (pp. 120–121)

Developed by the researchers while studying the influence of English testing in Sri Lanka, these hypotheses propose that language tests provoke a range of behavioral and cognitive effects related to several aspects of the language teaching and learning process, concerning teachers and learners alike. Research on washback, which focuses on the effects of tests in classrooms, represents one form of inquiry within the greater realm of testing consequences; other foci include impact, which represents the effect of tests at a larger societal level (Cheng, Sun, & Ma, 2015). It is important to consider that washback can be perceived as positive or negative, and as Cheng *et al.* (2015) note, such characterizations are “likely defined by stakeholders, possibly differentially, as they see how a test serves its purpose from their points of view” (p. 437).

Most of the research conducted on washback in applied linguistics has dealt with the effects of large-scale language tests, such as Shohamy, Donitsa-Schmidt, and Ferman’s (1996) study on national Arabic and English tests in Israel. Research on washback of classroom-based language assessments has been far less common; in their recent literature review, Cheng *et al.* (2015) identified only four classroom-based washback studies from the period 1993–2013. Although few in number, these studies have demonstrated that “the effects of classroom-based assessment seem to function in a similar way to those of large-scale testing” (p. 461). As stated above, only one study (Adair-Hauck *et al.*, 2006) has reported empirical data pertaining to washback of the IPA, a classroom-based language assessment.

### Research on the IPA and Washback

Adair-Hauck *et al.*’s (2006) study is the only one to date to address washback of the IPA.<sup>2</sup> Their study

---

<sup>2</sup> Adair-Hauck *et al.*’s (2013) manual *Implementing Integrated Performance Assessment* contains testimonials about IPA use in a range of foreign language classrooms, yet these are not empirical data.

was conducted during a government-funded pilot project in the late 1990s whose goals were (a) to develop a prototype IPA, (b) to conduct research on the IPA's feasibility and effectiveness for measuring student learning of the Standards, and (c) to explore washback of the IPA. Thirty foreign language teachers and 1,000 students in grades 3–12 participated in the project, in Chinese, French, German, Italian, and Spanish. Data regarding washback were collected via questionnaires administered to participating teachers, with an *N* of 22 or 23 (it is unclear what the actual response number was). Participants stated that the IPA positively influenced their teaching (89%) and their design of future assessments (91%), although it remains unexplained what “positive” means in this case. It may be that comments provided in the questionnaires such as “reaffirmed effective teaching techniques” (p. 372) fall under this rubric. The comments the researchers elicited reflected a variety of influences of the IPA, many in line with Alderson and Wall's (1993) hypotheses, such as “incorporate the videotape interpersonal assessment” (p. 372), which constitutes an influence on how teachers teach. Challenges associated with using the IPA were also captured, such as “teaching the students how to communicate and ‘think on their feet’ without pre-scripted dialogues” (p. 373). Overall, the researchers determined that “the IPA served as a consciousness-raising technique for standards-based language learning” (p. 373).

It bears mentioning that much remains to be understood regarding the IPA and its use in foreign language classrooms. For example, there are currently no published validation studies. To date, research on the IPA has examined student outcomes (Davin *et al.*, 2011; Glisan, Uribe, & Adair-Hauck, 2007); the discourse between an IPA evaluator and a student during a feedback conference (Adair-Hauck & Troyan, 2013); students' and teachers' perceptions of the IPA (Altstaedter & Kroszl, 2018; Martel, 2018; Zapata, 2016); the balance of the modes of communication during IPA implementation (Kissau & Adams, 2016); and teachers' attitudes toward IPA implementation in an intensive summer language program at the tertiary level (Martel & Bailey, 2016).

## Methods

### Context

The present study was conducted in the Summer Intensive Language Program (SILP) at Middlebury Institute of International Studies at Monterey. SILP is an eight-week, non-residential program that offers courses at elementary, intermediate, and sometimes advanced levels in Arabic, Chinese, French, Russian, and Spanish. The program features four-and-a-half hours of classes, optional free tutoring, and three to four hours of homework per evening during weekdays; this in addition to bi-weekly co-curricular activities, such as excursions, cooking classes, and guest lectures. Instructors and students are asked to use their respective target languages 90% of the time during program events, in line with ACTFL's position statement on target language use in the classroom (ACTFL, 2010), meaning there is no 24-hour target language requirement. During the 2017 session, when data were collected for this study, there were five language coordinators, 17 instructors, several support staff, and 141 students. The students in the program are almost exclusively adults (the average age is usually around 26 years old) and mostly come from the United States. They aspire to learn languages for a variety of purposes, such as pursuing an internationally focused master's degree at the Middlebury Institute, completing language requirements at their home institutions, and expanding their professional skill sets, as well as for personal enjoyment.

Proficiency targets for each course are set by the program; for example, zero beginners are pushed to reach Novice High by the end of the eight weeks, based on ACTFL's Proficiency Guidelines. Instructors, in collaboration with their language coordinators, are free to use whichever course materials they prefer, as long as they maintain a primary emphasis on content-based instruction. Since

2015, all instructors are required to use the IPA format for their midterm and final exams (see Martel & Bailey, 2016 for further information about the initial IPA implementation process). These exams occur at the end of weeks four and eight of the program, respectively, and are each worth 15% of students' final grades. Beyond this requirement, instructors individually determine which elements they want to include in their formal grading structures, which normally comprise an amalgam of formative assessments (e.g., weekly quizzes) and effort-based elements (e.g., homework). As such, performance on the IPA influences, but does not necessarily determine, students' final grades. It should also be mentioned that the program's placement test, administered on the first day of student orientation, is IPA-based.

## **Participants**

All 17 instructors from the program's 2017 session were invited to participate in the study. Ultimately, data were analyzed for three participants: one in Arabic, one in Chinese, and one in Spanish. These three instructors were chosen because they had no substantial experience with the IPA before their work in the program. The three participants – Adara, in Arabic, Nian, in Chinese, and Jane, in Spanish (all pseudonyms) – were all highly experienced language teachers who worked in other institutions during the academic year. 2017 was Adara and Nian's first summer with SILP, and Jane's third. Adara and Nian are native speakers of their teaching languages, while Jane's first language is English.

## **Data Collection and Analysis**

Data were collected for this study through interviews on three occasions: during the first week of the program, just before the midterm exam, and just before or just after the final exam. Since the principal investigator served as the program's associate director, interviews were conducted by one of his colleagues in order to minimize potential discomfort and supervisory bias, e.g., the participants shaping their contributions based on the researcher's role as evaluator and hiring manager (see Merriam & Tisdell, 2016). Interviews were conducted in the colleague's office and were captured using a digital voice recorder. Interview questions were designed to address participants' thinking about language assessment and changes they might have made to their practices over the course of the summer in light of the IPA requirement (see Appendix for full lists of questions).

Once collected, the data were transcribed, uploaded to Dedoose, and analyzed according to the following procedure. To begin with, all transcripts were read through, and references to influences on the participants' thinking and teaching were noted. Then, Alderson and Wall's (1993) hypotheses were consulted to create overarching categories, including influences on what teachers teach, influences on how teachers teach, influences on the rate and sequence of teaching, and attitudes toward various aspects of teaching. The influences detected in the initial reading of the transcripts were sorted into these categories where possible, and those that did not fit were used to generate new categories. The three cases were analyzed independently at first and then later pooled together, permitting comparisons across cases. Analysis was thus a multi-stage, iterative process that involved continual rereading of the data and consulting of the literature in order to refine the resulting analysis.

## **Findings**

The categories generated in the analysis of the data are presented in turn in the paragraphs below.

## The What of Teaching

Regarding the what of teaching (i.e., the type of content chosen for their curricula), all three participants either expressed intentions to add activities to their pedagogical repertoires or described activities that they had added over the course of the summer. These activities are captured in Table 1.

**Table 1** IPA-related activities

| Participant | Intention to Add  | Added   |
|-------------|---|---|
| Adara       | Group work<br>Presentations   | Interpretive listening activities<br>Author's perspective questions<br>Role plays   |
| Nian        | Written presentational activities   | Guessing meaning from context questions<br>Inferential questions<br>Oral presentations<br>Interpersonal speaking activities<br>Writing activities |
| Jane        | Oral presentational activities<br>Group recording activities<br>Inferential interpretive activities |   |

Not surprisingly, many of these activities were related to the constructs underlying the IPA. For example, the types of questions Nian added to her curriculum (i.e., guessing meaning from context and inferential questions) correspond directly with elements on the IPA's comprehension guide template for the interpretive task. Interestingly, Jane set goals for adding activities during interviews, but did not identify any specific activities she had added.

Only one participant, Nian, described eliminating activities from her curriculum. She noted that due to time constraints, she had to remove supplementary reading activities that she would have normally given to her students.

## The How of Teaching

Two participants expressed changes to ways in which they delivered their curricula, i.e., the how of teaching. To begin with, Adara modified her approach to focusing on form throughout the summer. During her third interview, she commented:

*I was addressing cultural perspective instead of kind of concentrating on just structure, structure, structure. The structure kind of came about. Yes, there were questions – “Why is this? Why is that?” – but it came about in the form of like the cultural understanding rather than kind of like... so, the structure kind of led itself eventually to be there through context, not just outside of context. (Adara, Interview, August 7, 2017)*

Here, Adara explained how the forms on which she focused related to the encoding of cultural meaning in the texts she had her students interpret. Such an approach is emergent, in tune with students' communicative needs, and contrasts with studying forms based on a pre-determined grammatical syllabus. Beyond focus on form, the IPA also influenced how Adara taught vocabulary, taught listening, asked questions during interpretive activities, and viewed the purpose of writing in the Arabic classroom.

Nian noted that the IPA influenced how she gave feedback on student work. During an exchange in her second interview, she said:

Interviewer: Is there anything else about the effect of the IPA testing on your teaching that you'd like to share with me at this time or comment on?

Nian: Well, I think IPA offers me a great tool to make suggestions to the students. In the past, after each test, when we have academic counseling, I just give students some general suggestions. "Oh, you should practice more" or "You should do translations" or something like that.

[Brief interruption.]

Nian: So, now I think the structure has offered me more insights into analyzing students' weaknesses and strengths.

Interviewer: I understand.

Nian: For example, after the midterm, I will grade each task, I have a form that [the associate director] gave us, so on each task, I will put their grades and my suggestions. So, when I look at the five tasks, I know, "Oh, this is the weakness. This is something students need to work on." So, I give them more targeted suggestions. (Nian, Interview, July 18, 2017)

In this excerpt, Nian noted that learning about the IPA helped her to give more useful feedback to her students. During the same interview, she explained that this was due to studying the criteria associated with the IPA's various tasks (e.g., language function).

### **The Rate and Sequence of Teaching**

Only one participant, Adara, made comments related to the rate and sequence of teaching. She stated during her first interview that she intended to build student vocabulary faster and in her second interview that she would like to cover twelve chapters of the textbook she was working with, despite student complaints about the pace. It was not fully clear, however, whether her comments about pace were influenced principally by the IPA or by the intensive nature of the summer program.

### **Attitudes toward Teaching**

The summer language program's IPA requirement shaped both Adara and Nian's attitudes toward language teaching and learning, notably in regard to assessment. During Adara's first interview, she stated:

*So, the one thing I hate about it is I feel I'm teaching to the test, but not necessarily teaching to the test, because I'm introducing concepts that they need to cover. So, in one way I feel I'm changing how I teach because I know I need to give this test in this particular way, but I'm finding that the students are getting knowledge a little bit more efficiently than if I had just spent time using drills to teach a concept. (Adara, Interview, July 20, 2017)*

Then, during her second interview, she said:

*[The IPA] was just like a standard testing system. I would test what I would teach in class and that's it. For the IPA, this was kind of like my experience with it. At the beginning, I thought that I was teaching for the test but no, I'm not. Today, I don't think*

*I'm teaching for the test. I'm teaching to assess knowledge in general framed in a different way and I feel that I've been enriched by the experience, you know? (Adara, Interview, August 7, 2017)*

In this interesting progression, Adara seemed to become more comfortable over the course of the summer with the IPA as a test to “teach to.” In the first excerpt, she was skeptical, but noted that the IPA was at least to some extent effective, while in the second excerpt, she seemed convinced, seeing the IPA in a rather positive light. It would thus seem that the change in Adara’s thinking was less about standardized testing writ large and more about the IPA framework specifically.

During her third interview, Nian stated:

*I think now I can tell how assessment can guide my teaching. In the past, assessment for me was just a way to force students to put effort in. Every morning, we would have a quiz. The major purpose is to make sure that they reviewed the materials ahead of time. That doesn't really tell me anything about my teaching. It doesn't really inform anything, just to see if the students are hardworking. But now I think IPA gives me a very comprehensive way of evaluating students' proficiency and to tell me which area I can design more activities and put more time in it. (Nian, Interview, August 7, 2017)*

Here, Nian indicated that the IPA shifted her attitude toward assessment. She moved from seeing it primarily as a way of pressuring students to study to seeing it as tool for generating useful information about student learning, which can in turn inform future teaching.

### **Additional Influences**

Besides the categories listed above, the participants highlighted further effects related to the program’s IPA requirement. Two participants, Adara and Nian, noted an increase in their knowledge about language teaching and learning. Adara characterized the summer as “the biggest professional development session of my life” (Adara, Interview, August 7, 2017). She also considered how she might bring the IPA to her home teaching context, stating:

*I'm grateful because I felt that I'm learning more about this structure so that I can carry with me to my regular classroom later, even though it's not intensive and we meet only three hours for a three credit course, but I can now figure out how to organize it in a sense to see if I can get the students to have a task, a theme, rather than something else. (Adara, Interview, July 20, 2017)*

Nian said that she had learned new terminology and rethought her conceptualization of interpersonal speaking. She noted, however, that she needed to seek out further knowledge when she stated, “I think I need more training on how to use the rubrics to grade students. I’m a little bit confused about the interpretive tasks because there’s only one rubric and I have two different levels.” (Nian, Interview, August 8, 2017).

In addition, working with the IPA compelled Adara to reflect on the effort required to prepare her classes. She said:

*I felt that I am doing more now with this, but it also made me work harder a little bit because it's my first time with the material. I feel like I need to prepare the situation for them to be able to roleplay. I can't just come in the classroom, because that's what I*



*started to do at the beginning. "Okay, come in the classroom. Today we're going to do this..." and I started writing information on the board and expecting that they will practice it with me, and no, they need more time to process, I realized. So, I started preparing the sheet ahead of time for the situation, giving it to them and then coming back. So, it made me work a little more to prepare for it, but I felt definitely that the time was important for the students to be able to do that. (Adara, Interview, July 20, 2017)*

Here, Adara noted how using the IPA framework required more preparation time not only because it was new to her, but also because it asked students to learn in ways to which she was not accustomed.

Finally, the IPA served a validating function for Jane. During her third interview, she mentioned:

*I don't think that my teaching and strategies changed. I feel like it was just that my teaching was more validated by having an assessment that reflected what I was already doing in the classroom and, in fact, it was kind of weird at the end of today. [...] ...it was really nice to have an assessment that really closely matched what I was doing already, so I didn't change anything. I more just sort of went for it, I feel like, yeah, in my teaching. (Jane, Interview, August 7, 2017)*

Rather than compelling her to change her teaching, the IPA system confirmed for Jane that the curriculum she delivered to her students was effective.

## Discussion

The present study explored washback of ACTFL's IPA on instructors in an intensive summer language program at the tertiary level. The participants shared ways in which the IPA influenced their teaching, including what they chose to teach, how they taught it, the rate at which they taught it, their attitudes toward testing, their knowledge about language teaching and learning, their approach to preparing for teaching, and their feelings about their teaching. Overall, the findings presented here indicate positive washback of the IPA from the participants' perspectives. It should be noted, however, that the IPA had seemingly less of an impact on Jane than on Adara and Nian, which could signal that Jane's practice was more standards-based at the outset of the program than Adara's or Nian's. Nevertheless, the data demonstrate that the IPA can indeed "inform and improve the curriculum, teaching and learning practices beyond the test" (Adair-Hauck *et al.*, 2013, p. 103) at the higher education level.

The perspectives captured in this study mirror many of those from Adair-Hauck *et al.*'s (2006) study, the only other to have reported empirical data about washback of the IPA. The participants in both studies felt generally good about the changes to their practices and thinking, which fell largely in line with the Standards. Participants in both studies also experienced a similar consciousness-raising effect of the IPA, notably in setting intentions to add certain types of activities to their pedagogical repertoires. In terms of challenges, participants in both studies stated that it was difficult to convert the rubrics into letter grades. This continued difficulty indicates that it may be time for the creators of the IPA framework to design a feedback strategy that is more accessible to its users.

From a practical standpoint, language program administrators (LPAs) could consider using the IPA as a way of fostering standards-based instruction among faculty. However, as Fournier-Kowaleski (2005) notes about curricular change, "for it to occur, more is required than simply integrating a test into a curriculum" (p. 46). This being the case, LPAs who bring the IPA to their programs should carve out ample time during the workweek for instructors to discuss IPA-oriented design and instructional

practices (say, during a weekly faculty meeting) and to observe each other implementing those practices. LPAs should also make sure to provide adequate formative feedback on instructors' teaching with the IPA in mind, in addition to the evaluative data they need to collect for hiring purposes. Finally, LPAs could work with instructors on the level of their identities as they relate to the IPA as an innovative framework in language education, as hinted to in Adara's comments about teaching to the IPA. Such identity work could be stimulated through discussion questions such as "What do you value when it comes to assessing students' language abilities?" and "What is the role of innovation in your work as a language teacher?" (see Martel, 2017).

This study is limited by the fact that the principal investigator was also the summer language program's associate director, which could have compelled the participants to provide answers to interview questions that they thought he would have wanted to hear, despite the fact that interviews were conducted by one of his colleagues (Merriam & Tisdell, 2016). The study is also limited in that it used only one source of data (interviews). Subsequent studies on washback of the IPA should include additional sources of data, as recommended by Adair-Hauck *et al.* (2006) and Adair-Hauck *et al.* (2013), notably observations of classroom practice both *before* teachers learn about and implement the IPA and then *while* they are implementing it (Alderson & Wall, 1993). The combination of observations and interviews (as well as other data sources, such as journals) would make stronger and potentially more focused claims about washback of the IPA than were possible in the present and previous studies. Such triangulation could show whether participants' perceptions of their practices are in alignment with their actual behaviors.

Furthermore, those interested in studying washback of the IPA should consider using longitudinal designs. Longer engagements in the field could reveal whether influences of the IPA on teachers' practices and cognitions endure over the long-term and/or change over time, and even across instructional contexts. For example, in this study, Adara wondered how she might bring the IPA back to her home teaching context. Will she? And will the standards-based changes she made during the summer program endure?

## Conclusion

According to Grant Wiggins, "assessment reform is the key...to long lasting school reform" (WGBH, 2003). Using washback as its principal lens, the present study captured evidence of positive washback of the IPA on teachers in the focal summer language program, demonstrating that the IPA can be used to foster reform at the tertiary level, manifested in increased standards-based instruction. As such, the study invites the field to reconsider its generally fraught stance toward teaching to tests, which surfaced in Adara's comments, as well as in previous studies (e.g., Davin *et al.*, 2011; Martel, 2018). If a test like the IPA aligns with current research and theory on language teaching and learning, provides meaningful data on student language use in real-life situations, and offers design flexibility, allowing for teachers to enact agency within a common framework, why not teach to it?

## References

- American Council on the Teaching of Foreign Languages (ACTFL). (2010). *Position statement on use of the target language in the classroom*. Retrieved from <https://www.actfl.org/news/position-statements/use-the-target-language-the-classroom>
- American Council on the Teaching of Foreign Languages (ACTFL). (2011). *Decade of standards: Influence, impact. Survey report*. Alexandria, VA: ACTFL.

- Adair-Hauck, B., Glisan, E., Koda, K., Swender, E., & Sandroock, P. (2006). The Integrated Performance Assessment (IPA): Connecting assessment to instruction and learning. *Foreign Language Annals*, 39 (3), 359–382.
- Adair-Hauck, B., Glisan, E. W., & Troyan, F. J. (2013). *Implementing integrated performance assessment*. Alexandria, VA: ACTFL.
- Adair-Hauck, B., & Troyan, F. J. (2013). A descriptive and co-constructive approach to Integrated Performance Assessment feedback. *Foreign Language Annals*, 46 (1), 23–44 .
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14 (2), 115–129.
- Altstaedter, L. L., & 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.
- Andrews, S. (2004). Washback and curriculum innovation. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing* (pp. 59–72). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bailey, K. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing* 13 (3), 257–279.
- Bailey, K., & Curtis, A. (2015). *Learning about language assessment* (2<sup>nd</sup> ed.). Boston, MA: National Geographic Learning/Heinle.
- Center for Applied Second Language Studies. (2010). *What proficiency level do high school students achieve?* Eugene, OR: University of Oregon.
- Cheng, L., Watanabe, Y., & Curtis, A. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*, 48 (4), 436–470.
- Davin, K., Troyan, F. J., Donato, R., & Hellman, A. (2011). Research on the Integrated Performance Assessment in an early foreign language learning program. *Foreign Language Annals*, 44 (4), 605–625.
- Fournier-Kowaleski, L. (2005). *Depicting washback in the intermediate Spanish language classroom: A descriptive study of teachers' instructional behaviors as they relate to tests* (Unpublished doctoral dissertation). The Ohio State University, Columbus, OH.
- Frey, N., & Fisher, D. (2011). *The formative assessment action plan: Practical steps to more successful teaching and learning*. Alexandria, VA: ASCD.
- Glisan, E. W. (2012). National standards: Research into practice. *Language Teaching*, 45 (4), 515–526.
- Glisan, E. W., Uribe, D., & 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.
- Kissau, S., & Adams, M. J. (2016). Instructional decision making and IPAs: Assessing the modes of communication. *Foreign Language Annals*, 49 (1), 105–123.
- Martel, J. (2018). Postsecondary students' and instructors' evaluative comments about ACTFL's Integrated Performance Assessment. *Journal of Applied Language Learning*, 28(1), 1–18.
- Martel, J. (2017). Identity, innovation, and learning to teach a foreign/second language. In G. Barkhuizen (Ed.), *Reflections on language teacher identity research* (pp. 87–92). New York: Routledge.
- Martel, J., & 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.
- Merriam, S., & Tisdell, E. (2016). *Qualitative research: A guide to design and implementation* (4th ed). San Francisco, CA: Jossey-Bass.

- National Standards Collaborative Board. (2015). *World-readiness standards for learning languages* (4<sup>th</sup> ed.). Alexandria, VA: ACTFL.
- Shohamy, E., Donitsa-Schmidt, S., and Ferman, I. (1996). Test impact revisited: washback effect over time. *Language Testing*, 13 (3), 298–317.
- Wiggins, G., & McTighe, J. (2005). *Understanding by design*. Columbus, OH: Pearson Education, Ltd.
- WGBH. (2003). *Teaching foreign languages K–12: A library of classroom practices*. Web videos. Retrieved from <http://www.learner.org/resources/series185.html>
- Zapata, G. C. (2016). University students' perceptions of Integrated Performance Assessments and the connection between classroom learning and assessment. *Foreign Language Annals*, 49 (1), 93–104.

### Author biodata

**Jason Martel** is Associate Professor of TESOL/TFL at the Middlebury Institute of International Studies at Monterey, where he teaches courses in foreign and second language pedagogy. His publications can be found in *Foreign Language Annals*, *Journal of Applied Language Learning*, *CATESOL Journal*, and *French Review*.

## Appendix: Interview Questions

### First Interview

- What sorts of activities are typical in lessons you teach?
- What sorts of assessment procedures do you usually use in your teaching?
- What is your philosophy about the link between teaching and assessment?
- How do you plan to prepare your students for the IPA-based midterms and finals you will administer this summer?
- How will you prepare them for the specific sections (interpretive, interpersonal presentational) of the midterms and finals?
- Have you use the IPA system before?
  - If no: Given that you haven't used the IPA before working for this program, what is your initial reaction to it?
  - If yes: Since you've used it before, what do you think about it? What impact has it had on your teaching?
- Is there anything else about the midterm and final testing program or about the course in general that you would like to comment on or share with me?

### Second Interview

- Have you made any adjustments to your teaching these past four weeks in anticipation of the IPA-based midterm? If so, which?
- Have you eliminated any activities from class sessions that you had initially planned on incorporating due to the IPA midterm expectation? If so, which activities, and why?
- How has your knowledge of the IPA framework affected the way you plan for and conduct class instruction?
- Are there areas in which you feel you need to change your instruction in order to help students better prepare for demands of the final? If so, which areas, and why?

- Is there anything else about the effect of IPA testing on your teaching that you would like to comment on or share with me?

### **Third Interview**

- Looking back on the summer, did your teaching and assessment strategies change as a result of the administration of the IPA midterm and final? If so, how?
- Having administered two IPA-based tests this summer, what do you believe to be the best way(s) to help the students prepare for them?
- Do you feel like your curriculum adequately prepared students for the midterm and final?
- Has your understanding of the link between teaching and assessment changed this summer? If so, how?
- Is there anything else about the IPA midterm and final testing program or about the course in general that you would like to comment on or share with me?