

# Percepciones de los alumnos acerca de las evaluaciones integrales de desempeño: hacia un modelo de evaluación eficaz enfocado en las destrezas comunicativas

LAURA LEVI ALTSTAEDTER  
*East Carolina University*  
*levialtstaedterl@ecu.edu*

MAURO FALASCA  
*College of Business*  
*East Carolina University*  
*falascam@ecu.edu.*

Recibido: 10 de octubre de 2019 – Aceptado: 5 de noviembre de 2019

**Resumen:** En el campo de la enseñanza de lenguas segundas y extranjeras, las tendencias actuales en diseño curricular se enfocan en la importancia de crear instrumentos que evalúen el progreso del alumno en cuanto al desarrollo de sus destrezas comunicativas en tres modos de comunicación: interpretativo (comprensión auditiva o lectora sin posibilidad de negociación de sentido con el creador del mensaje), presentacional (comunicación escrita u oral sin posibilidad de negociación de sentido con el lector/interlocutor) e interpersonal (comunicación oral o escrita con posibilidad de negociación de sentido con el lector/interlocutor) (The National Standards Collaborative Board 2015). Un tipo de evaluación que permite este enfoque es la evaluación integral de desempeño, cuya efectividad ha sido analizada en estudios recientes (Adair-Hauck, Glisan et al. 2006, Davin, Troyan et al. 2011, Adair-Hauck & Troyan 2013, Zapata 2016, Levi Altstaedter & Krosi 2018, entre otros). Un aspecto de suma importancia en el proceso enseñanza-aprendizaje son las percepciones del alumno, ya que éstas pueden influir en su desempeño en tareas y evaluaciones. Si bien estudios anteriores se han enfocado en el análisis de las percepciones de los alumnos acerca de las evaluaciones integrales de desempeño (Zapata 2016, Levi Altstaedter & Krosi 2018), es necesario ahondar aún más esta línea de investigación. Así, el presente estudio propone un análisis de la relación de variables tales como las percepciones de efectividad de las evaluaciones, las percepciones de ansiedad frente a las evaluaciones, las percepciones del diseño de instrumentos y las percepciones de satisfacción del alumno y ofrece un modelo teórico que presenta la conexión entre dichas variables.

**Palabras clave:** evaluación - destrezas comunicativas – ansiedad percepciones

## **Students' perceptions of Integrated Performance Assessments: Towards an effective assessment model focused on communicative ability**

**Abstract:** In the field of foreign and second language teaching, current trends in curriculum development highlight the importance of creating performance-based assessment instruments focused on the three modes of communication: *interpretive* (aural or reading comprehension without the possibility of negotiating meaning with the creator of the message), *presentational* (written or oral communication without the possibility of negotiating meaning with the reader/interlocutor) and *interpersonal* (oral or written two-way communication with the possibility of negotiating meaning with the reader/interlocutor) (The National Standards Collaborative Board 2015). As previous studies (Adair-Hauck, Glisan et al. 2006, Davin, Troyan et al. 2011, Adair-Hauck & Troyan 2013, Zapata 2016, Levi Altstaedter & Krosi 2018; among others) have shown, Integrated Performance Assessments, are a viable option for this kind of assessment approach. A crucial aspect of the teaching-learning process are students' perceptions, as these can impact their performance on tasks and assessments. Although previous studies have focused on students' perceptions of IPAs, (Zapata 2016, Levi Altstaedter & Krosi 2018), more research is needed on this topic. Therefore, this study investigates the relationship among variables such as perceptions of effectiveness of IPAs, assessment anxiety, perceptions of instrument design, and students' satisfaction, while offering a theoretical model that presents a connection among those variables and offers a theoretical model that includes the connections among those variables.

**Keywords:** assessment - communicative ability – anxiety - perceptions

### **Introducción**

En el proceso de enseñanza-aprendizaje de lenguas segundas y extranjeras, la evaluación juega un papel esencial, no solo para el alumno sino también para el docente. Los resultados de las evaluaciones tienen una función dual: por un lado, le indican al alumno si alcanza o no los objetivos de aprendizaje, y por otro le proporcionan al docente información acerca del logro (o falta de logro) de esos objetivos, a la vez que aportan datos que el docente luego usa para reflexionar acerca del desempeño y progreso de los alumnos y, de ser necesario, hacer ajustes a planificaciones futuras. El diseño de evaluaciones debe enfocarse en indagar lo que el alumno sabe acerca de la lengua y cultura y lo que puede hacer con el idioma en términos de su habilidad de comunicarse. Así, se logra obtener información acerca del progreso del alumno en cuanto a los saberes, destrezas y actitudes que ha desarrollado, y aquello que todavía necesita seguir desarrollando.

Wiggins & McTighe (2005) propusieron un enfoque de diseño curricular centrado en un proceso inverso, en el que en primer lugar se redactan objetivos concretos y medibles, luego se diseña un instrumento de evaluación sumativa que permita obtener evidencia de si el alumno ha logrado (o no) esos objetivos, y por último se procede a diseñar planificaciones, que ofrecerán la formación necesaria para que alumno esté preparado para demostrar lo que ha logrado a través de su desempeño en la evaluación.

En el área de la enseñanza-aprendizaje de lenguas segundas y extranjeras, es esencial diseñar evaluaciones que permitan al alumno demostrar no solo su conocimiento declarativo de la lengua, sino también sus destrezas y habilidades en cuanto al uso de la lengua. Por lo tanto, se debe seleccionar cuidadosamente el tipo de evaluación que se va a diseñar, asegurándose de que condiga con los objetivos fijados. Un tipo de instrumento de evaluación que permite al alumno demostrar su dominio de la lengua en cuanto a la comprensión, presentación de información y comunicación interpersonal son los *Integrated Performance Assessments (IPAs* por sus siglas en inglés) o Evaluaciones Integrales de Desempeño (Adair-Hauck, Glisan et al. 2006, Davin, Troyan et al. 2011, Adair-Hauck & Troyan 2013, Zapata 2016, Levi Altstaedter & Kroszl 2018, entre otros). Estas evaluaciones miden el desempeño del alumno en una serie de tareas comunicativas interrelacionadas: una tarea de interpretación, una tarea de comunicación interpersonal, y una tarea de presentación de información. El presente estudio explora las percepciones de alumnos universitarios en clases de español como lengua extranjera acerca de las evaluaciones integradas de desempeño.

## Estado de la cuestión

En los Estados Unidos, el campo de la enseñanza-aprendizaje de lenguas segundas y extranjeras esta regido mayormente por los Estándares de Preparación Global para el Aprendizaje de Lenguas (*World Readiness Standards for Learning Languages*, por su nombre en inglés) (The National Standards Collaborative Board 2015) y sus predecesores, los Estándares para la Enseñanza de Lenguas Extranjeras en el Siglo 21 (*Standards for Foreign Language Learning in the 21<sup>st</sup> Century*, por su nombre en inglés) (National Standards in Foreign Language Education Project 1999). Estos estándares marcaron un hito de suma importancia en el campo de la enseñanza de lenguas segundas y extranjeras, ya que proponen un marco que presenta pautas en cuanto al desarrollo curricular, el diseño de unidades temáticas, así como la planificación de actividades, tareas, y evaluaciones enfocadas en lo que el alumno puede hacer con la lengua y en sus habilidades interculturales. Así, con un enfoque en cinco Cs, que representan la *Comunicación, Culturas, Conexiones, Comunidades y Comparaciones*, el modelo pretende ofrecer una visión abarcadora del proceso enseñanza-aprendizaje de las lenguas segundas y extranjeras, alejándose de enfoques y métodos basados en la precisión lingüística sin atención a la comunicación funcional ni a las otras cuestiones presentadas en estos estándares.

## Evaluación en el campo de las lenguas segundas y extranjeras

La evaluación es un aspecto crucial en el proceso de enseñanza-aprendizaje, ya que permite al docente recopilar datos relativos a los saberes, destrezas, y actitudes del alumno en cuanto a la lengua y cultura en cuestión. Así, como indica Purpura (2016), la evaluación tiene como objetivo la recopilación sistemática de información y datos que permitan al docente no solo determinar el nivel y alcance de las destrezas, saberes y actitudes del alumno, sino también utilizar esa información para tomar decisiones en cuanto al currículo.

Según el modelo de diseño curricular inverso propuesto por Wiggins & McTighe (2005), las evaluaciones se deben crear inmediatamente después de determinarse los objetivos que se pretende que el alumno alcance, y por ende deben adecuarse en forma y exigencia a los objetivos propuestos. Así, si un objetivo indica que el alumno podrá desarrollar exitosamente una determinada función lingüística, la evaluación debe diseñarse de forma tal que presente un contexto lingüístico y sociocultural en el que el alumno pueda demostrar que puede desarrollarla (o no). Asimismo, si se adopta un currículo alineado con los estándares nacionales, debe existir una conexión directa entre estos y las evaluaciones (Kaplan 2016).

En su análisis de la trayectoria en la profesión en cuanto a la implementación de diferentes modelos de evaluación, Purpura (2016) reconoce el valor de las evaluaciones de destrezas integradas, ya que mediante estas se consigue obtener evidencia más completa de la capacidad del alumno de integrar y sintetizar lo aprendido. Así, las evaluaciones integrales de desempeño cumplen con este supuesto a la vez que presentan una forma de evaluación que permite una articulación con los estándares nacionales.

### Evaluaciones integrales de desempeño

**Origen e importancia.** Las evaluaciones integrales de desempeño surgen en el campo de la enseñanza-aprendizaje de lenguas segundas y extranjeras en los Estados Unidos a partir de la necesidad de diseñar evaluaciones que permitieran medir el progreso del alumno en destrezas alineadas con los estándares nacionales (Adair-Hauck, Glisan et al. 2006). Estos investigadores propusieron un prototipo de evaluación integral de desempeño centrado en los tres modos de comunicación que va más allá de la evaluación de cada destreza por separado, sino que las evalúa de manera integrada.

Asimismo, Adair-Hauck, Glisan et al. (2006) indicaron que las evaluaciones integrales de desempeño se clasifican dentro de las evaluaciones auténticas (Wiggins 1998), ya que: (a) son realistas, (b) requieren que el alumno sea innovador y haga algo concreto con la lengua más allá de repetir o regurgitar la información aprendida, (c) simulan contextos auténticos, (d) miden la capacidad del alumno de integrar todo lo aprendido, y (e) permiten que el alumno reciba retroalimentación y refine su desempeño. De esta manera, las evaluaciones integrales de desempeño cumplen una función dual ya que, por un lado, permiten medir el progreso del alumno en lo relativo a

los estándares nacionales en el campo de la enseñanza-aprendizaje de lenguas segundas y extranjeras en los Estados Unidos de forma integrada y, por otro, permiten establecer una conexión entre un enfoque de enseñanza basado en los estándares y la evaluación (Adair-Hauck, Glisan et al. 2006).

**Estructura.** Según el prototipo propuesto por Adair-Hauck, Glisan et al. (2006), las evaluaciones integrales de desempeño comienzan con un esquema, en el cual se presenta el tema y se contextualizan las tareas que realizará el alumno, al cual le siguen tres tareas unidas por un hilo conductor temático. La primera tarea, una tarea interpretativa, incluye la visualización, lectura o audición de una fuente auténtica, ya sea un texto escrito, un audio o un video, y su posterior análisis e interpretación. En esta tarea se miden destrezas como identificación del significado de palabras y frases clave, identificación de ideas principales y subordinadas, reconstrucción del texto a partir del reacomodamiento de sus partes, inferencias del significado de ciertos vocablos, inferencias de conceptos, así como identificación de perspectivas del autor y perspectivas culturales.

La segunda tarea, la interpersonal, supone la evaluación de la capacidad del alumno de participar de manera exitosa en una situación comunicativa en la que hay posibilidad de negociación de sentido. Esta tarea puede realizarse de forma escrita u oral –aunque normalmente se trata de tareas orales– y debe tener las siguientes características: que la comunicación sea espontánea, es decir, no preparada y memorizada con antelación ni leída; que tenga una función comunicativa; y que contenga un vacío de información. La tercera tarea, la de presentación de información, supone la evaluación de la capacidad del alumno de comunicar, ya sea en forma oral o escrita, de manera exitosa un mensaje en una situación comunicativa en la que no existe posibilidad de negociación de sentido. Es en esta tarea final que el alumno crea un producto original dentro de la temática de la evaluación integral de desempeño, integrando la información y contenido de las otras tareas.

**Percepciones.** Algunos estudios previos acerca de este tipo de evaluación se han centrado en el diseño de los instrumentos utilizados (Adair-Hauck, Glisan et al. 2006), el desempeño de los alumnos en este tipo de evaluación (Glisan, Uribe et al. 2007, Davin, Troyan et al. 2011), retroalimentación en la implementación de este tipo de evaluación (Adair-Hauck & Troyan 2013), y percepciones de docentes (Kissau & Adams 2016, Martel & Bailey 2016, Martel 2018) y alumnos (Zapata 2016, Levi Altstaedter & Kroszl 2018, Martel 2018).

En un reciente estudio enfocado en las percepciones de alumnos universitarios acerca de las evaluaciones integrales de desempeño, Zapata (2016) indicó que en general los alumnos tuvieron percepciones positivas acerca de este tipo de instrumento de evaluación. La investigadora recopiló datos entre alumnos universitarios de español como lengua extranjera en cursos de primer, segundo y tercer nivel. Los datos fueron recabados mediante una encuesta de tres preguntas en escala Likert de cinco niveles. En cuanto a la primera pregunta, si a los alumnos les había gustado el formato del instrumento, los resultados mostraron que en el caso de alumnos de primer y segundo nivel de clases de español como lengua extranjera las percepciones fueron positivas,

pero los alumnos de tercer nivel tuvieron percepciones significativamente inferiores a las de los otros dos niveles. Respecto de la segunda pregunta, si el instrumento se alineaba con las actividades y tareas realizadas en el aula, y la tercera, si las actividades extras asignadas estaban alineadas con el instrumento, se obtuvieron resultados positivos de los tres grupos de alumnos. Asimismo, se recabaron datos adicionales mediante preguntas abiertas. Los resultados cualitativos mostraron diferencias en la implementación; por ejemplo, algunos alumnos indicaron que había una descompaginación entre las actividades y tareas áulicas y la evaluación. Así, la investigadora puntualizó la importancia de que exista fidelidad de implementación y articulación en cuanto a las innovaciones curriculares.

Siguiendo con esta trayectoria de investigación, por medio del presente estudio se propone investigar si existe alguna relación entre las percepciones de efectividad de las evaluaciones, las percepciones de ansiedad frente a las evaluaciones, las percepciones del diseño del instrumento, y las percepciones de satisfacción de alumnos universitarios de español como lengua extranjera en cuanto a las evaluaciones integrales de desempeño.

Por otro lado, Martel (2018) en un estudio realizado en un programa intensivo de verano, recopiló datos entre profesores y alumnos en cursos de español, árabe, chino, francés y ruso como lenguas extranjeras. Los datos de los alumnos fueron recopilados a través de un cuestionario que contenía preguntas abiertas. Los resultados mostraron que a los alumnos les gustó la evaluación en general, especialmente su estructura holística, aunque también se puntualizó el desequilibrio en la dificultad que presentaron las diversas partes de la evaluación. Asimismo, otros alumnos resaltaron que la evaluación integral de desempeño se asemejaba a situaciones reales, aunque otros resaltaron que les habría gustado que se hubiera incluido un componente gramatical y que tenían demasiado contenido. Respecto de este último punto, el autor reconoció que esto podía deberse a su decisión de implementar un instrumento con cinco tareas que incluía tareas interpretativas tanto de audición como comprensión lectora, como tareas de presentación de información tanto de forma oral como escrita, en lugar de utilizar la estructura más común de tres tareas, con solo una para evaluar la capacidad de interpretar información y una para evaluar la capacidad de presentar información.

Levi Altstaedtler & Kroszl (2018) llevaron a cabo otro estudio reciente sobre el tema. En dicho estudio se administró una encuesta a alumnos que tomaban un curso de español básico a nivel universitario basada en la teoría Expectativa-Valor de Eccles & Wigfield (1995), (2002). Los resultados de ese estudio indicaron que las percepciones de los alumnos en cuanto a las evaluaciones integrales de desempeño fueron significativamente positivas, comparadas con el valor neutro de 4. Asimismo, los resultados indicaron que aun aquellos alumnos que se ubicaban en el percentil más bajo (25%) tuvieron percepciones cercanas o superiores a la neutra. Los resultados cualitativos mostraron varios temas identificados por los alumnos en lo relativo a este tipo de evaluación. Así, algunos alumnos indicaron que estas evaluaciones les permitieron ver la lengua en uso en contextos reales y otros resaltaron la viabilidad de

estos instrumentos como medidas de evaluación válidas. Asimismo, otros temas surgidos del análisis fueron la colaboración, la metacognición, y el desafío que presentó este tipo de instrumento.

Si bien estos estudios previos se enfocaron en las percepciones del alumno, a través del presente estudio se propuso ahondar aún más en el campo de las percepciones de los alumnos en cuanto a las evaluaciones integrales de desempeño investigando la relación entre constructos como percepciones de efectividad de las evaluaciones, percepciones de ansiedad frente a las evaluaciones, percepciones del diseño de instrumentos y percepciones de satisfacción del alumno, y ofreciendo un modelo teórico en lo relativo a la conexión entre estos constructos.

## Metodología

### Participantes

Los participantes en este estudio fueron cuarenta y seis alumnos de español como lengua extranjera en una universidad del sureste de los Estados Unidos. Los alumnos cursaban el segundo nivel de español como lengua extranjera. Los alumnos que cursan este nivel generalmente han tomado dos niveles de español en el secundario y/o un nivel en la universidad. A este nivel, el alumno puede comúnmente realizar funciones lingüísticas básicas en tiempo presente, como hablar de su rutina diaria, describir su entorno, incluyendo personas, lugares y cosas, y dar instrucciones, entre otras.

### Desarrollo de la Encuesta y Recolección de Datos

Los datos fueron obtenidos mediante el método de encuesta. Las preguntas, algunas de las cuales fueron adaptadas de Pereira, Niklasson et al. (2017) y Martel (2018), se agruparon en cuatro constructos: percepciones de efectividad de las evaluaciones (EE) percepciones de ansiedad frente a las evaluaciones (AE), percepciones del diseño del instrumento (DI) y percepciones de satisfacción del alumno (SA). La figura número 1 presenta el modelo teórico propuesto.

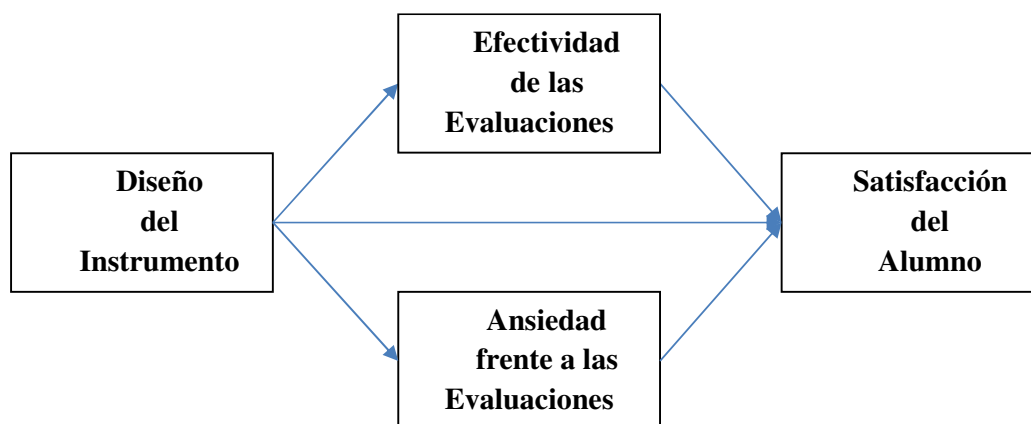


Figura 1. Modelo teórico propuesto

Escalas Likert de siete puntos (de 1 = ‘Totalmente en desacuerdo’ a 7 = ‘Totalmente de acuerdo’) fueron utilizadas para los distintos indicadores. Los indicadores utilizados en la encuesta representan medidas perceptuales de opiniones sobre desempeño, las cuales han demostrado satisfacer requisitos de confiabilidad y validez en la medición de constructos (Ketokivi & Schroeder 2004).

### Análisis de Datos

En primer lugar, se calcularon estadísticas descriptivas para todos los constructos e indicadores incluidos en el estudio. Estos resultados son presentados en la Tabla 1. Para evaluar el potencial sesgo por utilización de una fuente común de datos (*Common Method Bias*, o *CMB* por sus siglas en inglés) se utilizó la prueba de factor único de Harman (Harman 1976, Podsakoff, MacKenzie et al. 2003). Todos los indicadores utilizados en el estudio se incluyeron en un análisis factorial para determinar si la mayoría de la varianza en el modelo era explicada por solo un factor. El porcentaje máximo de varianza explicado fue igual al 26.1%. Dicho resultado sugiere que el sesgo por utilización de una fuente común de datos no representa un problema, ya que el porcentaje de varianza explicado por un factor general fue menor al 50 por ciento (Podsakoff, MacKenzie et al. 2003).

**Tabla 1.** Estadísticas Descriptivas de Constructos e Indicadores de la Encuesta

<i>Constructo/Indicadores</i>		<b>Media</b>	<b>Desvío Estándar</b>
<b><i>Diseño del Instrumento (DI)</i></b>		<b>5.241</b>	<b>0.975</b>
<b>DI1.</b>	Las instrucciones y expectativas del IPA fueron claras.	5.167	1.248
<b>DI2.</b>	Entiendo lo que pretendía evaluar el IPA.	5.778	1.192
<b>DI3.</b>	El IPA estaba altamente conectado con las actividades de clase.	5.315	1.319
<b>DI4.</b>	El contenido de la lectura era apropiado para mi nivel.	4.704	1.437
<b><i>Efectividad de las Evaluaciones (EE)</i></b>		<b>5.071</b>	<b>0.631</b>
<b>EE1.</b>	La evaluación es más efectiva si anima a aplicar conocimiento en contextos reales.	5.071	0.677
<b>EE2.</b>	La evaluación es más efectiva cuando da lugar a mejorar destrezas técnicas como escritura, lectura o interpretación.	4.976	0.749
<b>EE3.</b>	La evaluación es más efectiva cuando contribuye a profundizar lo aprendido.	5.167	0.696
<b><i>Ansiedad frente a las Evaluaciones (AE)</i></b>		<b>4.030</b>	<b>0.716</b>
<b>AE1.</b>	Me pongo nervioso antes de tomar un examen.	4.333	0.786
<b>AE2.</b>	Me es difícil enfocarme cuando me prepare para un examen.	3.571	1.107



Percepciones de los alumnos acerca de las evaluaciones integrales de desempeño

<b>AE3.</b>	Me es difícil enfocarme durante un examen.	4.071	1.068
<b>AE4.</b>	Mi desempeño en los exámenes depende de factores externos (como estado de animo, ansiedad, problemas personales) en vez de mi conocimiento sobre el contenido.	4.143	1.280
<b>Satisfaccion del Alumno (SA)</b>		<b>4.889</b>	<b>0.978</b>
<b>SA1.</b>	Disfruté usar el sistema de IPAs en el aprendizaje y evaluación de la lengua este semestre.	5.148	1.152
<b>SA2.</b>	Las tareas del IPA se relacionan con cómo me veo usando el español en el mundo real.	5.000	1.216
<b>SA3.</b>	El IPA me permitió demostrar claramente lo que aprendí en este curso.	5.296	1.143
<b>SA4.</b>	El IPA me permitió demostrar claramente lo que puedo hacer con la lengua.	4.111	1.737

A continuación, los investigadores siguieron las recomendaciones publicadas en la literatura para evaluar el modelo teórico propuesto. La evaluación del modelo incluyó un análisis de la confiabilidad de los indicadores de la encuesta, así también como la consistencia interna, la validez convergente y la validez discriminante de los constructos utilizados en el estudio. En primer lugar, los factores de cargas externas (*Outer Loadings*, en inglés) fueron analizados para evaluar la confiabilidad de los indicadores de la encuesta. Este análisis es presentado en la tabla número 2.

**Tabla 2.** Factores de Cargas Externas

<b>Indicador</b>	<b>Carga Externa</b>	<b>Valor <i>p</i></b>
<b>DI1</b>	0.757	< 0.001
<b>DI2</b>	0.602	0.001
<b>DI3</b>	0.769	< 0.001
<b>DI4</b>	0.830	< 0.001
<b>EE1</b>	0.798	0.016
<b>EE2</b>	0.916	0.001
<b>EE3</b>	0.942	< 0.001
<b>AE1</b>	0.734	0.008
<b>AE2</b>	0.926	0.004
<b>AE3</b>	0.838	0.009
<b>AE4</b>	0.679	0.029
<b>SA1</b>	0.699	< 0.001
<b>SA2</b>	0.774	< 0.001
<b>SA3</b>	0.863	< 0.001
<b>SA4</b>	0.647	0.005

Todos los factores de cargas externas fueron estadísticamente significativos (*i.e.*, valores  $p < 0.05$ ). Al mismo tiempo, todos los indicadores excepto cuatro ítems de la encuesta tuvieron cargas superiores al límite sugerido de 0.70 (Hair, Hult et al. 2016). A pesar de que cuatro de los ítems tenían cargas entre 0.60 y 0.70, dichos indicadores fueron retenidos ya que su eliminación no era necesaria para aumentar por encima de los límites sugeridos los índices de confiabilidad compuesta de los constructos. Los resultados de esta parte del análisis indicaron un nivel aceptable de confiabilidad para los diferentes indicadores incluidos en la encuesta (Hair, Hult et al. 2016).

La consistencia interna de los constructos se evaluó utilizando coeficientes Alfa de Cronbach, índices de confiabilidad  $\rho_A$  (Dijkstra & Henseler 2015) y coeficientes de confiabilidad compuesta (*Composite Reliability*, en inglés). Como se muestra en la Tabla 3, todos los índices estuvieron por encima del límite de 0,70 recomendado por Hair, Hult et al. (2016), lo que sugiere una consistencia interna aceptable para los cuatro constructos incluidos en el modelo.

**Tabla 3.** Fiabilidad y Validez de Constructos

Constructo	Alfa de Cronbach	$\rho_A$	Confiabilidad Compuesta	Varianza Media Extraída (AVE)
DI	0.733	0.757	0.831	0.554
EE	0.871	1.041	0.917	0.788
AE	0.823	1.062	0.875	0.640
SA	0.738	0.765	0.836	0.563

A continuación, se calcularon estimaciones de varianza media extraída (*Average Variance Extracted*, o AVE por sus siglas en inglés) para establecer la validez convergente de los constructos. Los valores de varianza media extraída (véase la Tabla 3) indican que las cuatro variables latentes incluidas en el estudio explican más de la mitad (*i.e.*, AVE > 0.500) de la varianza de sus indicadores. De esta forma, los resultados sugieren una validez convergente satisfactoria para todos los constructos.

Finalmente, se evaluó la validez discriminante de las variables latentes. Se utilizó el método de Fornell-Larcker (Fornell & Larcker 1981) para determinar si los constructos cumplían las condiciones básicas de validez discriminante. En este sentido, las raíces cuadradas de los valores AVE para cada uno de los cuatro constructos resultaron ser más altas que las correlaciones de dichos constructos con las otras variables latentes incluidas en el modelo (véase Tabla 4), sugiriendo que todos los constructos utilizados en el estudio representan medidas válidas de conceptos únicos (*i.e.*, los constructos son empíricamente distintos).

**Tabla 4.** Validez Discriminante: Criterio de Fornell-Larcker

Constructo	DI	EE	AE	SA
<b>DI</b>	<b>0.744</b>			
<b>EE</b>	0.147	<b>0.888</b>		
<b>AE</b>	0.132	0.089	<b>0.800</b>	
<b>SA</b>	0.629	0.238	0.077	<b>0.750</b>

*Nota:* Raíz cuadrada de los valores AVE en negrita, sobre la diagonal.

Una vez que la confiabilidad y la validez del modelo fueron establecidas, los investigadores analizaron las relaciones hipotéticas incluidas dentro del modelo. La siguiente sección presenta los resultados empíricos del estudio.

## Resultados

### Evaluación de Relaciones Hipotéticas

El modelo de ecuaciones estructurales (*Structural Equation Model*, o *SEM* por sus siglas en inglés) se evaluó utilizando SmartPLS, un paquete de software de modelado de ecuaciones estructurales basado en el método de mínimos cuadrados parciales (o *PLS-SEM* por sus siglas en inglés) (Ringle, Wende et al. 2015). *PLS-SEM* representa un enfoque basado en la varianza que funciona correctamente incluso con muestras pequeñas (es decir, muestras de menos de 100 observaciones) (Hair, Ringle et al. 2011).

Las relaciones hipotéticas fueron evaluadas simultáneamente en SmartPLS analizando el nivel de significación de los coeficientes estandarizados de las distintas relaciones. Los niveles de significación estadística de los coeficientes fueron estimados en SmartPLS utilizando el método de Bootstrapping (5.000 remuestras) (Efron & Tibshirani 1986, Davison & Hinkley 1997). El resumen de los resultados se presenta en la tabla número 5.

**Tabla 5. Resultados del modelo**

Relación	Coefficiente	<i>p</i>	Conclusión
DI → EE	0.147	0.440	No confirmada
DI → AE	0.132	0.532	No confirmada
DI → SA	0.609	<0.001	Confirmada
EE → SA	0.150	0.374	No confirmada
AE → SA	-0.017	0.932	No confirmada

Por un lado, se encontró que el Diseño del Instrumento no tiene un efecto significativo tanto en las percepciones de Efectividad de las Evaluaciones como en las percepciones de Ansiedad frente a las Evaluaciones. De esta forma, el análisis de los

resultados obtenidos no permite confirmar  $DI \rightarrow EE$  y  $DI \rightarrow AE$  a un nivel de significancia de 0.05.

Con respecto a la Satisfacción del Alumno, se determinó que el Diseño del Instrumento tiene un efecto significativamente positivo en la Satisfacción del Alumno a un nivel de significancia de 0.001. Por otro lado, se encontró que tanto las percepciones de Efectividad de las Evaluaciones como las percepciones de Ansiedad frente a las Evaluaciones no tienen un efecto directo significativo en la Satisfacción del Alumno. De este modo, los resultados no brindan apoyo para confirmar  $EE \rightarrow SA$  y  $AE \rightarrow SA$  con un nivel aceptable de significación estadística.

### **Análisis de Mediación**

En función de los resultados obtenidos, los investigadores procedieron a evaluar el rol de las percepciones de Efectividad de las Evaluaciones y las percepciones de Ansiedad frente a las Evaluaciones en la relación entre el Diseño del Instrumento y la Satisfacción del Alumno. La Tabla 6 muestra los resultados de esta parte del análisis.

**Tabla 6. Resultados del Análisis de Mediación**

<b>Relación</b>	<b>Efecto</b>		<b>Conclusión</b>
	<b>Indirecto</b>	<b><i>p</i></b>	
DI $\rightarrow$ EE $\rightarrow$ SA	0.022	0.609	No confirmada
DI $\rightarrow$ AE $\rightarrow$ SA	-0.002	0.962	No confirmada

Los resultados no permiten confirmar los efectos indirectos  $DI \rightarrow EE \rightarrow SA$  y  $DI \rightarrow AE \rightarrow SA$  con un nivel aceptable de significación estadística, sugiriendo que tanto las percepciones de Efectividad de las Evaluaciones y las percepciones de Ansiedad frente a las Evaluaciones no explican una porción significativa del efecto del Diseño del Instrumento sobre la Satisfacción del Alumno. De este modo, los resultados empíricos indican que tanto las percepciones de Efectividad de las Evaluaciones como las percepciones de Ansiedad frente a las Evaluaciones no median la relación entre el Diseño del Instrumento y la Satisfacción del Alumno.

### **Discusión**

El presente estudio examinó la relación entre el Diseño del Instrumento, las percepciones de Efectividad de las Evaluaciones, las percepciones de Ansiedad frente a las Evaluaciones y la Satisfacción del Alumno. Tal como se reporta en la Tabla 5, se determinó que una percepción elevada del diseño del instrumento tiene un impacto positivo en la satisfacción del alumno. Por otro lado, los resultados sugieren que tanto las percepciones de Efectividad de las Evaluaciones como las percepciones de Ansiedad frente a las Evaluaciones no tienen un rol crítico en la relación entre el Diseño del Instrumento y la Satisfacción del Alumno. De estos resultados se desprende que aun

cuando el alumno se presente en el contexto educativo con ansiedad frente a las evaluaciones o con percepciones bajas o negativas en cuanto a la efectividad de las evaluaciones en general, un buen diseño del instrumento de evaluación puede tener un impacto positivo en la satisfacción del alumno.

Según se desprende de estudios anteriores (Levi Altstaedter & Krosi 2018, Martel 2018, Zapata 2016), es de suma importancia estudiar las percepciones de los alumnos en cuanto a la implementación de innovaciones curriculares, como es el caso de las evaluaciones integrales de desempeño. Así, a través del presente estudio se intentó contribuir al entendimiento aún más acabado de cómo el diseño curricular en general y de evaluaciones en particular puede impactar la satisfacción del alumno en cuanto a determinada innovación, aun cuando este pudiera tener ansiedad o percepciones negativas o bajas en cuanto a las evaluaciones en general.

Los resultados del presente estudio contribuyen al campo de la investigación de diseño curricular y de evaluaciones en la enseñanza-aprendizaje de lenguas segundas y extranjeras en cuanto aporta al entendimiento de la relación entre la ansiedad frente a la evaluación, percepciones de efectividad de las evaluaciones, el diseño de instrumentos y la satisfacción del alumno. Así, en lo relativo a la percepción media del diseño de la evaluación utilizada en este estudio como de la satisfacción del alumno en cuanto a esta evaluación, los resultados demuestran que los alumnos tuvieron percepciones positivas. Como se desprende de los resultados, es imprescindible diseñar estratégicamente instrumentos de evaluación, ya que un buen diseño puede contrarrestar variables como la ansiedad frente a las evaluaciones o las percepciones negativas de las evaluaciones en general.

## Conclusión

Según surge del presente estudio, las evaluaciones integrales de desempeño presentan un modelo viable de evaluación en el campo de la enseñanza-aprendizaje de lenguas extranjeras. Este tipo de evaluación permite integrar en un mismo instrumento la evaluación de los tres modos de comunicación a la vez que propone una contextualización que se asemeja al mundo real. Así, por ejemplo, la siguiente cita resalta la visión desde el punto de vista del alumno de que a veces “la forma en que se enseña el español en la escuela no es cómo se usa en el mundo real” [traducción de los autores], mientras que de la cita a continuación surge evidencia de que este tipo de evaluación tiene conexión con el uso de la lengua en contextos auténticos: “la razón de cada tarea es poder escuchar, escribir, y leer en español, lo cual es necesario en el mundo del español” [traducción de los autores].

Como consta de la estructura de estas evaluaciones, la segunda tarea, la de comunicación interpersonal, exige que los alumnos participen en conversaciones de manera espontánea, sin posibilidad de ensayo previo ni uso de materiales de apoyo, como pudiera ser un diálogo previamente creado y luego memorizado. Esto se debe a que el objetivo de esta tarea es evaluar si el alumno está preparado para participar en este tipo de intercambio en una situación real que pudiera enfrentar, partiendo del

supuesto de que no habría en este tipo de situación posibilidad de preparación de lo que se va a decir con antelación. Este enfoque en comunicación interpersonal espontánea puede crear ansiedad en los alumnos, como surge de la siguiente contribución de uno de los alumnos, quien expresó: “no hubo tiempo para practicar lo que íbamos a decir” [traducción de los autores], lo cual por un lado demuestra la postura del alumno, que espera ese tiempo de práctica previo, y por otro sugiere que este tipo de evaluación pone al alumno en una situación en la que se ve obligado a usar la lengua de forma espontánea, tal como sucede en la vida real.

Como indicaron Levi Altstaedter & Krosi (2018), uno de los mayores beneficios de utilizar este tipo de instrumento es que provee tanto para el docente como para el alumno la posibilidad de determinar lo que realmente se ha aprendido, permitiendo que el alumno se dé cuenta de sus habilidades y conocimiento desarrollados hasta el momento de la evaluación. La siguiente contribución de unos de los participantes ilustra este punto: “he aprendido más de lo que pensaba” [traducción de los autores]. Esta afirmación apoya el argumento de que las evaluaciones integrales de desempeño representan una forma de evaluación viable en un contexto de enseñanza de lenguas segundas y extranjeras que se encuadra en un enfoque en el que prima el objetivo de preparar al alumno en este tipo de contexto a desarrollar habilidades comunicativas que le permitan desenvolverse en contextos auténticos, interpretando y presentando mensajes, así como intercambiando información de manera espontánea, todo en contextos auténticos, similares a aquellos a los cuales pudieran enfrentarse los alumnos en la vida real.

## Referencias bibliográficas

- ADAIR-HAUCK, B., E. W. GLISAN, K. KODA, E. B. SWENDER & P. SANDROCK (2006). "The Integrated Performance Assessment (IPA): Connecting assessment to instruction and learning." *Foreign Language Annals* 39(3): 359-382.
- ADAIR-HAUCK, B. & F. J. TROYAN (2013). "A descriptive and co-constructive approach to integrated performance assessment feedback." *Foreign Language Annals* 46(1): 23-44.
- DAVIN, K., F. J. TROYAN, R. DONATO & A. HELLMAN (2011). "Research on the Integrated Performance Assessment in an early foreign language learning program." *Foreign Language Annals* 44(4): 605-625.
- DAVISON, A. C. & D. V. HINKLEY (1997). *Bootstrap methods and their application*. Cambridge, UK, Cambridge University Press.
- DIJKSTRA, T. K. & J. HENSELER (2015). "Consistent partial least squares path modeling." *MIS Quarterly* 39(2): 297-316.
- ECCLES, J. S. & A. WIGFIELD (1995). "In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs." *Personality and social psychology bulletin* 21(3): 215-225.

- ECCLES, J. S. & A. WIGFIELD (2002). "Motivational beliefs, values, and goals." *Annual review of psychology* 53(1): 109-132.
- EFRON, B. & R. TIBSHIRANI (1986). "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy." *Statistical Science* 1(1): 54-75.
- FORNELL, C. & D. F. LARCKER (1981). "Evaluating structural equation models with unobservable variables and measurement error." *Journal of Marketing Research* 18(1): 39-50.
- GLISAN, E. W., D. URIBE & B. ADAIR-HAUCK (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.
- HAIR, J. F., G. T. M. HULT, C. RINGLE & M. SARSTEDT (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Los Angeles, CA, Sage Publications.
- HAIR, J. F., C. M. RINGLE & M. SARSTEDT (2011). "PLS-SEM: Indeed a silver bullet." *Journal of Marketing Theory and Practice* 19(2): 139-152.
- HARMAN, H. H. (1976). *Modern Factor Analysis*. Chicago, IL, University of Chicago Press.
- KAPLAN, C. S. (2016). "Alignment of world language standards and assessments: A multiple case study." *Foreign Language Annals* 49(3): 502-529.
- KETOKIVI, M. A. & R. G. SCHROEDER (2004). "Perceptual measures of performance: fact or fiction?" *Journal of Operations Management* 22(3): 247-264.
- KISSAU, S. & M. J. ADAMS (2016). "Instructional decision making and IPAs: Assessing the modes of communication." *Foreign Language Annals* 49(1): 105-123.
- LEVI ALTSTAEDTER, L. & M. KROSL (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.
- 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. & K. M. BAILEY (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.
- NATIONAL STANDARDS IN FOREIGN LANGUAGE EDUCATION PROJECT (1999). *Standards for foreign language learning in the 21st century*. Yonkers, NY.
- PEREIRA, D., L. NIKLASSON & M. A. FLORES (2017). "Students' perceptions of assessment: a comparative analysis between Portugal and Sweden." *Higher education* 73(1): 153-173.
- PODSAKOFF, P. M., S. B. MACKENZIE, J.-Y. LEE & N. P. PODSAKOFF (2003). "Common method biases in behavioral research: a critical review of the literature and recommended remedies." *Journal of Applied Psychology* 88(5): 879-903.

- PURPURA, J. E. (2016). "Second and Foreign Language Assessment." *The Modern Language Journal* 100(S1): 190-208.
- RINGLE, C. M., S. WENDE & J.-M. BECKER (2015). *SmartPLS 3*. Boenningstedt, Germany, SmartPLS GmbH.
- THE NATIONAL STANDARDS COLLABORATIVE BOARD (2015). *World-Readiness Standards for Learning Languages*. Alexandria, VA.
- WIGGINS, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA, Jossey-Bass Publishers.
- WIGGINS, G. P. & J. MCTIGHE (2005). *Understanding by design*. Alexandria, VA Association for Supervision and Curriculum Development.
- 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.