

8. Escala de autoeficacia en la enseñanza del inglés: evidencias del proceso de validez



EDUARDO VALLE VELÁZQUEZ*

MARÍA DE LOS ÁNGELES MAYTORENA NORIEGA**

DOI: <https://doi.org/10.52501/cc.402.08>

Resumen

Este capítulo documenta la elaboración, validación de contenido y prueba piloto de la Escala de Autoeficacia en la Enseñanza del Inglés (ECAEIL2), dirigida a futuros profesores de inglés en formación (FPI) en Hermosillo, México. Basada en el modelo de Chacón y Chacón (2010), la escala consta de 20 reactivos distribuidos en tres dimensiones: (1) Motivación Docente para la Enseñanza; (2) Manejo y Disciplina en el Aula; y (3) Evaluación de la Enseñanza. Para asegurar la claridad y pertinencia de los ítems, participaron cuatro jueces expertos con experiencia en investigación sobre autoeficacia educativa, quienes evaluaron cada reactivo en función de coherencia, relevancia, claridad y la suficiencia por dimensión. El Coeficiente de Validez de Contenido (CVC_c) fue de .88, lo que demuestra la validez de contenido de la ECAEIL2. Dos reactivos (13 y 16) con $CVC_{ic} \leq .79$ fueron reformulados para mejorar su comprensión. La prueba piloto, de diseño transversal no experimental, se aplicó a 18 FPI de la Escuela Normal Superior de Hermosillo. Se calcularon estadísticos descriptivos y coeficientes Alfa de Cronbach, obteniéndose $\alpha = .68$ en Motivación Docente, $\alpha = .56$ en Manejo y Disciplina, y $\alpha = .69$ en Evaluación de la Enseñanza. Aunque los

* Maestro en Innovación Educativa. Doctorando en Innovación Educativa, México. ORCID: <https://orcid.org/0009-0004-0307-263X>; correo electrónico: a223230146@unison.mx

** Doctora en Ciencias Sociales. Coordinadora del posgrado en Innovación Educativa. Universidad de Sonora, México. ORCID: <https://orcid.org/0000-0001-9792-6261>

valores fueron moderados, se interpretaron con cautela por el tamaño muestral. Además, un comentario cualitativo permitió ajustar un reactivo adicional. En conclusión, la ECAEIL2 presenta validez de contenido sólida y evidencia preliminar de confiabilidad, sentando las bases para su aplicación en muestras más amplias y estudios de validez de constructo.

Palabras clave: *futuros profesores de inglés, creencias de autoeficacia, prueba piloto, validez de contenido.*

Introducción

La autoeficacia docente, entendida como la creencia del profesor en su capacidad para lograr resultados educativos efectivos (Bandura, 1997), se ha consolidado como un constructo fundamental en el campo de la educación. Su relevancia se ha documentado como un componente de la identidad que desarrollan los docentes en diferentes contextos (Hanna et al., 2019). En el contexto específico de la enseñanza del inglés en México, la autoeficacia adquiere un papel central, dado que permite a los futuros docentes afrontar los retos asociados a la enseñanza de una lengua extranjera en entornos con recursos limitados o estructuras curriculares poco flexibles. Asimismo, ha sido reconocida como un elemento clave para mejorar la calidad educativa (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2023). No obstante, a pesar de contar con literatura sobre el estudio de esta variable en regiones como Asia y Europa, el caso de Latinoamérica ha sido poco explorado (Wyatt, 2018; Hoang, 2018).

El estudio de las creencias de autoeficacia en docentes se remonta a la década de 1970 con el trabajo de Armor et al. (1976), quienes, basándose en el modelo de locus de control de Rotter (1966), propusieron una escala para evaluar el papel de los maestros en el aprendizaje de sus estudiantes. A partir de este trabajo, se desarrollaron diversas escalas con el fin de validar y ampliar la medición de este constructo: la escala de Rose y Medway (1981), la de Guskey (1981), y la escala de Webb y Ashton (1986). Entre los instrumentos más influyentes destacan:

1. Viñetas de eficacia de Ashton et al. (1982): escala de 15 ítems sobre situaciones hipotéticas.
2. Escala de Gibson y Dembo (1984): mide eficacia personal y general en 30 ítems.
3. Teacher Sense of Efficacy Scale (TSES) de Tschannen y Woolfolk (2001): incluye 24 ítems distribuidos en tres factores: eficacia instruccional, gestión del aula y participación del alumnado. Existe también una versión corta de 12 ítems.

Estas escalas han sido utilizadas ampliamente en estudios sobre autoeficacia docente en la enseñanza de lenguas extranjeras (Chacón, 2005; Flynt, 2018), con diversos ajustes según el contexto o el idioma de aplicación (Wyatt, 2014; Klassen et al., 2009). Asimismo, se ha documentado una gran variedad de formatos de respuesta, desde escalas de 5 puntos (Han y Wang, 2021; Moradkhani y Haghi, 2017, 2019) hasta 100 puntos (Swanson, 2010a, 2010b; Swanson, 2012, 2013, 2014), aunque predominan las escalas de 9 puntos y el uso de la TSES (Karas et al., 2024). En algunos casos, la adaptación de la TSES a diferentes contextos solo implica traducir los reactivos al idioma nativo de los participantes, lo que dificulta medir el conjunto de creencias específicas asociadas al contexto de la enseñanza (Akbari y Tavassoli, 2014).

Pese a la proliferación de instrumentos, persisten desafíos importantes en torno a la validez psicométrica y contextual de las escalas utilizadas. En particular, en América Latina existe una producción limitada de estudios empíricos sobre autoeficacia docente en el ámbito de la enseñanza del inglés. Se realizó una revisión sistematizada de literatura, usando el framework de Codina (2018), y solo se identificaron nueve estudios que cumplieran criterios rigurosos en los últimos quince años, publicados en inglés y español. De este análisis surgen dos hallazgos clave: a) predominan los enfoques cualitativos y b) se observan ambigüedades importantes en la documentación de procesos de validación psicométrica.

Esto se alinea con lo reportado por Hoang (2018), quien también identificó tendencias fragmentadas en el estudio de la autoeficacia en contextos no anglófonos. Además, algunos estudios reportan haber sometido escalas a juicio de expertos o pruebas piloto, pero omiten los detalles

metodológicos de dichos procesos (Chacón y Chacón, 2010; Vergara et al., 2020), lo que limita su reproducibilidad y credibilidad científica (Lambert y Newman, 2023).

En retrospectiva, para la elaboración de este capítulo se propuso documentar parte del proceso validez de constructo de una escala construida en el marco de una tesis de maestría, con el objetivo de medir las creencias de autoeficacia docente en profesores de inglés en formación (FPI). La propuesta parte del modelo teórico de Chacón y Chacón (2010) y se diseñaron indicadores clave basados de escalas previas creadas para medir estas creencias en docentes de inglés y FPI (Akbari y Tavassoli, 2014; Karas, 2019).

Fundamentación teórica

En la reciente publicación de la obra *Building the Self-Efficacy Beliefs of English Language Learners and Teachers* (Wyatt y Faez, 2024) dedican un capítulo a describir el estado actual de los instrumentos de medición utilizados para medir Autoeficacia en la enseñanza del inglés. En este apartado los autores plantean cuatro preguntas que servirán como guía. En primera instancia se preguntan qué escalas se han usado en este contexto, seguido de cuáles han sido los niveles de Autoeficacia Docente en función de las escalas identificadas, posteriormente en qué grado afectan diferentes variables demográficas y metodológicas en los niveles de Autoeficacia Docente y, por último, qué se podría considerar para establecer puntos de cohorte y determinar aquellos “altos” o “bajos” niveles de Autoeficacia Docente según los niveles reportados en los trabajos revisados para dicho capítulo.

Atendiendo la primera interrogante los autores encontraron que las escalas utilizadas actualmente podrían categorizarse en tres grupos: Escalas “Originales” (n=50) basadas en la TSES, Escalas “Modificadas” (n=15) de la escala de TSE y “Otras” escalas (n=18) que no utilizaron la escala de las autoras mencionadas o alguno de los factores que proponen (Wyatt y Faez, 2024).

En lo que respecta a las escalas originales, se destaca que la mayoría respetan el formato de respuesta tipo Likert de 5 puntos, pero se reporta la presencia de escalas que utilizan un formato de respuesta tipo Likert de 7 puntos e incluso una escala con tipos de respuesta Likert de 101 puntos (i.e.,

0-100). Algunas de estas escalas reportan haberse administrado en el idioma original que fue publicado y en otros casos fue traducida al idioma de los participantes. De manera general se estimaron los índices de consistencia interna de todos los estudios considerados para dicha revisión, para lo que se reporta una media de Alfa de Cronbach de .87, lo que indica una alta consistencia interna de la TSES. Se reportaron alfas de Cronbach elevadas para las subescalas de Eficacia Instruccional ($M=.82$, $SD = .06$, $CI [.80, .85]$), Eficacia de Gestión/Control del Aula ($M=.84$, $SD = .06$, $CI [.81, .87]$) y Eficacia para Participación del Estudiante ($M=.81$, $SD = .06$, $CI [.79, .84]$), lo que denota una buena consistencia interna del instrumento.

Sin embargo, mejorar la especificidad de los ítems puede hacer que el trabajo cuantitativo de autoeficacia docente de lenguas extranjeras sea más práctico para los profesores de idiomas. Wyatt (2014) ha señalado que cuando los elementos tienen una perspectiva global, puede ser difícil entender completamente qué están evaluando realmente los profesores al completar las escalas. Sin embargo, al observar las escalas específicas del dominio de lenguas extranjeras y las escalas de competencia lingüística aún más específicas de Zhang (2018), Wyatt y Dikilitaş (2021), Akbari y Tavassoli (2014) y Karas (2019) encontramos elementos más importantes de tareas para que los profesores los interpreten. Hay numerosas otras áreas para que los investigadores de autoeficacia del profesor de idiomas exploren y las escalas que tienen un enfoque aún mayor podrían ser útiles para ayudar a los profesores a determinar áreas/tareas específicas de enseñanza de lenguas extranjeras en las que tienen una alta eficacia, o áreas/tareas donde podrían buscar mejorar y ganar más confianza. Un hallazgo importante para considerar en la contextualización de instrumentos de medición sobre Autoeficacia Docente es lo reportado por Thommen et al. (2022), quienes encontraron una alta correlación entre la calidad de enseñanza, referida por estudiantes de matemáticas, e indicadores de Autoeficacia Docente contextualizados a la enseñanza en comparación a indicadores generales de Autoeficacia Docente.

Como parte del objetivo principal de este trabajo, el modelo a adaptar es la propuesta publicada por Chacón y Chacón (2010) su medida sobre Autoeficacia Docente en profesores de inglés como lengua extranjera en secundaria desde el contexto venezolano. Los autores plantearon tres dimensiones derivadas de la revisión de literatura y con base en modelos de Autoeficacia

Docente (Armor et al., 1976; Riggs y Enoch, 1990). A grandes rasgos, se destaca la identificación de un modelo emergente que, aunque presenta solapamientos con modelos clásicos e inconsistencias en la descripción de la validez de contenido, ofrece un buen ajuste a los datos y sirve como punto de partida para futuras investigaciones. Este trabajo es particularmente relevante por ser uno de los pocos que aborda de manera específica la autoeficacia docente en el contexto de la enseñanza del inglés, destacando la necesidad de instrumentos de medición adaptados a contextos particulares. Hasta este momento, no se ha probado el modelo de Chacón y Chacón (2010) en otros países y aparentemente, este trabajo es el único intento por someterlo a prueba.

Definición operacional de Autoeficacia Docente en la Enseñanza del Inglés

En la Tabla 8.1 se describe el modelo teórico de Chacón y Chacón (2010) con nueve indicadores adicionales que se diseñaron para su adaptación al contexto de este estudio. Este modelo quedó conformado por un total de veinte reactivos agrupados en tres dimensiones: Autoeficacia y Motivación Docente para enseñar (7 reactivos), Autoeficacia y Manejo de la Disciplina en el Aula (7 reactivos), y Autoeficacia y Evaluación de la enseñanza (6 reactivos).

La primera dimensión, Autoeficacia y Motivación Docente para enseñar, alude a “la percepción del docente en función de sus capacidades para influir en el aprendizaje de los estudiantes” (Chacón y Chacón, 2010, p. 4) y lo componen siete ítems. Estos ítems aluden a la capacidad del docente para motivar a los estudiantes, influir en su percepción de sus propias capacidades, valorar el aprendizaje y mantener un compromiso constante con el proceso educativo. La motivación, según la teoría de la autoeficacia, determina las metas planteadas, el esfuerzo dedicado y el tiempo invertido en la perseverancia para vencer las dificultades, así como su resiliencia al fracaso.

La dimensión de Autoeficacia y Manejo de la Disciplina en el Aula se refiere como la eficacia percibida del docente en el diseño de estrategias para motivar y controlar la disciplina de los estudiantes conformado por siete reactivos (Chacón y Chacón, 2010). Los autores consideran esta dimensión como fundamental, puesto que el manejo de la disciplina y la capacidad para

promover un ambiente de aprendizaje efectivo son elementos considerados por Tschannen y Woolfolk (2001) para explicar la Autoeficacia Docente.

Tabla 8.1. Operacionalización del modelo adaptado de Chacón y Chacón (2010)

Variable	Definición Conceptual	Definición Operacional	Indicadores	Nivel de Medición
Motivación Docente para la Enseñanza. (Chacón y Chacón, 2010)	“Alude a la percepción del Docente sobre sus capacidades para influir en el aprendizaje de sus estudiantes.” Chacón y Chacón (2010)	7 indicadores con Formato de Respuesta Tipo Likert 1 (Nada) – 10 (Mucho). Sumas superiores a la media teórica (M = 35) indican una autoeficacia percibida alta para influir en el aprendizaje de los estudiantes. Sumas inferiores a la media teórica indican una baja autoeficacia percibida.	R2, R3, R4, R11, R14*, R17* y R19*.	1 Nada – 10 Mucho. Escala Intervalar. Nivel mínimo de 7 y Máximo de 70.
Manejo y Disciplina en el Aula. (Chacón y Chacón, 2010)	“Eficacia percibida en el diseño de estrategias para motivar y controlar la disciplina”. Chacón y Chacón (2010)	7 indicadores con Formato de Respuesta Tipo Likert 1 (Nada) – 10 (Mucho). Sumas superiores a la media teórica (M = 35) indican una autoeficacia percibida alta en el diseño de estrategias para motivar y controlar la disciplina. Sumas inferiores a la media teórica indican una baja autoeficacia percibida.	R1, R6, R7, R8, R12*, R13* y R20*.	1 Nada – 10 Mucho. Escala Intervalar. Nivel mínimo de 7 y Máximo de 70.
Evaluación de la Enseñanza. (Chacón y Chacón, 2010)	“Son aquellas creencias relacionadas a la capacidad de formular preguntas, generar estrategias de evaluación y de tomar medidas alternativas para evaluar el aprendizaje.” Chacón y Chacón (2010)	6 indicadores con Formato de Respuesta Tipo Likert 1 (Nada) – 10 (Mucho). Sumas superiores a la media teórica (M = 30) indican una alta autoeficacia percibida para formular preguntas, generar estrategias de evaluación y de tomar medidas alternativas para evaluar el aprendizaje. Sumas inferiores a la media teórica indican una baja autoeficacia percibida.	R5, R9, R10, R15*, R16* y R18*.	1 Nada – 10 Mucho. Escala Intervalar. Nivel mínimo de 6 y Máximo de 60.

Fuente: Elaboración propia.

La dimensión de Autoeficacia y Evaluación de la enseñanza se refiere a aquellas creencias relacionadas con la capacidad de formular preguntas, generar estrategias de evaluación y de tomar medidas alternativas para evaluar el aprendizaje (Chacón y Chacón, 2010). Esta es una dimensión

informada con cierta regularidad en los estudios que abordan el constructo de autoeficacia docente (Chacón, 2005; Tschannen y Woolfolk, 2001; Karas et al., 2024). Dicha dimensión se relaciona con la experiencia de los docentes y el conjunto de instrumentos o medios y estrategias utilizados para planificar la evaluación, así como la innovación de estrategias de aprendizaje. Además, se destaca que la evaluación basada en la negociación tiende a ser el modelo adoptado por los docentes que informan un alto sentido de autoeficacia docente (Chacón y Chacón, 2010) y lo conforman seis indicadores o reactivos.

Proceso de construcción de instrumento

El formato de respuesta que emplearon los autores originales fue de tipo Likert de 5 puntos donde 1=Nada y 5=Mucho en una muestra de docentes de inglés venezolanos (n=104). Sin embargo, para este trabajo se siguieron las recomendaciones sobre la relación respecto a la sensibilidad del instrumento y el aumento de las opciones de respuesta (Bisquerra y Pérez, 2015) por lo que se empleó un formato de respuesta tipo Likert de 10 puntos respetando los anclajes mencionados anteriormente.

Adicionalmente, se identificaron instrumentos que presentan indicadores relevantes en este contexto de enseñanza (Akbari y Tavassoli, 2014; Karas, 2019) por lo que se consideró pertinente diseñar 9 indicadores derivados de esta revisión de literatura y se distribuyeron en las tres dimensiones descritas en los apartados anteriores. Este proceso resultó en un modelo teórico donde el constructo de Creencias de Autoeficacia Docente en la Enseñanza del Inglés (CADEIL2) se define como un factor de segundo orden conformado por tres factores de primer orden donde se emplean un total de 20 indicadores para su medición que no requieren de ser codificados de manera inversa.

Validez de contenido

Previo a cualquier aplicación de la prueba, se revisó literatura pertinente sobre los diferentes métodos para calcular la Validez de Contenido en escalas tipo

Likert, por lo que resulta de interés discutir este tema. La validez de contenido alude a una etapa del proceso de diseño y validación de pruebas evaluativas en educación y psicología. Esta consiste en determinar la pertinencia de los ítems que conforman un instrumento de medición con la teoría y su congruencia para medir el constructo en cuestión, usualmente determinado inter-jueces, calculándose por índices de correlación con valores de 0 a 1, donde se interpretan como pertinentes aquellos ítems y dimensiones con valores cercanos a 1 o superiores a .8 (Mousalli, 2017; Maldonado y Santoyo, 2024; Pedrosa et al., 2014; Sánchez, 2021).

De las técnicas comúnmente empleadas se destacan los diferentes criterios y formatos de evaluación que se requieren para garantizar una aplicación adecuada del mismo, principalmente la cantidad de expertos que se necesitan. La técnica propuesta por Hernández (2002) resulta atractiva por diferentes características en su aplicación. El Coeficiente de Validez de Contenido (CVC_i) (Hernández, 2002) es un procedimiento utilizado en diversas disciplinas para satisfacer este criterio (Sánchez et al., 2023; Zaldívar et al., 2023) y requiere de la participación de al menos tres a cinco jueces expertos. Como resultado, el modelo teórico fue sometido a Validez de Contenido por un panel de expertos (n=4) con la intención de calcular el CVC_i de la Escala de Autoeficacia en la enseñanza del Inglés (ECAEIL2).

Los criterios para seleccionar a los candidatos fueron contar con al menos tres publicaciones en los últimos 5 años en materia de Autoeficacia, Autoeficacia Docente y/o Autoeficacia Docente en la enseñanza del inglés, así como tener más de 5 años desempeñándose como investigadores en México. Se les contactó por medio de correo electrónico siguiendo los lineamientos sugeridos para este proceso (Hernández, 2002; Escobar y Cuervo, 2008; Galicia et al., 2017). El perfil de los expertos que participaron en la validez de contenido se describe en la tabla 8.2, donde se destaca la participación de investigadores de tres instituciones de educación superior en México, con un promedio de 14 años (Min. = 6 y Max. = 20) de experiencia en investigación. Todos mencionaron pertenecer al Sistema Nacional de Investigadoras e Investigadores (SNI). Por último, los participantes reportaron su formación inicial en disciplinas como Psicología y Ciencias de la Educación.

Tabla 8.2. *Perfil de expertos participantes*

<i>Grado</i>	<i>Institución</i>	<i>Formación Inicial</i>	<i>Experiencia en AE.</i>	<i>Años en investigación.</i>
Doctorado	Instituto Tecnológico de Sonora.	Lic. en Psicología	3 publicaciones.	20 años.
Doctorado	Universidad Autónoma de Quintana Roo.	Lic. en Enseñanza del Francés	15 publicaciones.	20 años.
Doctorado	Universidad Autónoma de Coahuila.	S/información	5 publicaciones.	10 años.
Doctorado	Instituto Tecnológico de Sonora.	Lic. en Educación Especial	3 publicaciones.	6 años.

Fuente: Elaboración propia.

Para recabar la opinión de los jueces expertos, se diseñó una página web siguiendo las recomendaciones de Galicia et al. (2017), quienes utilizan el formato de Escobar y Cuervo (2008). Este consiste en tres indicadores que evalúa aspectos directamente sobre los reactivos o ítems y un indicador para evaluar las dimensiones del instrumento. Estos son: Coherencia, Relevancia, Claridad y Suficiencia. El primero alude a la relación lógica con la dimensión o indicador que se está midiendo; el segundo consiste en determinar si el ítem es esencial o importante y por lo tanto debe de ser incluido y, por último, para los indicadores que evalúan los reactivos de manera individual; el tercero alude a si el ítem se comprende fácilmente en el sentido de sintaxis y semántica son adecuadas. El cuarto indicador consiste en la evaluación de las dimensiones que conforman el instrumento en función, si los ítems que pertenecen a una misma dimensión bastan para obtener la medición de esta. Para responder el formato de evaluación se manejó una escala Likert de cuatro puntos (1- No Cumple y 4- Alto Nivel) a manera que los jueces pudieron valorar en qué medida los reactivos y dimensiones cumplen o no cumplen con los criterios de cada indicador evaluador; asimismo, se puso a disposición de los jueces participar con sugerencias u observaciones generales y específicas acerca de la escala en un banco de comentarios.

De igual manera se incluyó una sección sobre indicadores demográficos en la que se pregunta sobre la institución donde laboran, formación inicial, grado de estudios actual, si son miembros del SNII, experiencia en investigación sobre Autoeficacia y años de experiencia en investigación. Es posible observar la página web en el siguiente enlace: tarea.seison.com.mx. Posterior

al levantamiento de información, las respuestas de los expertos se cuantificaron para calcular el CVC_t (Hernández, 2002).

El CVC_t o Validez de Contenido del Instrumento se define como la proporción relativa con respecto del valor máximo de la escala del promedio de los puntajes entre jueces por cada ítem (CVC_{ic}), corregido por concordancia aleatoria o sesgo del panel (P_{ei}) por ítem (Hernández, 2002; Maldonado y Santoyo, 2024; Hurtado-Arenas et al., 2025). Es necesario seguir una serie de pasos para su cálculo.

Primero, debe realizarse la sumatoria de los puntajes asignados por cada juez (j) a cada uno de los ítems ($\sum x_i$). Una vez obtenida la suma de cada ítem, es necesario estimar el promedio en cada uno de los puntajes asignados a los reactivos ($Mx = \sum xi/j$). Posteriormente, dividir los promedios (Mx) entre el valor máximo que puede alcanzar cada ítem ($Vmax$) para definir la Validez de Contenido por ítem (CVC_i). Una vez identificado cada CVC_i se requiere corregir por la probabilidad de concordancia aleatoria (P_{ei}) y este se obtiene dividiendo 1 entre la cantidad de jueces participantes ($j=4$) elevado a la potencia de la cantidad de jueces expertos participantes, en este caso se representa por $P_{ei}=.04$ y es un valor constante; esta operación se representa como $CVC_{ic} = CVC_i - P_{ei}$. Por último, para definir el nivel de concordancia y Validez de la escala (CVC_t) se logra con el promedio de la sumatoria de los valores de CVC_{ic} y se representa como $CVC_t = \sum CVC_{ic} / N$.

El autor original sugiere un punto de cohorte superior a .8 para interpretar los resultados como satisfactorios, a manera que aquellos ítems con valores superiores a .8 poseen una alta concordancia inter-jueces (CVC_{ic}) y apropiada la Validez de Contenido del instrumento (CVC_t) (Hernández, 2002). Sánchez (2021) propone los siguientes puntos de cohorte:

- a) Menor a .6, validez y concordancia inaceptable.
- b) Igual o mayor de .6 y menor a .7, validez y concordancia deficientes.
- c) Mayor que .71 y menor o igual que .8, validez y concordancia aceptables.
- d) Mayor que .8 y menor o igual a .9, validez y concordancia buenas.
- e) Mayor a .9, validez y concordancia excelentes.

La ECAEIL2 obtuvo como resultado un $CVC_t = .88$, lo que sugiere una Validez de Contenido aceptable (Hernández, 2002), garantizando la estructura del modelo teórico durante este proceso. Para describir los resultados de este proceso se elaboró una tabla de frecuencia en función de los puntos de cohorte propuestos por Sánchez (2021) con respecto al CVC_{ic} por ítem (ver tabla 8.3): la mayoría de los ítems sometidos a validación por expertos obtuvieron índices de validez y concordancia buenos ($n=12$) y excelente ($n=6$) con lo que se demuestra que la estructura original permanece aun cuando dos ítems (Pregunta 13 y Pregunta 16) obtuvieron índices inferiores o iguales a .79.

Tabla 8.3. *Frecuencia de ítems según valores de Sánchez (2021)*

<i>Punto de Cohorte</i>	<i>Frecuencia de Ítems</i>
>.71 y ≤.8 (Validez y Concordancia Aceptable)	2
>.8 y ≤.9 (Validez y Concordancia Buena)	12
>.9 (Validez y Concordancia Excelente)	6
Total	20

Fuente: Elaboración propia.

Los expertos destacaron la claridad y pertinencia de los ítems que se evaluaron, sugiriendo modificaciones en la redacción de algunos para mejorar la precisión en la interpretación. Por ejemplo, se recomendó ajustar términos como ‘control del aula’ para hacerlo más específico al contexto de enseñanza del inglés, resultando en la frase ‘capacidad para mantener el enfoque en actividades de aprendizaje del idioma’.

En lo que respecta a los ítems con valores inferiores a .8 (Pregunta 13 y Pregunta 16) se realizaron ajustes en cuestión de gramática y redacción (ver tabla 8.4) sustanciales a partir de los comentarios de los expertos .8. Al revisar la matriz de respuestas se identificó que la proporción relativa con respecto del valor máximo de la escala del promedio de los puntajes entre jueces en estos ítems es afectado por las respuestas obtenidas en el indicador de Claridad, el cual alude a las características mencionadas al inicio de este enunciado. Al realizar estas modificaciones se establece la pertinencia de mantener estos ítems en la ECAEIL2, sin embargo, se recomienda tomar en cuenta estos hallazgos a la hora de realizar modificaciones a la estructura de la escala durante los análisis de confiabilidad requeridos para determinar la consistencia interna del instrumento en etapas posteriores de aplicación.

Tabla 8.4. *Redacción original de ítems y reactivo modificado*

<i>No.</i>	<i>Ítem original</i>	<i>Ítem modificado</i>
13.	¿Cuánto puedes hacer tú para que un alumno no interrumpa tus clases?	¿En qué medida crees que los límites o reglas que estableces pueden ayudar a que los estudiantes eviten interrumpir las clases?
16.	¿Cuánto puedes hacer tú para retroalimentar los resultados de tus evaluaciones a las autoridades escolares?	¿Cuánto crees que puedes hacer para presentar las retroalimentaciones de los resultados de tus estudiantes a las autoridades?

Fuente: Elaboración propia.

En conclusión, esta etapa del proceso de validez confirmó que la mayoría de las preguntas resultan coherentes, relevantes y claras al igual que suficientes en su conjunto para medir el constructo de Autoeficacia docente en la enseñanza del inglés. La versión final de la escala para la prueba piloto quedó conformada por 20 ítems tras las modificaciones propuestas.

Prueba piloto

Una vez alcanzado el criterio de validez de contenido, describiremos el proceso llevado a cabo durante la prueba piloto, no sin antes resaltar brevemente la importancia de esta etapa. En la literatura, las pruebas piloto se definen como estudios de factibilidad, a pequeña escala, de los métodos de recopilación y análisis de datos planeados para un estudio a gran escala (Lowes, 2019; Teresi et al., 2022). Si bien, cuando se diseña un instrumento es de esperarse que los puntajes presenten una consistencia interna (Hayes et al., 1987), estos estudios nos permiten anticipar problemas que podrían incurrir en la aplicación de una escala (van Teijlingen y Hundley, 2001), por lo que se recomienda que las pruebas piloto se lleven a cabo posterior a la validez de contenido del instrumento (Carpenter, 2018; Lambert y Newman, 2023). Su flexibilidad metodológica permite orientar su diseño bajo diferentes propósitos u objetivos (Díaz, 2020).

Por lo tanto, la prueba piloto se desarrolló no solo con la intención de afinar el instrumento, sino también de contribuir al fortalecimiento metodológico de estudios similares en la región, proporcionando una descripción detallada de los eventos que transcurrieron en este proceso. La metodología empleada en esta prueba se describe como un diseño no experimental de tipo

transversal con un muestreo no probabilístico por conveniencia (Kerlinger, 2002) dada la poca cantidad de FPI que disponen las instituciones participantes ($n=3$). En consonancia con lo revisado, el objetivo de la prueba piloto fue estimar la consistencia interna del instrumento, poner a prueba la escala y mejorar los reactivos de esta a raíz de los comentarios emitidos por los participantes en esta fase.

Contexto de trabajo

Resulta sumamente importante revisar el contexto en el que se llevó a cabo este trabajo, pues en la ciudad donde se realizó la prueba se lograron identificar cuatro Instituciones de Educación Superior (IES) que ofrecen diferentes planes de estudio orientados a formar profesores de inglés o inglés y francés. Por un lado, la asincronía entre las mallas curriculares de las instituciones identificadas radica en el punto en que los estudiantes comienzan a involucrarse en experiencias de desempeño, lo que de antemano podría señalar que se encuentren diferentes niveles de autoeficacia entre la población local. Esta falta de unificación curricular es una problemática reconocida en México, pues se ha visto que afecta la motivación y el desempeño en la enseñanza del inglés (Ramírez, 2023; Medel, 2023). De forma paralela, se reconoce la necesidad de incluir en los planes de estudio, de manera explícita, el uso de estrategias metacognitivas para favorecer la enseñanza del inglés en México (Martínez et al., 2016). Por otro lado, solo se estableció contacto con tres de las instituciones para invitar a estudiantes que estuvieron cursando Prácticum o materias equivalentes donde comienzan a desarrollar experiencias de desempeño que influyen en el desarrollo de creencias de autoeficacia (Cárcamo, 2023; Murrieta y Reyes, 2019).

Aunado a esto, para la prueba piloto se consideró a los Futuros Profesores de Inglés (FPI) de 5° y 7° semestre de la Escuela Normal Superior de Hermosillo ($n=21$) los cuales llevan un eje de práctica docente desde el primer semestre. Sin embargo, no se contó con la participación de tres estudiantes, quienes no pudieron acudir por cuestiones académicas (intercambios estudiantiles) y médicas, por lo que al final participaron 18 FPI (11 mujeres y 7 hombres) de 5° y 7° de la Lic. en Enseñanza del Inglés con una media de edad

de 21 años (min. = 19 y máx.= 24) de los cuales 11 llevan a cabo sus prácticas docentes en educación media superior, 1 en educación superior y el resto en educación básica.

En lo que concierne al proceso de recogida de información, se llevó a cabo en el aula mediante papel y lápiz en el mes de noviembre del 2024 en un tiempo estimado de 10-15 minutos por sesión. Durante este proceso, se consideró pertinente primero exponer la información recopilada en la Guía de Consideraciones Éticas/Consentimiento Informado (ver Anexo A). Esta guía contiene información sobre el título del estudio, instituciones participantes, propósito del estudio y de la prueba piloto, procedimiento, posibles riesgos y molestias y los beneficios que involucran su participación en el estudio. Una vez obtenido el consentimiento de manera verbal por parte de los estudiantes, se pasó a entregar las hojas de consentimiento informado para ser firmadas, en las que se declaran que los puntos de la Guía de Consideraciones Éticas/Consentimiento Informado fueron comprendidos y que su participación es confidencial y voluntaria. De igual manera se les hizo saber que podían abandonar el estudio en cualquier momento sin que esto comprometiera su situación académica o de algún otro tipo. Se les invitó a tomarse su tiempo para leer los reactivos, pues se les hizo saber que estarían participando en la prueba piloto y que parte del objetivo era que hicieran comentarios acerca de los reactivos, en caso de que alguno de ellos les causara algún tipo de confusión para que la escala fuera mejorada previo a su aplicación final.

Posteriormente, los instrumentos fueron capturados en el Paquete Estadístico para Ciencias Sociales (SPSS). La depuración de la base de datos no fue necesaria, pues se garantizó que los participantes no dejaran espacios en blanco en las hojas de respuesta, ya que se revisaron los instrumentos en el momento que los participantes terminaban de contestarlos. Una vez cargados los datos se corrieron análisis descriptivos, de distribución e índices de consistencia interna (Alfa de Cronbach) por dimensión. Se acota brevemente sobre la decisión de utilizar este estadígrafo, debido a su amplio reconocimiento para describir la confiabilidad-replicabilidad de los resultados en una investigación (Hernández et al., 2020; Cho, 2022; Lambert y Newman, 2023; Colorado et al., 2025).

Como resultado, durante la prueba piloto uno de los participantes señaló una inconsistencia en cuanto a la redacción en uno de los ítems (R17). Cabe

resaltar que fue el único comentario realizado durante la ejecución de la prueba piloto. Este ajuste fue realizado y tomado en cuenta para la fase posterior en el proceso de validación. Asimismo, los resultados de la estimación de la consistencia interna, el promedio de las sumas, la varianza y la cantidad de reactivos por dimensión se muestran en la tabla 8.5.

Tabla 8.5. *Estimación de consistencia interna por dimensión*

<i>Variable</i>	<i>R</i>	<i>Min</i>	<i>Max.</i>	<i>M</i>	<i>S²</i>	<i>α</i>
Motivación Docente para la enseñanza	7	1	10	59.06	26.40	.68
Manejo y Disciplina en el Aula	7	1	10	54.78	22.53	.56
Evaluación de la Enseñanza	6	1	10	50.17	36.97	.69

Fuente: Elaboración propia.

Reflexiones metodológicas

Esta prueba piloto aporta información importante sobre dos elementos clave para la toma de decisiones. En primer lugar, los resultados de los análisis de consistencia interna pueden ser alarmantes a simple vista, por sus valores debajo de lo mínimo aceptable. No obstante, el tamaño de la muestra, el tipo de estadígrafo, las características de los datos, entre otros factores resultan importantes en los resultados de este análisis. Retomando la decisión de usar Alfa de Cronbach (α) para reportar la consistencia interna de los datos, se reconoce que es un método comúnmente empleado (Malkewitz et al., 2023) y que muchas veces se utiliza erróneamente (Hussey et al., 2023), sin embargo, el reconocimiento que tiene la comunidad científica acerca de este estadígrafo fue suficiente para utilizarlo con fines de difusión y sin olvidar que uno de los objetivos de esta prueba piloto fue la estimación de la consistencia interna de los resultados.

Las disparidades entre los índices de confiabilidad podrían deberse en parte a lo reportado por Chacón y Chacón (2010) donde el modelo sufrió cambios sustanciales como el solapamiento de dos dimensiones en una sola y las modificaciones que sufrió la dimensión que permaneció o incluso por cómo se comporta el estadístico en muestras pequeñas ($n < 60$) (Cho, 2021) y por las propiedades de los datos (Trizano y Alvarado, 2016). Es por esto

último que no se consideró pertinente llevar a cabo modificaciones a la estructura del instrumento a pesar de haber revisado los coeficientes de confiabilidad que se obtendría si se elimina cierto ítem.

Otra causa por la que no se realizaron modificaciones en la estructura del instrumento fue la necesidad de ajustar los reactivos a raíz del comentario señalado anteriormente, lo que corrobora la importancia de someter a prueba el instrumento de medición que se desea emplear. Con este trabajo se pretende divulgar una parte del proceso sistemático llevado a cabo para desarrollar un modelo válido y confiable o bien la validación de constructo de las Creencias de Autoeficacia en FPI para su estudio en una región de México. Las sugerencias para realizar pruebas piloto y pre-tests varían en investigación, sin embargo, con este trabajo se pretende plasmar la importancia de valorar los reactivos por medio de entrevistas cognitivas en algún momento de la aplicación. Aun cuando estas fueron informales, la información verbal de los encuestados permite valorar la claridad en diferentes apartados del instrumento: consentimiento informado, instrucciones, apartado sociodemográfico, redacción de preguntas y formato de respuesta, por mencionar algunos. Gracias a esto se logró identificar un área a mejorar en nuestro instrumento, por lo que se recomienda ampliamente este ejercicio durante la etapa previa a la aplicación de cualquier escala.

Limitaciones del estudio

En primera instancia, resulta importante constatar que el diseño y muestreo no probabilístico intencional de tipo transversal empleado en este trabajo no permite la generalización de sus resultados, lo que implica un obstáculo para la interpretación de estos resultados a una población más amplia. En segundo lugar, el tamaño de la muestra ($n=18$) comprende otro aspecto a delimitar, pues se tiene registro de la relación estrecha que mantienen los estadígrafos y su potencia de análisis con respecto a la cantidad de observaciones que son sometidas a prueba (Malkewitz et al., 2023). En tercer lugar, al ser un trabajo donde parte de sus objetivos fue estimar la consistencia interna de una escala con validez de contenido indica que el instrumento de medida

aún se encuentra en una etapa preliminar para su uso. En consecuencia, los resultados que se muestran en este capítulo deben considerarse como una primera aproximación a las propiedades psicométricas de un modelo teórico del cual no se tiene registro previo de su uso en México, por lo que se recomienda interpretar estos resultados con precaución.

Conclusiones

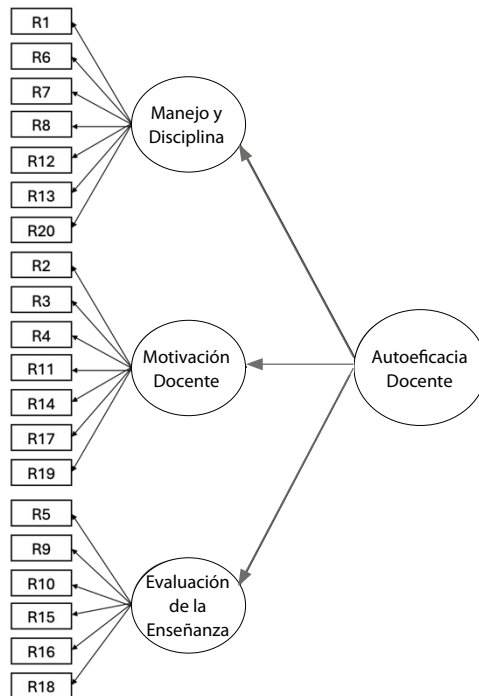
El presente estudio documenta de manera detallada el proceso de validación de contenido y la prueba piloto de la Escala de Autoeficacia en la Enseñanza del Inglés (ECAEIL2), adaptada a futuros profesores de inglés en formación (FPI) en el contexto de Hermosillo. A través de un panel de cuatro jueces expertos, se obtuvo un Coeficiente de Validez de Contenido total (CVC_t) de .88, lo cual indica una validez de contenido “buena” según los umbrales de Sánchez (2021). La evaluación de los ítems permitió identificar oportunidades de mejora en la redacción —especialmente en los reactivos 13 y 16— y culminó en una versión final de 20 reactivos coherentes, relevantes, claros y suficientes para medir el constructo de autoeficacia docente en inglés.

La prueba piloto con 18 FPI reveló valores de consistencia interna moderados ($\alpha = .68$ para Motivación Docente, $\alpha = .56$ para Manejo de Disciplina y $\alpha = .69$ para Evaluación de la Enseñanza), lo cual puede deberse, en gran medida, al tamaño reducido de la muestra, por lo que se interpretó con suma precaución. Además, el único comentario de un participante (reactivo R17) confirmó la utilidad de incorporar entrevistas cognitivas informales para afinar la comprensión de los reactivos antes de su aplicación a gran escala. En este sentido, es necesario hablar sobre el tamaño de la muestra. De antemano, se identificó la poca cantidad de FPI que estuvieran cursando Prácticum, efecto que se atribuye a la disparidad entre los planes de estudio que poseen las instituciones que formaron parte de la muestra de estudio; aspecto que resultó importante para la decisión de emplear específicamente esta población con la intención de no poner en riesgo el tamaño de la muestra en la fase posterior del proceso de validez.

Como línea futura, se recomienda aplicar el modelo teórico (ver figura 8.1) a una muestra representativa mayor para realizar Análisis Factorial

Confirmatorio (AFC), calcular coeficientes de fiabilidad alternativos (Omega de McDonald) y recabar evidencia de validez de constructo (convergente y divergente) con la intención de orientar esfuerzos por mejorar la formación de FPI con un modelo de medición válido y confiable en la región. En conjunto, este trabajo no solo aporta un instrumento específico para medir las creencias de autoeficacia en la enseñanza del inglés en contextos latinoamericanos, sino que también ejemplifica un camino metodológico transparente y replicable: desde la fundamentación teórica y la elaboración de ítems hasta la validación de contenido y prueba piloto. De esta manera, se sientan las bases para estudios posteriores que profundicen en las dimensiones psicométricas de la escala y, en esa medida, contribuyan a mejorar la formación y la práctica profesional de los FPI.

Figura 8.1. Método teórico adaptado de Chacón y Chacón (2010)



Fuente: Elaboración propia.

Referencias

- Akbari, R., Ghafar Samar, R., Kiany, G. R., y Tahernia, M. (2017). A Qualitative Study of EFL Teachers' Emotion Regulation Behavior in the Classroom. *Theory and Practice in Language Studies*, 7(4), 311-321. <http://dx.doi.org/10.17507/tpis.0704.10>
- Akbari, R. y Tavassoli, K. (2014). Developing an ELT Context-Specific Teacher Efficacy Instrument. *RELC Journal*, 45(1), 27-50. <https://doi.org/10.1177/0033688214523345>
- Armor, D., Conroy, P., Cox, M., King, N., McDonnell, L., Pascal, A., Pauly, E. y Zellman, G. (1976). *Analysis of the school preferred reading programs in selected Los Angeles minority schools*. RAND.
- Ashton, P., Olejnik, S., Crocker, L. y McAuliffe, M. (1982, 19-23 de marzo). *Measurement problems in the study of teachers' sense of efficacy* [Conferencia]. Annual meeting of the American Educational Research Association, New York, Estados Unidos.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
- Bisquerra, R. y Pérez, N. (2015). ¿Pueden las escalas Likert aumentar en sensibilidad? *Revista d'Innovació i Recerca en Educació*, 8, 129-147. <https://doi.org/10.1344/reire2015.8.2828>
- Cárcamo, B. (2023). El impacto de una experiencia de enseñanza en la autoeficacia de los profesores de Inglés como Lengua Extranjera (ILE) en etapa de formación. *Revista de estudios y experiencias en educación*, 22(50), 336-354. <http://dx.doi.org/10.21703/rexe.v22i50.1759>
- Carpenter, S. (2018). Ten Steps in Scale Development and Reporting: A Guide for Researchers. *Communications Methods and Measures*, 12(1), 25-44. <https://doi.org/10.1080/19312458.2017.1396583>
- Chacón, C. (2005). Teachers' perceived efficacy among English as a foreign language teacher in middle schools in Venezuela. *Teaching and Teacher Education*, 21(3), 257-272. <https://doi.org/10.1016/j.tate.2005.01.001>
- Chacón, E. y Chacón, C. (2010). Un Modelo para Medir el Sentido de Autoeficacia Docente en Profesores de Inglés como Lengua Extranjera en Secundaria. *Evaluar*, 10(1), 1-21. <https://doi.org/10.35670/1667-4545.v10.n1.456>
- Cho, E. (2021). Neither Cronbach's Alpha nor McDonald's Omega: A commentary on Sijtsma and Pfadt. *Psychometrika*, 86(4), 877-886. <https://doi.org/10.1007/s11336-021-09801-1>
- Cho, E. (2022). The Accuracy of Reliability Coefficients: A Reanalysis of Existing Simulations. *Psychological Methods*, 29(2), 331-349. <https://doi.org/10.1037/met0000475>
- Codina, L. (2018). Revisiones bibliográficas sistematizadas. Procedimientos generales y Framework para Ciencias Humanas y Sociales. *Universitat Pompeu Fabra*, 87. <https://repositori.upf.edu/items/224872e9-6f79-4666-8ad9-07aad198aa97?locale=es>
- Colorado, J., Romero, M., Salazar, M., Cabrera, G. y Castillo, V. (2025). Análisis Comparativo de los Coeficientes Alfa de Cronbach, Omega de McDonald y Alfa Ordinal en la Validación de Cuestionarios. *Estudios y Perspectivas Revista Científica y Académica*, 4(4), 2738-2755. <https://doi.org/10.61384/r.c.a..v4i4.836>

- Díaz, G. (2020). Metodología del estudio piloto. *Revista chilena de radiología*, 26(3), 100-104. <http://dx.doi.org/10.4067/S0717-93082020000300100>
- Escobar, J. y Cuervo, A. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición*, 6(1), 27-36.
- Flynt, A. (2018). *Tennessee ESL teachers' self-efficacy: A predictive correlational study* [Unpublished doctoral thesis]. The University of Memphis, Memphis, TE.
- Galicia, L., Balderrama, J. y Edel, R. (2017). Validez de Contenido por juicio de expertos: una propuesta de herramienta virtual. *Apertura*, 9(2), 42-53. <https://doi.org/10.32870/ap.v9n2.993>
- Gibson, S. y Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, 76(4), 569-582. <https://psycnet.apa.org/doi/10.1037/0022-0663.76.4.569>
- Guskey, T. (1981). Measurement of responsibility teachers assume for academic successes and failures in the classroom. *Journal of Teacher Education*, 32, 44-51. <https://doi.org/10.1177/002248718103200310>
- Han, Y. y Wang, Y. (2021). Investigating the Correlation Among Chinese EFL Teachers' Self-efficacy, Work Engagement, and Reflection. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.763234>
- Hanna, F., Oostdam, R., Severiens, S. y Zijlstra, B. (2019). Domains of teacher identity: A review of quantitative measurement instruments. *Educational Research Review*, 27, 15-27. <https://doi.org/10.1016/j.edurev.2019.01.003>
- Hayes, S., Nelson, R. y Jarrett, R. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist*, 42(11), 963-974. <https://doi.org/10.1037/0003-066X.42.11.963>
- Hernández, A., Hidalgo, M., Hambleton, R. y Gómez, J. (2020). International Test Commission guidelines for test adaptation: A criterion checklist. *Psicothema*, 32(3), 390-398. <https://doi.org/10.7334/psicothema2019.306>
- Hernández, R. (2002). El Coeficiente de Validez de Contenido (CVC) y el Coeficiente Kappa, en la determinación de la Validez de Contenido según la técnica de Juicio de Expertos. En R. Hernández (Ed.), *Contribuciones al Análisis Estadístico* (pp. 107-168). Mérida.
- Hoang, T. (2018). Teacher Self-efficacy Research in English as Foreign Language Contexts: A Systematic Review. *The Journal of AsiaTEFL*, 15(4), 976-990. <https://doi.org/10.18823/asiatefl.2018.15.4.6.976>
- Hurtado-Arenas P., Guevara, M. R. y González-Chordá, V. M. (2025). Comparación de índices de validez de contenido para investigación en enfermería clínica: Un caso práctico. *Enfermería Clínica*, 35(3). 502214. <https://doi.org/10.1016/j.enfcli.2025.502214>
- Hussey, I., Alsalti, T., Bosco, F., Elson, M. y Arslan, R. (2023). An aberrant abundance of Cronbach's alpha values at .70. *Advances in Methods and Practices in Psychological Science*, 8(1). <https://doi.org/10.1177/25152459241287123>
- Karas, M. (2019). *English Language Teacher Self-Efficacy Beliefs* [Doctoral thesis, Western University]. Electronic Thesis and Dissertation Repository. <https://ir.lib.uwo.ca/etd/6091>

- Karas, M., Uchihara, T. y Faez, F. (2024). Language teacher self-efficacy surveys: What have we learned? Where are we going? En F. Faez y M. Wyatt (Eds.), *Building the Self-Efficacy Beliefs of English Language Learners and Teachers* (pp. 21-43). Routledge.
- Kerlinger, F. (2002). *Investigación del comportamiento* (4ª edición). McGraw-Hill.
- Klassen, R., Bong, M., Usher, E., Chong, W., Huan, V., Wong, I. y Georgiou, T. (2009). Exploring the validity of a teachers' self-efficacy scale in five countries. *Contemporary Educational Psychology*, 34(1), 67-76. <https://doi.org/10.1016/j.cedpsych.2008.08.001>
- Lambert, L. S. y Newman, D. A. (2023). Construct Development and Validation in Three Practical Steps: Recommendations for Reviewers, Editors and Authors. *Organizational Research Methods*, 26(4), 574-607. <https://doi.org/10.1177/10944281221115374>
- Lowes, N. (2019). What is a pilot study? *JOGNN*, 48(2), 117-118. <https://doi.org/10.1016/j.jogn.2019.01.005>
- Maldonado, N. y Santoyo, F. (2024). Validez de contenido por juicio de expertos: Integración cuantitativa y cualitativa en la construcción de instrumentos de medición. *REIRE Revista d'Innovació I Recerca En Educació*, 17(2), 1-19. <https://doi.org/10.1344/reire.46238>
- Malkewitz, C., Schwall, P., Meesters, C. y Hardt, J. (2023). Estimating reliability: A comparison of Cronbach's α , McDonald's ω and the greatest lower bound. *Social Sciences & Humanities Open*, 7(1). <https://doi.org/10.1016/j.ssaho.2022.100368>
- Martínez, J., Killian, M. y Del Ángel-Cortés, L. (2016). Estudio autónomo y estrategias metacognitivas: sus implicaciones para mejorar el aprendizaje del idioma inglés en educación básica. *Revista de Psicología y Ciencias del Comportamiento de la Unidad Académica de Ciencias Jurídicas y Sociales*. 7(1), 63-75. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S200718332016000100005&lng=es&tlng=es
- Medel, R. (2023). Creencias pedagógicas y práctica docente de maestros de inglés en escuelas primarias de educación básica: ¿practican lo que predicán? *Revista Educscientia. Divulgación de la Ciencia Educativa*, 6(12), 57-80. <https://educscientia.com/index.php/journal/article/view/345>
- Moradkhani, S. y Haghi, S. (2017). Context-based sources of EFL teachers' self-efficacy: Iranian public schools versus private institutes. *System*, 67, 259-259. <https://doi.org/10.1016/j.tate.2017.06.019>
- Moradkhani, S. y Haghi, S. (2019). The impact of university-based teacher education on EFL teachers' self-efficacy: The case of bachelor and master programmes in Iran. *The Language Learning Journal*, 50(1), 45-58. <https://doi.org/10.1080/09571736.2019.1688856>
- Mousalli, G. (2017). *Los instrumentos de evaluación en la investigación educativa*. Mérida.
- Murrieta, G. y Reyes, M. (2019). Experiencias de desempeño y creencias de autoeficacia en profesores de lenguas extranjeras. *Sinéctica*, 52. [https://doi.org/10.31391/s2007-7033\(2019\)0052-010](https://doi.org/10.31391/s2007-7033(2019)0052-010)
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2023). *Informe mundial sobre el personal docente: afrontar la escasez de docentes; puntos clave*. https://unesdoc.unesco.org/ark:/48223/pf0000387400_spa?posInSet=1&-queryId=45550825-69a2-41df-ada1-5be2f7fad9ee

- Pedrosa, I., Suárez-Álvarez, J. y García-Cueto, E. (2014). Evidencias sobre la validez de contenido: avances teóricos y métodos para su estimación. *Acción Psicológica*, 10(2), 3-20. <http://dx.doi.org/10.5944/ap.10.2.11820>
- Ramírez, P. (29 de noviembre de 2023). *La enseñanza del inglés en la educación básica en México: ¿Por qué nos cuesta tanto?* Universidad Iberoamericana. <https://faroeducativo.ibero.mx/2023/11/29/apunte-de-politica-n-42-la-ensenanza-del-ingles-en-la-educacion-basica-en-mexico-por-que-nos-cuesta-tanto/>
- Riggs, I. M. y Enochs, L. G. (1990). Toward the Development of an Elementary Teacher's Science Teaching Efficacy Belief Instrument. *Science Education*, 74(6), 625-637. <https://psycnet.apa.org/doi/10.1002/sce.3730740605>
- Rose, J. y Medway, F. (1981). Measurement of teachers' beliefs in their control over student outcome. *Journal of Educational Research*, 74(3), 185-190. <https://doi.org/10.1080/00220671.1981.10885308>
- Rotter, J. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80(1), 1-28. <https://psycnet.apa.org/doi/10.1037/h0092976>
- Sánchez, R. (2021). El tema de validez de contenido en la educación y la propuesta de Hernández-Nieto. *Latin-American Journal of Physics Education*, 15(3), 1-5. http://lajpe.org/sep21/15_3_09.pdf
- Sánchez, R., Decimavilla, D., Ortiz, M. y Torres, D. (2023). Las TIC y el desarrollo cognitivo de los estudiantes de tecnologías: una valoración desde la perspectiva del docente. *Prohominum*, 5(1), 77-94. <https://doi.org/10.47606/ACVEN/PH0172>
- Swanson, P. (2010a). Teacher efficacy and attrition: Helping students at introductory levels of language instruction appears critical. *Hispania*, 93(2), 305-321. <https://dx.doi.org/10.1353/hpn.2010.a382877>
- Swanson, P. (2010b). Efficacy and language teacher attrition: A case for mentorship beyond the classroom. *NECTFL Review*, 66, 48-72.
- Swanson, P. (2012). Second/foreign language teacher efficacy and its relationship to professional attrition. *The Canadian Modern Language Review*, 68(1), 78-101. <https://doi.org/10.3138/cmlr.68.1.078>
- Swanson, P. (2013). From teacher training through the first year on the job: Changes in foreign language teacher efficacy. *Electronic Journal of Foreign Language Teaching*, 10(1), 5-16. <https://e-flt.nus.edu.sg/wp-content/uploads/docroot/v10n12013/swanson.pdf>
- Swanson, P. (2014). The power of belief: Spanish teachers' sense of efficacy and student performance on the National Spanish Examinations. *Hispania*, 97(1), 5-20. <https://dx.doi.org/10.1353/hpn.2014.0015>
- Teresi, J., Yu, X., Stewart, A. y Hays, R. (2022). Guidelines for Designing and Evaluating Feasibility Pilot Studies. *Med Care*, 60(1), 95-103. <https://doi.org/10.1097/mlr.0000000001664>
- Thommen, D., Grob, U., Lauermann, F., Klassen, R. y Praetorius A. (2022). Different Levels of Context-Specificity of Teacher Self-Efficacy and Their Relations with Teaching Quality. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.857526>

- Trizano, I. y Alvarado, J. (2016). Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00769>
- Tschannen, M. y Woolfolk, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783-805. [https://doi.org/10.1016/S0742-051X\(01\)00036-1](https://doi.org/10.1016/S0742-051X(01)00036-1)
- Van Teijlingen, E. y Hundley, V. (2001). The Importance of Pilot Studies. *Social Research Update*, (35), 1-4. <http://sru.soc.surrey.ac.uk/SRU35.pdf>
- Vergara, J., Díaz, C., Tagle, T. y Ortiz, M. (2020). Adaptación de la Escala de Creencias de Autoeficacia Docente en estudiantes universitarios chilenos. *ACADEMO Revista De Investigación en Ciencias Sociales y Humanidades*, 7(2), 105-114. <https://revistaciencia.uamericana.edu.py/index.php/academo/article/view/329>
- Webb, R. y Ashton, P. (1986). Teacher Motivation and the Conditions of Teaching: A Call for Ecological Reform. *Journal of Thought*, 21(2), 43-60. <http://www.jstor.org/stable/42589178>
- Wyatt, M. (2014). Towards a re-conceptualization of teachers' self-efficacy beliefs: Tackling enduring problems with the quantitative research and moving on. *International Journal of Research and Method in Education*, 37(2), 166-189. <https://doi.org/10.1080/1743727X.2012.742050>
- Wyatt, M. (2018). Language Teachers' Self-efficacy Beliefs: A review of the Literature (2005-2016). *Australian Journal of Teacher Education*, 43(4), 92-120. <http://dx.doi.org/10.14221/ajte.2018v43n4.6>
- Wyatt, M., y Dikilitaş, K. (2021). English language teachers' self-efficacy beliefs for grammar instruction: Implications for teacher educators. *The Language Learning Journal*, 49(5), 541-553. <https://doi.org/10.1080/09571736.2019.1642943>
- Wyatt, M. y Faez, F. (2024). *Building the Self-Efficacy Beliefs of English Language Learners and Teachers* (1st ed.). Taylor Francis. <https://www.perlego.com/book/4318855/building-the-selfefficacy-beliefs-of-english-language-learners-and-teachers-new-perspectives-for-research-teaching-and-learning-pdf>
- Zaldívar, M., Morales, L. y Valdés, D. (2023). Adaptación lingüística y cultural de la escala de trastorno de ansiedad generalizado para pacientes con epilepsias en el contexto cubano. *Medisur*, 21(6), 1263-1271. http://scielo.sld.cu/scielo.php?pid=S1727897X2023000601263&script=sci_arttext&tlng=en
- Zhang, B. (2018). *English Language Teachers' Required Knowledge and Self-Efficacy Beliefs about Pronunciation Instruction* [Master's thesis, The University of Western Ontario]. Electronic Thesis and Dissertation Repository. <https://ir.lib.uwo.ca/etd/5290>

Anexo A

Guía sobre Consideraciones Éticas/Consentimiento Informado.

Título del Estudio: Creencias de Autoeficacia Docente en Futuros Profesores de Inglés

Instituciones Participantes: Escuela Normal Superior Plantel Hermosillo.

Propósito del Estudio: adaptación y validación de un instrumento para medir creencias de Autoeficacia en Futuros Profesores de Inglés.

Procedimientos: se tomó en cuenta a aquellos estudiantes inscritos en la materia de Prácticas y/o materias equivalentes a Prácticas Profesionales para su participación. Para proteger la identidad de los participantes, no será necesario anotar su nombre en la hoja de respuestas; se les asignará un folio dependiendo del orden en el que sean recibidas por el investigador responsable. Una vez firmado el consentimiento, se les entregará una hoja en la que responderán una serie de preguntas cuya resolución tomará 15 minutos aproximadamente. Los participantes tienen la libertad de anular su participación o abandonar el estudio en cualquier momento, sin la necesidad de explicar el motivo de su abandono. Se agradece a los participantes la mayor sinceridad posible al responder el cuestionario, puesto que no existen respuestas buenas o malas.

Riesgos y Molestias: el presente estudio se cataloga como de “riesgo mínimo” según el artículo 17 del Reglamento de la Ley General de Salud en Materia de Investigación para la Salud en México. Por una parte, no se realizará ningún tipo de intervención. Por otra parte, al firmar el consentimiento informado se autoriza al investigador a utilizar los resultados con fines exclusivamente científicos y conservando el anonimato de los participantes en todo momento.

Beneficios: es importante señalar que los beneficios obtenidos de este estudio tienen la intención de promover aspectos psicosociales importantes que influyen en la enseñanza y aprendizaje del inglés en México. No se compensará a los participantes con ningún estímulo que propicie su participación en el estudio, la intención es que participen bajo su propia voluntad aceptando y comprendiendo lo mencionado anteriormente. De

igual forma, los participantes están en todo derecho de solicitar los resultados de la investigación.

Datos Adicionales: la investigación es dirigida por el maestrante en Innovación Educativa (Universidad de Sonora) Eduardo Valle Velázquez, bajo la dirección de la Dra. María de los Ángeles Maytorena Noriega (Universidad de Sonora) en codirección con la Dra. Ma. del Rosario Reyes-Cruz (Universidad Autónoma de Quintana Roo). En caso de haber alguna duda respecto al estudio, favor de levantar la mano para ser atendida o acercarse al investigador en cualquier momento del estudio.

Contacto de Investigador: a223230146@unison.mx