

Certificación de alemán en contexto universitario mexicano: la correlación entre el *Österreichische Sprachdiplom* y la *Oral Proficiency Interview by Computer*

*German certification in Mexican university context: the correlation between the
Österreichische Sprachdiplom and the Oral Proficiency Computer interview*

RESUMEN: En este artículo el autor hace uso de un enfoque cuantitativo para analizar la correspondencia que existe entre los resultados arrojados por la prueba oral de dos exámenes estandarizados de alemán con reconocimiento internacional; el *ÖSD Zertifikat A2* y la *Oral Proficiency Interview by Computer* (OPIC). Al basarse en los criterios de análisis de tests propuestos por Bachman y Palmer (1996), este trabajo se encarga de proponer los coeficientes necesarios para la implementación de una ecuación que permita crear un modelo de regresión que prediga el desempeño de un candidato en cualquiera de las dos pruebas orales basándose en los resultados que este haya obtenido en alguna de ellas previamente.

PALABRAS CLAVE: correspondencia, correlación, fiabilidad, teoría de tests, evaluación oral, tests estandarizados.

ABSTRACT: In this article, the author uses a quantitative approach to analyze the correspondence between the results of the oral tests of two standardized German certifications with international recognition: the *ÖSD Zertifikat A2* and the Oral Proficiency Interview by Computer (OPIC). Based on the test-analysis criteria proposed by Bachman and Palmer (1996), this work conceives an equation that allows creating a regression model that predicts the performance of any candidate in one of these oral tests based on the results previously obtained in any of them.

KEYWORDS: correspondence, correlation, reliability, test theory, oral assessment, standardized tests.

Carlos C. Solís Becerra
csolis.becerra@outlook.com
Universidad de Leipzig

Recibido: 22/09/2019
Aceptado: 11/11/2019
VERBUM ET LINGUA
NÚM. 14
JULIO / DICIEMBRE 2019
ISSN 2007-7319

Introducción

La certificación de competencias en una lengua extranjera en el ámbito universitario es una necesidad a la que hoy se enfrentan las instituciones de educación superior a nivel global. El intercambio de conocimiento, de académicos y estudiantes, así como la cooperación en proyectos internacionales han revelado la importancia

de la expedición de certificados con reconocimiento en el extranjero que den fe de las capacidades de un individuo en una lengua meta y que se basen en evaluaciones confiables, válidas y objetivas.

En este sentido, han surgido dos herramientas clave que sirven como cimiento en la construcción y diseño de instrumentos de evaluación a nivel regional e incluso global. Por un lado, la publicación del *Marco Común Europeo de Referencia para las Lenguas* (MCERL) por el Consejo de Europa en 2001 ofreció precisamente un nuevo punto de partida para la concepción de certificaciones internacionales de idioma dentro y fuera de Europa (Trim, 2012: 19), como el *Diplôme Élémentaire de Langue Française* (DELF) y el *Diplôme Approfondi de Langue Française* (DALF) de la lengua francesa, el certificado CELI (*Certificazione della Lingua Italiana*) propuesto por el Ministerio de la Instrucción, la Universidad y la Investigación de Italia o el *Österreichische Sprachdiplom* (ÖSD) y el *Goethe Zertifikat* en el caso del alemán. Por otra parte, desde 1982 las *Proficiency Guidelines* del *American Council on the Teaching of Foreign Languages* (APGL) han jugado un rol determinante en el diseño de tests de lengua extranjera y su implementación en todos los niveles educativos de Estados Unidos (Fulcher, 2003: 11); a tal grado que su uso comienza a evidenciarse también fuera de esta región.

Lejos de representar una competencia, la existencia de estos dos marcos de referencia para la evaluación de lenguas ha dado lugar a la coordinación de esfuerzos entre instituciones y profesionales para desarrollar un modelo que logre empatar los constructos y productos de ambos do-

cumentos y que, por lo tanto, permita el reconocimiento mutuo de los resultados arrojados por pruebas basadas en ellos.

La necesidad de este ejercicio de diálogo entre ambos marcos es la razón por la que se llevó a cabo en 2010 la *ACTFL-CEFR Alignment Conference* en el Instituto Herder de la Universidad de Leipzig; evento del que se desprendió la posterior publicación de la obra *Aligning Frameworks of Reference in Language Testing* en 2012 y que ha servido como inspiración para el desarrollo del trabajo que a este estudio atañe, como el caso de la obra de A. Mosher, Tschirner, Swender y Bärenfänger, que serán tratadas a detalle más adelante.

Es por esta razón que la Universidad de Guadalajara (UDG), que ostenta el título de *Prüfungszentrum* (centro evaluador) para el ÖSD desde 1999 y que además está autorizada para implementar evaluaciones de lengua del *American Council on the Teaching of Foreign Languages* (ACTFL), se revela como un escenario ideal para desarrollar un estudio que permita el cálculo del índice de correlación entre las pruebas orales desarrolladas bajo los lineamientos del MCERL y de las APGL.

Este artículo se concentra, por lo tanto, en el cálculo del índice de correlación entre dos modelos específicos de evaluaciones de la producción oral. El primero, basado en los estándares del MCERL, la prueba oral del *ÖSD Zertifikat A2* y el segundo diseñado por *Language Testing International* bajo los parámetros de las APGL, la *Oral Proficiency Interview by Computer* (OPIc). Asimismo, el autor crea un modelo de regresión lineal que permite utilizar una ecuación que prediga el desempeño que un examinado pueda tener en cualquiera de las dos pruebas,

basándose en los resultados que este haya obtenido previamente en alguna de ellas.

Antecedentes

Las certificaciones de idioma, como se mencionó anteriormente, se han convertido en el medio más solicitado que demuestra la competencia lingüística de un individuo frente a alguna autoridad o comunidad. Es por esta razón que se considera que los usuarios finales de estas certificaciones no son los candidatos que presentan una prueba, sino las autoridades gubernamentales o educativas que las requieren (Mègre y Riba, 2014: 13), por lo que, al responder a necesidades institucionales, deben ser tan objetivas y transparentes como sea posible.

Precisamente, una forma de favorecer estas dos condiciones es que las pruebas que llevan a la obtención del diploma en cuestión se basen en un conjunto de parámetros que permitan estandarizar su implementación y la interpretación de sus resultados. Además, aunque actualmente es posible encontrar distintos documentos que sirvan como referencia para la creación de tests de lengua, la realidad es que son solo dos los constructos que predominan en el diseño de evaluaciones: el MCERL y las APGL (*American Council for the Teaching of Foreign Languages*, 2018: 1).

Como se mencionó en el apartado anterior, la búsqueda de correspondencia entre estos dos documentos ya ha sido trabajada en distintos momentos, siendo uno de los casos más concretos las jornadas organizadas en 2010 en el Instituto Herder de la Universidad de Leipzig en colaboración con *Language Testing International* y el *Institut für Testforschung und Testentwicklung*

(ITT) y que desembocaron en la publicación de la antología *Aligning Frameworks of Reference in Language Testing: The ACTFL Proficiency Guidelines and the Common European Framework of Reference for Languages* en 2012. En este tomo, A. Mosher contribuye con el artículo titulado *The determination of equivalent ranges on the ACTFL Proficiency Guidelines-Speaking and the Scale Used for Speaking on the Test of German as a Foreign Language (TestDaF)* (2012: 139-149), donde emplea la fórmula de Spearman en el cálculo de índice de correlación para determinar en qué medida se corresponden los resultados de las dos pruebas analizadas y si es posible establecer una ecuación que prediga el puntaje de uno de los tests con base en el desempeño que un candidato haya obtenido en una primera evaluación, lo que normalmente se conoce como regresión lineal (Fields, 2009: 198):

The results of the present study seem to indicate that the ACTFL OPI is a more accurate predictor of TestDaF levels than the TestDaF ratings are of OPI ratings. Since the ratings according to the two procedures do correlate significantly, some tentative conclusions may be drawn. Because of the distribution of the evaluation results, the equivalences at the higher end (TDN-5 / Advanced Low to Superior) may be seen as more probable than at the lower end (TDN-3 / Intermediate Mid). Table 10 [Figura 1 en este texto] shows correspondences between the TestDaF and ACTFL OPI levels as suggested by this study. (Mosher, 2012: 147)

Figura 1

Posibles correspondencias entre los niveles de las pruebas TestDaF (GER) y ACTFL OPI (Mosher, 2012: 147)

TestDaF	ACTFL OPI
TDN-3	Intermediate Mid
TDN-4	Intermediate High
TDN-5	Advanced Mid

Adicionalmente, en esta misma obra, destaca el trabajo desarrollado por Tschirner, Swender y Bärenfänger (2012) al comparar los resultados de dos exámenes ya no de expresión oral, sino de comprensión escrita: el *NATO Benchmark Advisory Test* (precursor de los tests basados en las APGL) y el *Reading Test for Professionals* creado en la Universidad de Leipzig bajo los lineamientos del MCERL (2012: 136). En este trabajo, los autores determinan que la correspondencia entre ambas pruebas debe ser analizada con un método empírico y cuantitativo, utilizando herramientas estadísticas para este propósito. De esta forma es como el texto concluye con el establecimiento de correspondencia, en este caso específico, entre los niveles intermedio, avanzado y superior de ACTFL con los niveles A, B y C1 del MCERL respectivamente; haciendo hincapié en la necesidad de estudios similares con otras versiones de pruebas basadas en estos dos marcos de referencia para confirmar esta correlación.

Por otro lado, también Tschirner (2005) hace notar el desfase que existe entre las escalas de ACTFL y del MCERL en cuanto a progresión no solo gramatical, sino también de carga horaria necesaria para el desarrollo de la competencia oral en cada nivel de ambos sistemas. Esta contraposi-

ción de ambas escalas fue retomada posteriormente por Martínez (2008: 468) para su uso en un estudio similar, en el que el autor calcula la correspondencia de los resultados de la prueba OPI de ACTFL con aquellos arrojados por una entrevista oral desarrollada por el Centro de Lenguas Extranjeras de la Universidad de Granada siguiendo los lineamientos del MCERL. En su trabajo, Martínez concluye que cada nivel de la escala europea corresponde a dos o hasta tres niveles en el sistema de notación norteamericano.

Figura 2

Comparación de los valores mínimos entre MCERL-ACTFL (Martínez, 2008: 481)

Dado un nivel del MCER...	...se obtendría un mínimo en ACTFL de:
C2 >	> AH
C1 >	> AL
B2 >	> IH
B1 >	> IM
A2 >	> NH
A1 >	> NM

Posteriormente y en una nueva obra, Tschirner (2018: 113-114) obtiene resultados que resultan de especial relevancia para la progresión en la habilidad de producción oral en alemán de 251 estudiantes universitarios de siete universidades norteamericanas. Al implementar la evaluación OPIc a una población de 177 estudiantes, que corresponde a aquellos que en su momento cursaban el cuarto semestre de sus estudios superiores, los resultados obtenidos demuestran que los candidatos oscilaban entre los niveles *Intermediate Low* (IL) e *Intermediate Mid* (IM), lo que corresponde, según una

Figura 3

Comparación de escalas MCERL y ACTFL (Tschirner, 2005: 51)

Europarat	A1	A2	B1	B2	C1	C2
ACTFL	Nov High	Int Mid	Int High	Adv Mid	Adv High	Superior

tabla de conversión propuesta por el propio Tschirner, al nivel A2 alcanzado en la escala del MCERL que los estudiantes de la UDG obtienen tras cursar cuatro semestres de alemán como lengua extranjera.

Para el interés del presente estudio, es de especial relevancia que los trabajos precedentes convergen en la necesidad latente de datos sustentados en procedimientos empíricos que se ocupen del análisis de los resultados arrojados por los instrumentos de evaluación desarrollados bajo los criterios del MCERL y de ACTFL. Únicamente de esta forma podrá establecerse con certeza la correlación que existe entre ambas escalas para la valoración de las competencias productivas y receptivas. Por lo que, en el afán de contribuir a un entendimiento entre los dos sistemas aquí expuestos, este estudio se ocupa del cálculo del índice de correlación entre la evaluación oral del *ÖSD Zertifikat A2* y la prueba OPIc. Ambas pruebas forman parte de la oferta para certificación de alemán en la UDG, que se desempeña como *Prüfungszentrum del Österreichische Sprachdiplom* y que además ha obtenido la licencia de LTI-ACTFL para implementar sus evaluaciones de lengua en la totalidad de su red universitaria.

Evaluación y certificación de lenguas extranjeras en la Universidad de Guadalajara

En concordancia con las necesidades actuales para la cooperación universitaria

internacional, la UDG publicó en 2016 su propia Política de Lenguas Extranjeras (PLE), inserta en el Plan de Desarrollo Institucional 2014-2030 (2016b: 4) y que establece explícitamente la lengua inglesa como prioridad en la adquisición de lenguas para la currícula de sus planes de estudio, mencionando no obstante “otras lenguas” que pueden ser de utilidad para distintos fines académicos (2016b: 4).

Por esta razón, la universidad pretende homogeneizar los planes de trabajo y estudio en lo que a adquisición de una segunda lengua se refiere, de tal forma que sus estudiantes consigan desempeñarse de forma autónoma como individuos bilingües en el ámbito profesional y académico (2016b: 5). De gran utilidad para este fin resulta entonces el uso de un marco de referencia que determine la pauta a seguir en este proceso de estandarización, como es el caso del MCERL; que ha sido retomado por la Secretaría de Educación Pública (SEP) para alinear los niveles meta en cada nivel educativo con la carga horaria requerida (Figura 4).

La publicación de esta PLE en la red universitaria implicó que los distintos centros de idioma de la UDG se ajustaran a los requerimientos expresados en este documento, siguiendo la escala del MCERL para la enseñanza y evaluación de lenguas dentro de la institución. Sin embargo, la firma de un convenio también en 2016 con ACTFL para la implementación de tests

Figura 4

Objetivos lingüísticos por nivel educativo en la UDG (2016b: 6)

MCER	Nivel educativo	Horas de estudio
A2 Usuario elemental	Medio superior	201-350
B1 Usuario intermedio	Licenciatura	351-500
B2 Usuario intermedio superior	Posgrado	501-700

de distintos idiomas (2016a), implicó que la Universidad pudiera valorar las competencias comunicativas de los estudiantes bajo los parámetros de ya no solo el documento europeo, sino también de aquellos del marco de referencia norteamericano, poco conocido tanto por docentes, como por estudiantes de la UDG.

El estudio del que se desprende este artículo responde pues a la necesidad de comprender la correspondencia que puede existir entre dos escalas que entienden el dominio de una lengua extranjera y la progresión en su aprendizaje de forma aparentemente distinta; de tal forma que la oferta de certificaciones y cursos al interior de la universidad sea lo más transparente posible.

Metodología

Los tests son instrumentos de medición que permiten conocer, en valores numéricos, el progreso que un individuo ha realizado en la adquisición de un idioma o que demuestran el dominio que una persona tiene de este. En consecuencia, el diseño de estas herramientas debe estar guiado por parámetros que garanticen su objetividad y que disminuyan la probabilidad de mediciones erróneas. Para este efecto, Bachman y Palmer (1996) proponen cuatro criterios de calidad que deben de ser observados durante la construcción de un test: la validez, la fiabilidad del instrumento, su objetividad y su utilidad.

Esta naturaleza cuantitativa de los tests favorece el uso de un enfoque estadístico para analizar los datos recolectados. El empleo de un procedimiento empírico demuestra de forma objetiva y precisa si las expectativas de uno o varios instrumentos han sido demasiado altas o más bien bajas. Bachman, por ejemplo, se vale de la ecuación propuesta por Spearman para desarrollar un estudio de correlación entre dos instrumentos de evaluación (Bachman, 2004: 87-88). El cálculo del índice de correlación puede hacerse también, sin embargo, con la fórmula de Pearson. No obstante, para los objetivos de este trabajo, se eligió la ecuación de Spearman por ser la más empleada para el cálculo de un rango de coeficiente de correlación, además de que ofrece datos más precisos y se adapta de mejor forma al trabajo con más de una variable en una escala ordinal (Kuckatz et al., 2013: 217):

Figura 5: Fórmula de Spearman para el cálculo del índice de correlación

$$r_s = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)}$$

r_s: coeficiente de correlación por rangos de Spearman

D: Diferencia entre los rangos entre *X* y *Y*

n: población total

Esta fórmula, como ya se ha dicho, ha sido usada de forma exitosa en distintos estudios que pertenecen a distintas disciplinas por medio de herramientas como *software* de estadística. Precisamente este ha sido el caso para el tratamiento de la información recolectada durante este estudio, donde se ha empleado el programa SPSS (*Statistical Package for the Social Sciences*) que ha presentado buenos resultados en el campo de las ciencias sociales.

Procedimiento del estudio

La UDG ofrece desde 1980 cursos de alemán como lengua extranjera en distintos formatos. Luego de que en 2006 se comenzara a adoptar el MCERL como documento estandarizador de la planeación y evaluación de cursos de idioma, se ofreció por primera vez la modalidad de curso intensivo. Asimismo, a partir de 2008 existe también la posibilidad de tomar cursos sabbáticos de alemán para el público en general. En ambos casos, los asistentes tienen un total de 85 horas de clase por semestre, con una carga horaria de cinco a seis horas por semana.

Los datos recolectados en este estudio provienen de voluntarios de estas dos modalidades de cursos que ya han finalizado el nivel A2, siendo la mayoría de los casos estudiantes de un curso intensivo de B1. En total, la población analizada consta de 14 estudiantes entre hombres y mujeres que oscilan entre los 22 y los 30 años de edad. Cabe destacar que en este procedimiento no solo se recolectó información de estudiantes del Centro Universitario de Ciencias Sociales y Humanidades (CUCSH), sino que hubo participantes provenientes de otras facultades de la red universitaria de la UDG.

Además, cabe mencionar que es hasta el curso A2.1 que los estudiantes ven reflejados créditos en la carga curricular de sus licenciaturas. Los niveles posteriores, al no ser cursos oficiales, son solamente tomados por los estudiantes que desean continuar con el aprendizaje del alemán para obtener algún beneficio académico, como puede ser un semestre de movilidad en alguna universidad germanohablante, las posibilidades de continuar sus estudios de posgrado en el extranjero o el acceso a literatura en lengua original. Es por esta razón que la cantidad de individuos que se ha podido analizar durante esta investigación se muestra reducida, pero a la vez representativa, tomando en cuenta la relación que existe entre población total, que asciende a aproximadamente 30 estudiantes por semestre en el nivel B1.1, y población estadística (n).

El proceso de colecta de datos se desarrolló en dos sesiones por individuo. Durante el primer encuentro con el evaluador, con una duración de entre diez y quince minutos por persona, se llevó a cabo la prueba oral del *ÖSD Zertifikat A2*, simulando el protocolo que tanto el *ÖSD* como *ALTE* (*Association of Language Testers in Europe*) estipulan y en presencia de un evaluador certificado por este órgano, quien fue el encargado de asignar el puntaje total de cada estudiante siguiendo el formato estandarizado de la prueba oral *ÖSD*.

En una segunda sesión con cada voluntario, el investigador les hizo saber el procedimiento de la prueba *OPIc*; que se desarrolló en línea con ayuda de los equipos de cómputo del Centro de Autoacceso de Idiomas del Departamento de Lenguas Modernas en el CUCSH. Cada una de las

entrevistas se realizó de forma individual, con un nombre de usuario y contraseña proporcionados a cada estudiante por LTI (*Language Testing International*), institución encargada de desarrollar la prueba. La duración de cada prueba fue de aproximadamente veinticinco minutos por persona y una vez que los candidatos finalizaron el proceso, las entrevistas se enviaron de forma automática al centro de evaluación de LTI, que daría acceso a los resultados luego de un par de días en su sitio web por medio de una cuenta protegida asignada al investigador.

Finalmente, es necesario recalcar que en este último paso, LTI proveyó al investigador con las notas de los candidatos empleando la escala europea y no aquella de las APGL. Lo anterior, debido a que el convenio firmado entre la UDG con AC-TFL estipula que, basados en los valores de equivalencia mencionados en la figura 3, LTI podrá ofrecer los resultados de las pruebas realizadas bajo los parámetros de las APGL en cualquiera de las dos escalas.

Resultados

Análisis estadístico general

El análisis estadístico de los resultados obtenidos por los catorce candidatos en ambas pruebas se desarrolló, como se ha mencionado antes, con ayuda del software

SPSS. La figura 6 ilustra un primer análisis general de los datos recolectados. Importante es hacer notar que los resultados del test OPIc fueron proveídos por LTI basándose en la escala del MCERL; por lo que se tradujeron estos datos en forma numérica, correspondiendo el nivel A1 al valor 1, el nivel A2 al valor 2 y nivel B1 al valor 3. Por otra parte, la prueba oral del *ÖSD Zertifikat A2* ofrece una calificación máxima de 20 puntos, siendo 10 el valor mínimo requerido para acreditar esta competencia durante la certificación de este nivel.

En esta primera fase del análisis, se advierte que la prueba *ÖSD A2* fue acreditada por todos los candidatos, con valores iguales o mayores a los 10 puntos mínimos requeridos en esta certificación. No obstante, llama la atención que en el caso de la prueba OPIc los estudiantes evaluados oscilaron entre el nivel A1 y B1, lo que sugiere que puede haber un desfase entre los procesos de evaluación de ambos tests y que algunos estudiantes superaron los requerimientos del nivel A2 al obtener la nota máxima en la prueba *ÖSD* y acreditar el nivel 3 (B1) en la entrevista OPIc. Los valores de las medias para ambas pruebas sugieren un desempeño bueno en expresión oral para el nivel A2, siendo 2.07/3.0 en el caso de la prueba OPIc y 16.21/20.0 para la evaluación *ÖSD A2*.

Figura 6

Descriptivos estadísticos generales del estudio

	N	Mínimo	Máximo	Media	Desv. Desviación
OPIc	14	1	3	2.07	.829
ÖSD	14	10	20	16.21	2.860
N válido (por lista)	14				

La distribución de los resultados por candidato se puede ver de forma detallada en la figura 7, donde cuatro de los candidatos que acreditan la competencia oral en la prueba ÖSD A2 son clasificados con nivel 1 (A1) luego de presentar una evaluación OPIc. Por otra parte, es notorio que, además de los cinco estudiantes catalogados en el nivel 2 (A2) según la OPIc, los cinco restantes superaron las expectativas y lograron acreditar un nivel superior al previsto.

Las curvas de frecuencia en el gráfico de la figura 8 confirman en primer lugar la tendencia ascendente de los valores de ambas variables, que tienden a presentar los puntajes más altos en el lado izquierdo de cada agrupación de datos. Un segundo aspecto por destacar en un análisis más detallado de este gráfico es la presencia de

valores anómalos en este estudio. Como podrá verse al comparar las primeras dos columnas con el espacio designado al nivel 3 de OPIc, destaca la presencia de un caso clasificado en este nivel (B1) por la prueba basada en lineamientos de ACTFL y que corresponde, sin embargo, a un estudiante que demostró un desempeño promedio durante la prueba oral del ÖSD A2, obteniendo una nota de 15/20 y que parece corresponder más bien a valores típicos del nivel 2 (A2) de OPIc. La presencia de esta anomalía se evidencia de forma más concisa en el diagrama de cajas de la figura 9, donde una marca correspondiente al candidato con el número trece aparece fuera del esquema de resultados del resto de los voluntarios que realizaron ambas pruebas.

Figura 7
Resumen de procesamiento de casos

	OPIc	Válido		Casos perdidos		Total	
		N	Porcentaje	N	Porcentaje	N	Porcentaje
ÖSD	1	4	100.0	0	0.0	4	100.0
	2	5	100.0	0	0.0	5	100.0
	3	5	100.0	0	0.0	5	100.0

Figura 8
Histograma de frecuencia de los casos

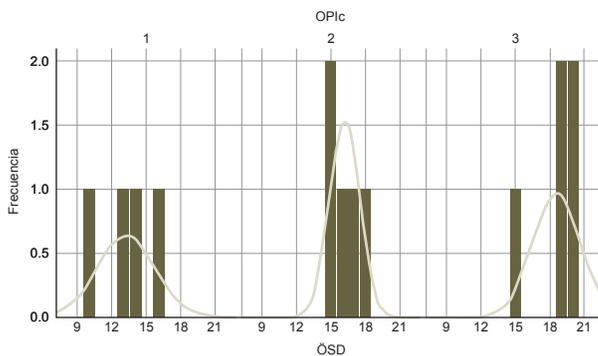
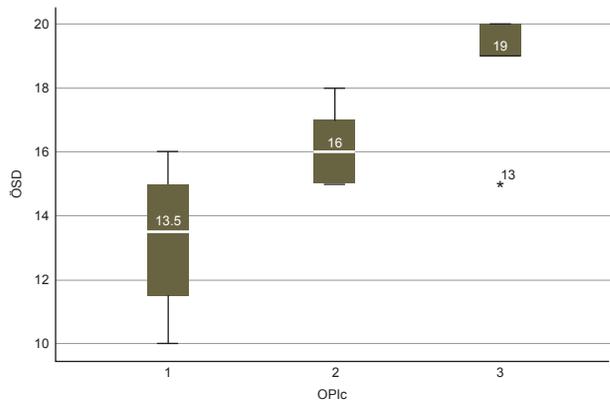


Figura 9
Diagrama de caja
OPIc – ÖSD A2



Adicionalmente, este esquema de cajas confirma la tendencia ascendente que ya se demostraba en la figura 8; además de que ilustra cómo el rango de resultados en la prueba ÖSD se acota en los niveles correspondientes a A2 y B1 de OPIc, llegando incluso a evidenciarse el desfase del desempeño de estudiantes clasificados en el nivel A1 de OPIc respecto a aquellos que obtuvieron el nivel B1 en la misma prueba. El caso concreto del valor anómalo identificado en esta misma categoría será discutido posteriormente.

Análisis de correlación

En el apartado anterior ya se ha establecido la naturaleza no paramétrica de este análisis, que ha develado la presencia de una tendencia ascendente, pero con valores anómalos. Para determinar la correlación existente entre ambas variables se hace entonces necesario el cálculo del índice de correlación con la fórmula de Spearman, que la tabla en la figura 10 ilustra y donde destaca un coeficiente de correlación (ρ) de 0.759 entre las dos pruebas orales analizadas. Esto implica una correlación de casi

Figura 10
Correlación entre variables (ρ de Spearman)

		ÖSD		OPIc
Rho de Spearman	ÖSD	Coeficiente de correlación	1.000	.759**
		Sig. (bilateral)		.002
		N	14	14
	OPIc	Coeficiente de correlación	.759**	1.000
		Sig. (bilateral)	.002	
		N	14	14

** La correlación es significativa en el nivel 0.01 (bilateral).

el 76% entre ambas variables y sugiere que ambas pruebas evalúan o “miden” constructos similares. Es por esta razón que se puede hablar de una covarianza positiva y, en consecuencia, de una correlación positiva de los datos, que el gráfico de la figura 11 ejemplifica.

La recta que cruza este gráfico revela la tendencia ascendente y positiva que ya se explicaba anteriormente, lo que significa que aunado al índice de correlación antes mencionado, es de esperar que a mejores resultados en cualquiera de las dos pruebas analizadas, el desempeño del mismo estudiante sea igualmente bueno en la evaluación restante. Cabe mencionar, además, que el bajo valor de $P(\alpha)$ en la figura 10 demuestra que la población analizada (N) puede ser considerada como confiable, lo que implica que en condiciones parecidas, un estudio de las mismas características debería arrojar resultados también similares.

En su propio estudio al calcular la correlación entre la OPI y el TestDaf, Mosher obtiene resultados similares a los presentados en este trabajo: “A statistical analysis shows that the OPI and the TestDaf ratings correlate significantly (Pearson $r = .616$, $p < .001$; Spearman’s $R = .672$,

$p < .001$)” (2012: 142). Con un índice de correlación de Spearman de 0.759, la correlación del presente estudio es incluso mayor a la obtenida en el cálculo de Mosher, aunque este ciertamente analiza una población mayor; lo que podría explicar un mayor índice de significatividad en el caso de este autor ($\alpha = .001$ vs $\alpha = .002$).

En resumen, ambas investigaciones demuestran la presencia de una correlación significativa entre las variables. Es por ello que, tras el cálculo del índice de correlación en el trabajo de Mosher y en el presente estudio, es posible establecer una similitud entre los constructos de las pruebas orales basadas en los lineamientos del MCERL o de las APGL.

Análisis de regresión

Tomando en cuenta los resultados obtenidos en el análisis de correlación de esta investigación y aquellos del trabajo realizado por Mosher, es posible desarrollar un análisis de regresión que permita utilizar los resultados arrojados por la prueba OPIc como predictores del desempeño que los examinados en una prueba basada en el MCERL como el *ÖSD Zertifikat A2*. Para este propósito, se calculó el valor de R^2 por medio del software

Figura 11
Gráfico de dispersión

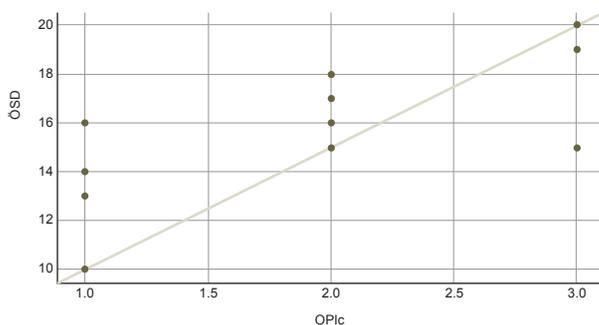


Figura 12
Cálculo de R²

Resumen del modelo ^b				
Modelo	R	R ²	R ² ajustado	Error estándar de la estimación
1	.772 ^a	.596	.562	1.893

^a Predictores: (Constante), OPIc.

^b Variable dependiente: ÖSD.

SPSS, como puede constatarse en la figura 12. En este procedimiento, se concluye que la probabilidad de predicción de resultados basándose en el nivel obtenido en una prueba OPIc (R²) es de casi 57%, lo que implica a su vez que más del 40% de los casos en que se realicen predicciones podrían presentar alguna incompatibilidad, caso que será discutido más adelante.

No obstante, las posibilidades de realizar un pronóstico exitoso de los resultados en la certificación ÖSD A2 son de más del 50%, como ya se ha dicho anteriormente. Las figuras 13 y 14 ilustran de forma más concreta los datos obtenidos para la construcción de este modelo y su funcionamiento.

Según los coeficientes presentados en la tabla anterior, los candidatos que presentan la prueba oral del *ÖSD Zertifikat A2* habiendo acreditado al menos el nivel A1 en la

OPIc, obtendrán por lo general una calificación de entre 10 y 11 puntos (10.7). Esto quiere decir que, para los individuos evaluados durante este estudio estaba garantizado un puntaje mínimo de 10 puntos en la prueba austriaca al haber acreditado al menos el nivel 1 del OPIc. Asimismo, el análisis realizado en SPSS arrojó una nota mínima de 2.66 puntos para las pruebas OPIc, lo que implica un punto medio entre los niveles A2 y B1. Los valores límite de 7.63 y 13.76 en la tabla representan la frontera del rango de puntajes mínimos que un estudiante podría obtener en la prueba ÖSD.

Es así como en el diagrama de dispersión de la figura 14 podemos encontrar la fórmula que ayuda al cálculo de un posible resultado en la prueba europea:

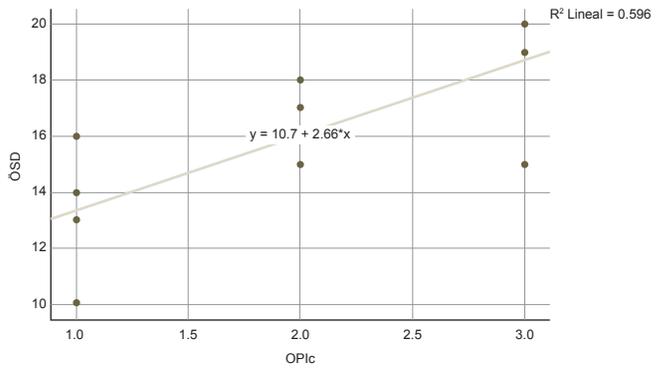
$$Y = 10.69 + (2.66 * x)$$

Figura 13
Coeficientes del modelo de regresión

Modelo	Coeficientes ^a							
	Coeficientes no estandarizados		Coeficientes estandarizados		t	Sig.	95.0% intervalo de confianza para B	
	B	Des. Error	Beta				Límite inferior	Límite superior
1	(Constante)	10.696	1.406		7.606	.000	7.632	13.760
	OPIc	2.664	.633	.772	4.206	.001	1.284	4.044

^a Variable dependiente: ÖSD.

Figura 14
Diagrama de dispersión
según modelo de regresión



Es de esta forma como el posible puntaje de una prueba *ÖSD Zertifikat A2* (Y) se puede calcular con la suma del coeficiente de regresión 10.69 y la multiplicación del valor mínimo de OPIc, 2.66, con el nivel OPIc acreditado por algún candidato (X). De esta forma, la predicción para un estudiante que haya alcanzado el nivel A2 de OPIc se llevaría a cabo de la siguiente manera:

$$Y = 10.69 + (2.66 * 2)$$

$$Y = 16.01$$

Es importante señalar, sin embargo, que la aparición de valores anómalos podría representar una perturbación en el cálculo y la predicción de resultados en este modelo de regresión, lo que finalmente afecta la fiabilidad del modelo, razón por la cual la regresión lineal se ha calculado nuevamente sin tomar en cuenta los datos atípicos, como veremos en el apartado siguiente.

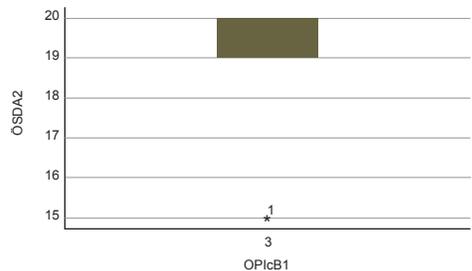
Casos de anomalías

Durante la primera fase del análisis estadístico, las figuras 8 y 9 ayudaron a identificar la presencia de valores anómalos que pueden afectar la validez del estudio. Con la fi-

abilidad de recrear un modelo de análisis de regresión y una correlación entre variables fiables, el autor reconstruyó el modelo esta vez sin considerar la información que pudiera viciar el análisis y que queda en evidencia al analizar el diagrama de caja en del apartado 5.1 a detalle, como se muestra en la figura 15.

En este nuevo diagrama se constata que la presencia de un valor equivalente a 15 puntos en la certificación oral *ÖSD* dista de la tendencia normal que hay entre los estudiantes, quienes luego de presentar esta misma prueba fueron catalogados en el nivel B1 según los resultados de la entrevista

Figura 15
Dato anómalo



OPIc. La relación que este puntaje irregular presenta con los resultados de otros estudiantes con el mismo nivel en la prueba ÖSD y clasificados en niveles inferiores según la OPIc será discutida posteriormente.

Como puede apreciarse en la figura 16, los datos estadísticos descriptivos de este nuevo modelo sufren modificaciones poco relevantes, con un ajuste de las medias equivalente a un descenso de 0.07 para el caso de la prueba OPIc y de un aumento de 0.10 para la ÖSD.

En un nuevo diagrama de caja, se puede constatar que el único cambio resultante de la eliminación del dato anómalo es el aumento de la media para el caso del grupo OPIc 3, que asciende de 19 a 19.5, lo que confirmaría que los estudiantes con

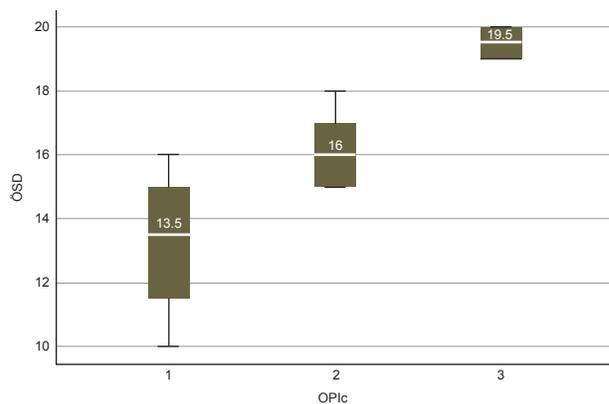
puntajes altos en la prueba oral del ÖSD *Zertifikat A2* normalmente acceden a un nivel superior en la examinación OPIc.

Por otra parte, la eliminación de la irregularidad identificada representó un aumento significativo en el cálculo del índice de correlación de Spearman, como puede verse en la figura 18; lo que implicaría una correlación aún más fuerte entre ambas pruebas orales. A sabiendas de lo anterior, es además de esperar que, junto con el índice de correlación, el valor de R^2 también se vea afectado en forma ascendente. Esta hipótesis se ve confirmada en la tabla de la figura 19, donde el valor de la predicción sube aproximadamente 15 puntos en relación con el cálculo presentado en el apartado 5.3, donde el valor de R^2 equivalía a

Figura 16
Estadísticos descriptivos sin valores irregulares

	N	Mínimo	Máximo	Media	Desv. Desviación
OPIc	13	1	3	2,00	.816
ÖSD	13	10	20	16,31	2,955
N válido (por lista)	13				

Figura 17
Diagrama de caja sin valores irregulares



0.596; frente a 0.746 en el presente modelo. Este hecho se traduce en una posibilidad de predicción de casi el 75% utilizando los resultados de la prueba OPIc como predictores para aquella de ÖSD A2.

Para el análisis de regresión, los coeficientes que resultan de la eliminación de anomalías se ven reflejados en la tabla (figura 20), de la que resulta la ecuación:

$$Y = 10.06 + (3.13 * X)$$

De tal forma que, tomando el mismo ejemplo que se dio para el modelo de regresión anterior, en el caso de que un estudiante acreditara el nivel A2 (2) luego de una examinación OPIc, la predicción de sus resultados para una prueba oral ÖSD A2 sería la siguiente:

$$Y = 10.06 + (3.13 * 2)$$

$$Y = 16.32$$

Figura 18

Cálculo de correlación de Spearman (rho) sin valores irregulares

			ÖSD	OPIc
Rho de Spearman	OPIc	Coefficiente de correlación	1.000	.883**
		Sig. (bilateral)		.000
		N	13	13
	ÖSD	Coefficiente de correlación	.883**	1.000
		Sig. (bilateral)	.000	
		N	13	13

Figura 19

Cálculo de R² sin valores irregulares

Resumen del modelo ^b				
Modelo	R	R ²	R ² ajustado	Error estándar de la estimación
1	.864 ^a	.596	.723	1.556

^a Predictores: (Constante), OPIc.

^b Variable dependiente: ÖSD.

Figura 20

Coefficientes para análisis de regresión sin valores irregulares

Coefficientes ^a						
Modelo		Coefficientes no estandarizados		Coefficientes estandarizados		
		B	Desv. Error	Beta	t	Sig.
1	(Constante)	10.058	1.182		8.508	.000
	OPIc	3.125	.550	.864	5.679	.000

^a Variable dependiente: ÖSD.

Es así como se puede concluir que, aunque la variación de resultados que arroja el modelo de regresión en una prueba oral ÖSD A2 basándose en el desempeño de una prueba OPIc, se ve poco afectada ante la eliminación de valores irregulares, la probabilidad de que esta predicción sea correcta sí aumenta significativamente, confiriéndole a este modelo de regresión mayor fiabilidad.

Discusión y conclusiones

Como ya se ha hecho mención, la existencia de dos marcos de referencia principales para la evaluación de lenguas ha provocado el interés de distintos autores por identificar el nivel de similitud que ambos pueden llegar a presentar. Un ejemplo de lo anterior son trabajos como el de Clifford (2012: 50-55), que dejan al descubierto la naturaleza pragmática del marco de referencia norteamericano, que encuentra su misión y origen al cubrir las necesidades laborales del mercado de trabajo en EE.UU, promoviendo la adquisición de habilidades propias del “mundo real y funcional”; algo que según Clifford el MCERL relega a una función secundaria al favorecer una comunicación oral y comunicativa para situaciones cotidianas y no necesariamente relacionadas con el ámbito laboral.

Sin embargo, el análisis estadístico desarrollado durante este estudio permitió al autor establecer una correlación moderada y positiva ($\rho = 0.759$) entre los resultados que las pruebas orales analizadas arrojaron. Adicionalmente, un nuevo análisis de varianza permitió a su vez identificar el aumento de este índice de correlación ($\rho = 0.883$) cuando el estudio fue desarrollado nuevamente sin contemplar la información irregular que destacó en los gráficos corres-

pondientes a la figura 9 y 14 de este artículo. Es importante señalar que la presencia de un índice de correlación superior a los 0.800 puntos, se puede traducir en una fuerte relación de validez entre la prueba oral de la certificación ÖSD A2 y la OPIc, lo que sugiere que ambas pruebas, contrario a lo que podría llegar a pensarse en algunos casos, evalúan constructos similares en el caso de la producción oral de adultos con nivel A2 de alemán.

Esta alta correlación entre ambos tests fue además el principal indicador que permitió desarrollar un modelo fiable de regresión para la predicción de resultados de la prueba oral ÖSD A2 con base en el desempeño de un candidato en la OPIc. Esto queda demostrado en el presente trabajo, que arrojó un valor de hasta 0.746 para R^2 , lo que se traduce en una certeza cercana al 75% para el modelo desarrollado y que podría quedar confirmada en un estudio que abarque una población más amplia.

La relevancia de este modelo de regresión queda clara en el caso de la UDG y de cualquier otra institución en México que evalúe o mida el desempeño de sus estudiantes bajo los lineamientos del MCERL. El uso cada vez más popularizado de las APGL no solo para la planeación de un curso de lengua, sino para la evaluación de la misma, queda evidenciado con la firma del convenio que en 2016 firmaran la UDG y ACTFL y que representa el primer caso de un centro evaluador que emplee las APGL en Latinoamérica. El modelo de regresión aquí propuesto se revela como un primer paso hacia una herramienta para la comprensión del alcance y funcionamiento del marco norteamericano tanto para las autoridades universitarias, como para los candidatos.

Como se mencionó en el apartado 2, un estudio similar al presente fue realizado en el 2003 en la Universidad de Leipzig (Mosher, 2012); revelando en sus resultados una correlación moderada 0.672, lo que significa 0.211 puntos por debajo del índice encontrado en este trabajo al eliminar valores anómalos ($\rho = 0.883$). No obstante, el alcance del estudio de Mosher es mayor al de la presente investigación, con una población de 35 candidatos evaluados con la prueba oral TestDaF y la *Oral Proficiency Interview* (OPI), que representa la variante no digitalizada de la prueba oral tematizada en este artículo. Tomando en cuenta lo anterior, queda clara la necesidad de realizar un nuevo estudio con una población más numerosa que pueda confirmar los datos arrojados en este trabajo y que podría incluso concentrarse en la correlación de la producción no solamente oral, sino tam-

bién escrita en las certificaciones de alemán *ÖSD Zertifikat A2* y *ACTFL-LTI*.

No se debe pasar por alto, que tanto el estudio de 2003 como el presentado aquí, coinciden en la presencia de una correlación positiva y de moderada a fuerte entre una prueba oral basada en el constructo del MCERL y otra bajo los lineamientos de las APGL. Por esta razón ambas investigaciones contemplan la posibilidad de desarrollar un estudio de regresión integral que sirva como predictor de resultados y que, por lo tanto, permita que los usuarios de estos dos tests comprendan de mejor forma cómo fueron evaluados y cómo pueden interpretar su desempeño en ambas escalas. Lo anterior solamente será posible con nuevas investigaciones que establezcan la correlación que existe entre el resto de las competencias comunicativas evaluadas en las pruebas *ÖSD Zertifikat* y sus equivalentes en la versión *LTI*.

Bibliografía

- American Council on the Teaching of Foreign Languages (2012). *ACTFL proficiency guidelines*, Alexandria, VA: ACTFL.
- American Council for the Teaching of Foreign Languages (20/03/2018). *Assigning CEFR Ratings to ACTFL Assessments*. Consultado el 26/04/2018 en <https://www.actfl.org/publications/additional-resources/assigning-cefr-ratings-actfl-assessments>
- American Council on the Teaching of Foreign Languages (2015). *NCSSFL-ACTFL Can-Do Statements Performance Indicators for Language Learners*. Consultado en <https://www.actfl.org/>

- publications/guidelines-and-manuals/ncssfl-actfl-can-do-statements
- Bachman, L. F. (2004). *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.
- Bachman, L. F. y Palmer, A. S. (1996). *Language Testing in Practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Barkowski, H. y Krumm, H. (Ed.). (2010). *Fachlexikon Deutsch als Fremd- und Zweitsprache*. Tübingen: Narr Francke Attempto.
- Bolton, S. (1991). *Évaluation de la compétence communicative en langue étrangère*. Paris: Hatier et Didier.
- Bordón, T. y Liskin-Gasparro, J. E. (2006). *Evaluación*. En Lacorte, M. (Coord.).

- Lingüística aplicada del español*. Madrid: Arco/Libros. Pp. 211-247
- Brown, H. D. y Abeywickrama, P. (2010). *Language Assessment. Principles and Classroom Practices*. New York: Pearson Longman.
- Clifford, R. T. (2012). It is easier to malign tests than it is to align tests. En Tschirner, E. (Ed.). *Aligning Frameworks of Reference in Language Testing: The ACTFL Proficiency Guidelines and the Common European Framework of Reference for Languages*. Tübingen: Stauffenburg Verlag. Pp. 49-56
- Douglas, D. (2010). *Understanding Language Testing*. London: Hodder Education.
- Europarat für kulturelle Zusammenarbeit (Ed.) (2001): *Gemeinsamer europäischer Europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin, Múnich: Langenscheidt.
- Fulcher, G. (2003). *Testing Second Language Speaking*. Londres: Routledge.
- Glaboniat, M. (1998). *Kommunikatives Testen im Bereich Deutsch als Fremdsprache: Eine Untersuchung am Beispiel des Österreichischen Sprachdiploms Deutsch*. Innsbruck-Viena: Studien Verlag.
- Glaboniat, M. et al. (2012). *A2 Grundstufe Deutsch 2, inklusive Z-Variante. Handreichung für Prüferinnen und Prüfer*. Viena y Klagenfurt: Österreichisches Sprachdiplom Deutsch.
- Glisan, E. W.; Swender, E. y Surface, E. A. (2013). Foreign Language Teacher Candidates: Current Findings and Future Research Directions. En *American Council for the Teaching of Foreign Languages: Foreign Language Annals*, 46, 2. Pp. 264-289
- Grotjahn, R. (2000). Testtheorie: Grundzüge und Anwendungen in der Praxis. En *Materialien Deutsch als Fremdsprache*, 53. Regensburg: FaDaF. Pp. 304-341
- Kromrey, H. (2001). Evaluation - ein vielschichtiges Konzept: Begriff und Methodik von Evaluierung und Evaluationsforschung; Empfehlungen für die Praxis. En *Sozialwissenschaften und Berufspraxis*, 24; 2. Pp: 105-131
- Kuckartz, U.; Rädiker, S.; Erbert, T. y Schehl, J. (2013) *Statistik. Eine verständliche Einführung* Wiesbaden: Springer.
- Martínez Baztán, A. (2008). *La evaluación oral: una equivalencia entre las guidelines de ACTFL y algunas escalas del MCER*. Universidad de Granada.
- Mègre, B. y Riba, Patrick. (2014). *Démarche qualité et évaluation en langues*. Paris: Hachette.
- Mosher, A. (2012). The determination of equivalent ranges on the ACTFL Proficiency Guidelines-Speaking and the Scale Used for Speaking on the Test of German as a Foreign Language (TestDaF). En Tschirner, E. (Ed.). *Aligning Frameworks of Reference in Language Testing: The ACTFL Proficiency Guidelines and the Common European Framework of Reference for Languages*. Tübingen: Stauffenburg Verlag. Pp. 139-149
- Neuland, E. y Peschel, C. (2013). *Einführung in die Sprachdidaktik*. Stuttgart: J. B. Metzler
- Puren, C. (1988). *Histoire des méthodologies d'enseignement des langues vivantes*. Consultado el 24/06/2018 en <https://www.aplv-languesmodernes.org/docrestreint.api/1849/>

- b1a776bacb5d6ccb0a692b19bd
88566e4b5a707/pdf/puren_histoire_
methodologies.pdf
- Swender, E.; Tschirner, E. y Bärenfänger, O. (2012). Comparing ACTFL/ILR and CEFR Based Reading Tests. En Tschirner, E. (Ed.). *Aligning Frameworks of Reference in Language Testing: The ACTFL Proficiency Guidelines and the Common European Framework of Reference for Languages*. Tübingen: Stauffenburg Verlag. Pp. 123-137
- Thompson, G. L.; Cox, T. L. y Knapp, N. (2016). Comparing the OPI and the OPIc: The effect of test method on oral proficiency scores and student preference. En *American Council for the teaching of foreign languages. Foreign Language Annals*, 49, 1. Pp. 75-92
- Trim, J. (2012): Provo Adress. In Tschirner, E. (Ed.). *Aligning Frameworks of Reference in Language Testing: The ACTFL Proficiency Guidelines and the Common European Framework of Reference for Languages*. Tübingen: Stauffenburg Verlag. Pp. 19-22.
- Tschirner, E. (2005). Das ACTFL OPI und der europäische Referenzrahmen. En *Babylonia*, 2005, 2. Pp. 50-55
- Tschirner, E. (2018). Language Testing: Current Practices and Future Developments. En *Die Unterrichtspraxis/ Teaching German*, 51. Pp. 105-120. doi:10.1111/tger.12067
- Universidad de Guadalajara. (2006). Dictamen Núm. I/2006/378
- Universidad de Guadalajara. (2016a). Convenio Marco para servicios de exámenes de idiomas.
- Universidad de Guadalajara. (2016b). Política de Lenguas Extranjeras para la Universidad de Guadalajara. Consultado el 30/04/2018 en http://flip.cga.udg.mx/?page_id=206