

LA EVALUACIÓN DE EO Y EE EN ELE: RETOS Y PROPUESTAS

Coordinadores

Mtra. Marcela Nieto Rodríguez
Mtro. Edgar René Pacheco Martínez

María del Rosario Angulo Guerra • Klaudia Yvonne Buzo Hachec
Maribel Carmona Herrera • Sandra Luz Cruz Ituribarri
Jaime Osvaldo Islas Farías • Moisés López Olea
Ileana Betzabeth Lugo Martínez • Diana Helena Martínez Cruz
Jorge Salinas Hurtado



CEPE

CENTRO DE ENSEÑANZA
PARA EXTRANJEROS

LA EVALUACIÓN DE EXPRESIÓN ORAL Y EXPRESIÓN ESCRITA EN ESPAÑOL COMO LENGUA EXTRANJERA: RETOS Y PROPUESTAS

Proyecto PAPIME 407120

Coordinadores

Mtra. Marcela Nieto Rodríguez
Mtro. Edgar René Pacheco Martínez



CEPE

CENTRO DE ENSEÑANZA
PARA EXTRANJEROS

Lista de autores

Ileana Betzabeth Lugo Martínez
Diana Helena Martínez Cruz
María del Rosario Angulo Guerra
Jaime Osvaldo Islas Farías
Klaudia Yvonne Buzo Hachec
Moisés López Olea
Sandra Luz Cruz Iturribarri
Jorge Salinas Hurtado
Marcela Nieto Rodríguez
Maribel Carmona Herrera
Edgar René Pacheco Martínez



Colaboradores

Mariana Sofía Caleano Luna
César Jardines Suazo

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Dr. Enrique Luis Graue Wiechers

Rector

Dr. Leonardo Lomelí Vanegas

Secretario General

Dr. Alberto Ken Oyama Nakagawa

Secretario de Desarrollo Institucional

Dr. Francisco José Trigo Tavera

Coordinador de Relaciones y Asuntos Internacionales

CENTRO DE ENSEÑANZA PARA EXTRANJEROS

Dr. Domingo Alberto Vital Díaz

Director

Dra. Rosa Esther Delgadillo Macías

Secretaria General

Mtro. Luis Miguel Samperio Sánchez

Secretario Académico

Clasificación:

LCC PC4065

(Libro electrónico)

DDC

468.0071-dc23

Catalogación en la publicación UNAM. Dirección General de Bibliotecas

Nombres: Nieto Rodríguez, Marcela, editor, autor. | Pacheco Martínez, Edgar René, editor, autor. | Lugo Martínez, Ileana Betzabeth, autor. | Martínez Cruz, Diana Helena, autor. | Angulo Guerra, María del Rosario, autor. | Islas Fariás, Jaime Osvaldo, autor. | Buzo Hachec, Klaudia Yvonne, autor. | López Olea, Moisés, autor. | Cruz Iturrigarri, Sandra Luz, autor. | Salinas Hurtado, Jorge, autor. | Carmona Herrera, Marcela, autor. | Caleano Luna, Mariana Sofía, autor. | Jardines Suazo, César, autor.

Título: La evaluación de EO y EE en ELE: retos y propuestas / Marcela Nieto Rodríguez y Edgar René Pacheco Martínez, coordinadores.

Otros títulos: La evaluación de expresión oral y expresión escrita en ELE: retos y propuestas.

Descripción: Primera edición. | Ciudad de México: Universidad Nacional Autónoma de México, Centro de Enseñanza para Extranjeros, 2021.

Identificadores: LIBRUNAM 2103703 (libro electrónico) | ISBN 9786073043908 (libro electrónico).

Temas: Español -- Estudio y enseñanza -- Hablantes extranjeros. | Español -- Gramática -- Problemas, ejercicios, etc. | Español -- Escritura -- Problemas, ejercicios, etc.

Primera edición: mayo de 2021.

D.R. © 2021 Universidad Nacional Autónoma de México

Ciudad Universitaria, Alcaldía Coyoacán, C.P. 04510, Ciudad de México.

Centro de Enseñanza para Extranjeros

Avenida Universidad 3002, Alcaldía Coyoacán, C.P. 04510, Ciudad de México.

ISBN 978-607-30-4390-8

Responsable editorial: Guadalupe Rangel Esparza

Cuidado editorial: Brenda J. Vázquez Cantú

Corrección de estilo: Brenda Elizabeth Sosa Quintero, Pablo Aftab Gálvez Gutiérrez

Diseño editorial y formación: Susana Zavala Torres en Aguamiel Creativo

Esta edición y sus características son propiedad de la Universidad Nacional Autónoma de México.

Prohibida la reproducción total o parcial por cualquier medio sin la autorización escrita del titular de los derechos patrimoniales.

Hecho en México.

Introducción	6
Sección I. Introducción al diseño y construcción de instrumentos para la evaluación de ELE	12
Capítulo 1. Principios de la medición para construir instrumentos de evaluación del aprendizaje de ELE	13
Capítulo 2. Proceso metodológico para el diseño y construcción de instrumentos de evaluación de la EIE y la EIO	25
Sección II. La expresión oral	45
Capítulo 3. La habilidad oral y su evaluación	46
Capítulo 4. Lineamientos para la creación de instrumentos de la evaluación oral	65
Capítulo 5. Análisis de los actuales procedimientos e instrumentos del CEPE para la evaluación de la expresión oral	84
Sección III. La expresión escrita	107
Capítulo 6. La habilidad de expresión escrita y su evaluación	108
Capítulo 7. Lineamientos para la creación de instrumentos para la evaluación de la expresión escrita	118
Capítulo 8. Análisis de los actuales procedimientos e instrumentos del CEPE para la evaluación de la expresión escrita	128
Sección IV. Las tecnologías de aprendizaje y conocimiento (TAC) en la evaluación de EO y ELE	148
Capítulo 9. Breve historia de los desarrollos tecnológicos aplicados a ELE en el CEPE	149
Capítulo 10. Plataforma digital para la aplicación de exámenes de aprovechamiento y para la investigación en ELE: una propuesta tecnológica para la evaluación de las habilidades de expresión oral y escrita	159
Capítulo 11. Evaluación de las habilidades de expresión oral y escrita mediante las TAC	175
Glosario	188

INTRODUCCIÓN

Este libro es uno de los frutos del proyecto PAPIME con clave PE407120, titulado *Nuevos procedimientos para la evaluación de las habilidades de expresión escrita y expresión oral en estudiantes de español como lengua extranjera (LE) en el Centro de Enseñanza para Extranjeros (CEPE)*. Los autores que en él colaboramos tenemos como primer objetivo presentar la investigación que hemos llevado a cabo (a manera de antecedente teórico-práctico) para el desarrollo del producto final de este proyecto: una plataforma digital que aloje, a mediano plazo, propuestas de los exámenes de aprovechamiento, de expresión tanto oral como escrita; asimismo, planteamos propuestas de procedimientos e instrumentos para la evaluación de dichas habilidades.

En cuanto a los métodos implementados actualmente para ponderar las habilidades productivas con que trabajamos en el CEPE, tras un análisis cuidadoso, determinamos que dependen en gran medida de la subjetividad de los evaluadores. Esto, en términos rigurosos de la disciplina para la evaluación, reduce el grado de validez y confiabilidad de los resultados, pues favorece el margen de subjetividad y reduce el de objetividad. No es de extrañar que entre los colegas profesores, aun contando con rúbricas para evaluar las destrezas productivas, primen criterios muy variables, donde las perspectivas personales adquieren un peso considerable.

En relación con lo anterior, observamos que dentro de la subjetividad que implica evaluar la producción oral y escrita (y adelantándonos a un análisis más exhaustivo), posiblemente se omite considerar los objetivos de medición de cada una de las habilidades. Es decir, al examinar las deficiencias de esta metodología, reflexionamos en torno a los constructos de los instrumentos de evaluación con los que trabajamos. Asimismo, hemos notado que las creencias infundadas del profesorado repercuten en cómo se evalúa a ciertos grupos de estudiantes, con base en su origen étnico; por ejemplo, alumnos chinos, japoneses y coreanos, quienes a veces se considera que tardan más en desarrollar ciertas habilidades comunicativas en lengua española.

Por principio de cuentas, este alto grado de subjetividad en el modo de evaluar las expresiones oral y escrita, nos interesó profundamente, porque ocasiona que las calificaciones de un mismo estudiante resulten muy dispares, al ser diagnosticado por dos profesores; lo cual, sin duda incide de manera directa, incluso en la vida personal de los alumnos. Evidencia de esto es que, en ocasiones, el puntaje de la evaluación formativa que otorga el profesor del curso de lengua en el CEPE, puede diferir mucho del resultante de los exámenes de aprovechamiento, aplicados por otros profesores.

Debido a lo anterior, estimamos que se debe replantear la evaluación formativa en nuestros cursos, haciendo énfasis en la aplicación cotidiana y efectiva de procedimientos continuos, los cuales asistan tanto al profesor para brindar un seguimiento efectivo al aprendizaje de sus estudiantes, como a éstos mismos para que adquieran consciencia de sus propios avances en el dominio del español. No obstante, consideramos que también es imperativo atender la cuestión de las evaluaciones de aprovechamiento, mediante una propuesta de nuevos procedimientos, a través de los cuales se reduzca lo más posible el factor subjetivo y se estandaricen eficazmente los criterios de evaluación entre los miembros de la planta docente del CEPE.

A manera de antecedente

A partir de la década de los ochenta, se concibió la necesidad de evaluar el aprendizaje de una lengua extranjera desde otras perspectivas, además de únicamente la gramatical; es decir, contemplar la competencia lingüística comunicativa. Por ese entonces, también surgió la iniciativa de orientar la docencia hacia el uso de la lengua como vehículo de comunicación, por medio de las actuaciones orales y escritas en contextos de la vida real. El criterio que prevaleció después de que las destrezas productivas ocuparan un lugar central, fue el de que la expresión oral se dejara de evaluar de forma indirecta, a través de ítems escritos.

Si bien en el salón de clases ya se calificaban las tareas de interacción oral, era más frecuente que se hiciera una evaluación de la expresión, es decir, de las tareas de discurso monológico: con éstas se pretendía cubrir un discurso dialógico, o sólo responder las preguntas que formulaba el profesor, las cuales no se apegaban a las reglas básicas de la interacción en un contexto real. Por otro lado, también la producción escrita se evaluaba dentro del salón de clase, mediante tareas sencillas sobre actividades cotidianas, cartas o textos de exposición, que hacían énfasis sobre todo en lo concerniente a la gramática.

Con la premisa de justipreciar dichas actuaciones, tanto oral como escrita, proponemos, como parte de este proyecto, implementar las tareas incorporadas a la nueva plataforma digital, para que así se aseguren la validez y confiabilidad en la puntuación de dichas competencias. Es decir, que figuren como una muestra representativa de las posibles tareas que se espera que el estudiante sea capaz de producir. Además, creemos importante redactar tareas que eliciten del estudiante actuaciones en las que pueda mostrar todo lo que es capaz de hacer. Por tal motivo, creemos que la validez y confiabilidad de las pruebas de expresión deben estar concebidas en función de unos criterios de evaluación debidamente consensuados y socializados, que describan con claridad lo que se espera que el estudiante

realice frente al reactivo que representa una tarea determinada; asimismo, que las tareas de producción sean elaboradas con base en los documentos que aseguren la validez y confiabilidad de la evaluación.

La tecnología y la evaluación

Las nuevas tecnologías, sin duda alguna, han modificado la forma de leer y escribir de la gente del siglo XXI, lo cual afecta obvia y naturalmente el modo en que se aprenden las lenguas extranjeras. Así que, de acuerdo con los cambios tecnológicos que se viven hoy día, es necesario aplicar las actualizaciones pertinentes que demanda el campo de la evaluación, para su adaptación a las nuevas realidades virtuales y electrónicas. Del mismo modo, las recientes tecnologías han representado un gran empuje en lo que se refiere a los softwares disponibles y el desarrollo de programas para la elaboración y administración de pruebas. En este sentido, es importante destacar que la aplicación de exámenes en computadora, nos permitirá realizar análisis psicométricos de los resultados de forma más fina, lo que influirá en la mayor corrección y confiabilidad de dichos exámenes.

Como hemos señalado, el objetivo de este proyecto está orientado hacia la resolución de un problema muy puntual que enfrentamos los profesores de español del CEPE: actualmente, en los procedimientos para la evaluación de las habilidades, tanto de expresión escrita como oral, la subjetividad de los profesores –como evaluadores– posee un peso excesivo.

Debido a esto, hallamos conveniente realizar una evaluación exhaustiva de dichos procedimientos; identificar sus debilidades, fortalezas y áreas de posible mejora para, posteriormente, elaborar una propuesta de nuevos procedimientos de evaluación que procuren ser más objetivos y que permitan, tanto a los docentes como al Departamento de Evaluación, contar con mayores datos para el progreso continuo de nuestra enseñanza y del aprendizaje de nuestros alumnos.

Aunado a lo anterior, la plataforma es el principal producto del proyecto, ya que soportará las herramientas necesarias para los nuevos procedimientos de evaluación de las habilidades comunicativas de producción (escrita y oral): servirán como evidencias el texto escrito por los estudiantes y el video registrado con su actuación durante los exámenes orales. Además, se contará con rúbricas de evaluación para los profesores que calificarán dichas actuaciones; con herramientas para que los docentes registren el seguimiento del progreso de los estudiantes durante los cursos; y a mediano y largo plazo, la posibilidad de contar con herramientas de autoevaluación y coevaluación, para que los estudiantes mismos tengan la oportunidad de observar los avances de su propio aprendizaje, para así, estimular el aprendizaje autodirigido. Asimismo, consideramos que esta

herramienta puede servir como modelo para una futura plataforma que otras entidades de la UNAM implementen para la aplicación de exámenes a distancia.

Consideramos fundamental que los resultados de nuestras investigaciones sean difundidos, tanto entre colegas del CEPE como entre investigadores nacionales y extranjeros. Con esto, nuestra finalidad es que el proceso de enseñanza-aprendizaje del español en el CEPE mejore día con día; buscamos establecer un diálogo con nuestros profesores, así como con investigadores de otras latitudes, a través del cual se ponga sobre la mesa el tema de la evaluación del español: sus dificultades y sus retos, pero también las posibles soluciones y propuestas de mejora.

La estructura del libro

El hilo conductor de los capítulos que integran el presente volumen es la evaluación de las habilidades de expresión, oral y escrita, en los exámenes de aprovechamiento de ELE. Nuestros objetivos específicos son: introducir a los profesores de ELE en los conceptos fundamentales de la psicometría, ciencia de gran importancia para la disciplina de la evaluación; proporcionarles herramientas teórico-prácticas para reflexionar sobre cómo elaborar instrumentos de evaluación de expresión oral y escrita para sus clases; finalmente, presentar cómo las herramientas tecnológicas pueden integrarse en la evaluación de ELE, para optimizarla. Por tal razón, las secciones en que se divide el libro son cuatro, de las que se derivan sus respectivos capítulos específicos.

La primera sección se titula *Introducción al diseño y construcción de instrumentos para la evaluación de ELE*. Se compone del “Capítulo 1. Principios de la medición para construir instrumentos de evaluación del aprendizaje de ELE”, en el cual los autores abordan y definen el concepto de evaluación y el modo en que se vincula con el aprendizaje de ELE; asimismo, revisan los instrumentos de evaluación y los diferentes tipos de validez, concepto esencial para la evaluación en el ámbito docente. En el “Capítulo 2. Proceso metodológico para el diseño y construcción de instrumentos de evaluación de la expresión e interacción escrita (EIE) y la expresión e interacción oral (EIO)”, se describe el proceso metodológico que garantiza las buenas prácticas en la creación de tareas para la evaluación de las habilidades mencionadas; además, trata acerca de la elaboración de tablas de especificaciones, el análisis psicométrico y sus teorías, los puntos de corte y, finalmente, los lineamientos técnicos y las reglas para elaborar tareas eficaces en la evaluación de las expresiones oral y escrita.

La segunda sección se titula *La expresión oral*. Está integrado, en primer lugar, por el “Capítulo 3. La habilidad oral y su evaluación”; aquí, los autores se enfocan en la evaluación de la habilidad de expresión oral en el ámbito del aprovechamiento, es decir, al final de un curso de lengua. Con este propósito en mente, se definen los constructos pertinentes para ello: la habilidad oral, la expresión, la interacción y la conversación; además, atienden al tema de los criterios y los parámetros para diseñar una evaluación oral válida y confiable. En segundo lugar, está el “Capítulo 4. Lineamientos para la creación de instrumentos para la evaluación de la competencia oral”, cuyo objetivo es práctico: tal como su nombre lo indica, busca proporcionar una orientación cuya finalidad es elaborar herramientas para evaluar eficazmente la habilidad de expresión e interacción oral (EIO), principalmente su aprovechamiento. Y, por último, cierra esta sección el “Capítulo 5. Análisis de los actuales procedimientos e instrumentos del CEPE para la evaluación de la expresión oral”, donde se describen los actuales procedimientos para evaluar el aprovechamiento de EO, al interior del CEPE; se analizan, además, sus fortalezas y debilidades, presentando propuestas de mejora.

La tercera sección se enfoca en la expresión escrita. Está compuesta por el “Capítulo 6. La habilidad escrita y su evaluación”, cuyos autores se enfocan en exponer el modo en que se determina el aprovechamiento de la habilidad de expresión escrita; es decir, la que tiene como finalidad discernir hasta qué punto un estudiante ha alcanzado los objetivos específicos de aprendizaje de un curso de lengua, con base en los rasgos propios tanto de la escritura como de la intención, el contexto y el código compartido, así como los constructos de la escritura. El “Capítulo 7. Lineamientos para la creación de instrumentos de evaluación escrita”, cuyo objetivo es similar al del capítulo 4: brindar una guía que sirva en la elaboración de herramientas para la evaluación del aprovechamiento EE. Finalmente, se encuentra el “Capítulo 8. Análisis de los actuales procedimientos e instrumentos del CEPE para la evaluación de la expresión escrita”, donde se describen los actuales procedimientos que esta dependencia implementa para tal propósito, con el fin de analizar tanto sus fortalezas como sus debilidades, y presentar propuestas de mejora.

Por último, tenemos la cuarta sección, que lleva por título *Las tecnologías de aprendizaje y conocimiento (TAC) en la evaluación de EO y EE*, y está integrada por el “Capítulo 9. Breve historia de los desarrollos tecnológicos aplicados a ELE en el CEPE”, donde se presenta una revisión histórica de los diversos desarrollos tecnológicos realizados en nuestro centro, poniendo especial énfasis en la evaluación de las habilidades de expresión oral y escrita. Enseguida, el “Capítulo 10. Plataforma digital para la aplicación de exámenes de aprovechamiento y para la investigación en ELE: una propuesta tecnológica para la evaluación de las habilidades de

expresión oral y escrita” presenta las características y utilidades de la plataforma, producto final del proyecto PAPIME, así como los detalles de su administración y los beneficios que implicará su puesta en funcionamiento, tanto para la aplicación de exámenes, como para la recolección de datos con fines de investigación sobre ELE. Para cerrar el volumen, tenemos el “Capítulo 11. Evaluación de las habilidades de expresión oral y escrita mediante las TAC”, el cual tiene como objetivo exponer las características que adquiere la evaluación de las habilidades mencionadas a través de las tecnologías de aprendizaje y conocimiento (TAC); además, los retos que esto supone, y los pros y contras de dicho cambio.

Asimismo, el libro cuenta con un glosario de términos clave para la evaluación de la expresión oral y escrita, que consideramos será de gran utilidad para los docentes y profesores en formación que tengan interés en la disciplina de la evaluación de lenguas.

No queremos concluir esta introducción sin antes agradecer el apoyo del programa PAPIME, que facilitó los recursos para la realización de este proyecto. De igual modo, también queremos extender nuestra infinita gratitud a todos los colegas del Departamento de Evaluación del CEPE, quienes creyeron en esta iniciativa –¡gracias por creer y confiar en nosotros!– y aceptaron aportar con su experiencia y conocimientos para redactar los capítulos que aquí entregamos.

Finalmente, deseamos también hacer una mención especial a nuestros becarios del proyecto, Mariana Sofía Caleano Luna y César Jardines Suazo, quienes elaboraron de forma solícita, inteligente y tecnológicamente experta el glosario de conceptos clave. Este libro, sin duda, es un gran logro colectivo que nos impulsa a mantener la convicción de que sólo la unión hace la fuerza. De nuevo, gracias infinitas.

Marcela Nieto Rodríguez
Edgar René Pacheco Martínez

SECCIÓN I. INTRODUCCIÓN **AL DISEÑO Y CONSTRUCCIÓN DE** **INSTRUMENTOS PARA LA EVALUACIÓN DE ELE**

CAPÍTULO 1

PRINCIPIOS DE LA MEDICIÓN PARA CONSTRUIR INSTRUMENTOS DE EVALUACIÓN DEL APRENDIZAJE DE ELE

Ileana Betzabeth **Lugo Martínez**
María del Rosario **Angulo Guerra**
Diana Helena **Martínez Cruz**

1.1. La evaluación del aprendizaje de ELE

En cualquier proceso de enseñanza se requiere saber si los estudiantes están aprendiendo correctamente lo que se espera que aprendan. Por tal motivo, es necesario realizar una evaluación que permita identificar en qué medida el aprendiz alcanza o logra los objetivos planteados en un curso. En la adquisición de segundas lenguas (o lenguas extranjeras), la evaluación es un proceso que ha cobrado especial relevancia, ya que permite determinar el grado en el que los aprendientes van adquiriendo el dominio y nivel de competencia.

Según Bordón (2004), la evaluación dentro del currículum de la enseñanza de segundas lenguas dio sus primeros pasos en los años sesenta del siglo xx; a la fecha, ha ganado reconocimiento y un lugar importante dentro del campo de la lingüística aplicada, gracias al auge del modelo de la competencia comunicativa. Éste hace hincapié en la enseñanza del uso de la lengua y propone la evaluación de la misma a partir del desarrollo de tareas de examen, denominadas “tareas auténticas”. Cabe destacar que actualmente existe una gran preocupación por asegurar la validez y la confiabilidad de la evaluación. De ahí que se haya incorporado la psicometría, tal como se podrá ver en diversos instrumentos desarrollados en la última década, cuyo propósito es evaluar y certificar el dominio de la lengua española.

Si bien, como puede apreciarse, la evaluación es inherente al quehacer educativo en cualquier campo disciplinario; desafortunadamente, en el ámbito académico se le reduce a emitir una calificación numérica, o a la aplicación de un examen y, en el mejor de los casos, se asume como la valoración y regulación del proceso de aprendizaje. Por ello, cobra gran relevancia que quienes realicen una evaluación, tengan claro lo que significa *evaluar*.

Por ejemplo, Stufflebeam (1987) plantea que existen distintas definiciones del concepto de “evaluación”. Algunas de ellas varían sólo en aspectos menores, pero otras presentan diferencias significativas. Mientras que, por ejemplo, algunas afirman que la evaluación supone comparar objetivos y resultados, otras exigen una conceptualización más amplia. El *Joint*

Committee on Standards for Educational Evaluation pondera: “la evaluación es el enjuiciamiento sistemático de la valía o el mérito de un objeto”. Dicha acepción se centra en el término *valor* e implica que la evaluación siempre supone la emisión de juicios de esta índole.

Stufflebeam, estima que, en un sentido óptimo, la evaluación debe ser un proceso en el que se requiera recopilar información segura e imparcial; esto es, si un estudio no informa de cuán buena o mala es una cosa, no se trata de una evaluación apropiada de la misma. Así pues, este concepto puede ser entendido como un proceso que implica la obtención de información útil, la emisión de juicios de valor y la toma de decisiones al respecto.

Además de tener claro lo que implica y para qué sirve evaluar, también es fundamental saber que existen diversos tipos de evaluación. De acuerdo con Díaz Barriga (2002), en la evaluación de los aprendizajes se distinguen dos tipos de funciones: la pedagógica (que también se conoce como formativa) y la social (o sumativa). La función pedagógica hace referencia a la comprensión, regulación y mejora de la situación de enseñanza y aprendizaje, por lo que la evaluación sirve para saber lo que ocurre con las estrategias de enseñanza y con el aprendizaje de los alumnos; para que, en ambos casos, se realicen las mejoras y ajustes necesarios. En ese sentido, la función pedagógica se integra al proceso de enseñanza como una genuina *evaluación continua* dirigida a tomar decisiones, por lo que los objetos de la evaluación no son sólo los procesos de aprendizaje, sino también el proceso mismo de la enseñanza. Por su parte, la función social de la evaluación, se refiere a los usos que se le da a dicho procedimiento, más allá de la situación de enseñanza y aprendizaje y que tiene que ver con cuestiones tales como la selección, la promoción, la acreditación, la certificación, la información a otros, etc.

Desafortunadamente, si bien se reconoce la importancia y el valor de la evaluación pedagógica, ha tendido a prevalecer la función social, debido a la idea de que evaluar significa aplicar exámenes al final del proceso de instrucción, o que es calificar y asignar un número que certifique si se ha aprendido o no; o lo que es más crítico: se ha llegado a pensar que se está realizando una evaluación pedagógica, tan sólo porque se realizan a lo largo del proceso de enseñanza. Sin embargo, únicamente se emiten juicios y se le otorga una calificación al desempeño de los estudiantes, sin tomar acciones para mejorar su aprendizaje. Dichas calificaciones incluso constituyen prácticamente el único parámetro para decidir si el estudiante es promovido a un siguiente curso o no.

Según Gallardo *et. al.* (2015), en la actualidad, la evaluación ha pasado de ser una evaluación del aprendizaje a una evaluación *para* el aprendizaje, lo cual la coloca en otra dimensión del campo de la investigación: de considerarla en un principio como un medio para la emisión de juicios del aprendizaje a

posteriori, ha pasado a transformarse en un proceso entendido como activo y significativo, en tanto que permite que los estudiantes comprendan su proceso de aprendizaje *a priori*. En tal sentido, la evaluación sumativa tiene lugar al término de un determinado proceso de enseñanza-aprendizaje y está orientada a valorar si la instrucción ha sido satisfactoria en relación con los objetivos planteados, puesto que evalúa los productos que demuestran que los objetivos se han alcanzado. La evaluación formativa, por su parte, se realiza durante todo el proceso aludido, pero con la conciencia de que la información que se obtiene, debe forzosamente servir para reorientar el planteamiento inicial de la formación, cuando es necesario. Ambos tipos de evaluación requieren definir criterios válidos y confiables (conceptos que serán explicados más adelante), conforme a los objetivos de cada nivel de enseñanza. Estos serán aplicados por un equipo entrenado previamente para desarrollar una sensibilidad común que permita interpretar de modo similar dichas categorías cualitativas.

En el ámbito de la enseñanza de segundas lenguas, se espera que los aprendientes transiten de un estado de menor a otro de mayor conocimiento, para adquirir determinado nivel de competencia. De ahí que la evaluación, como queda dicho, haya cobrado un papel preponderante y, por tanto, que tanto la evaluación pedagógica como la social tengan la misma importancia. Como se aprecia en el modelo académico de los cursos de español y cultura mexicana del CEPE de la UNAM, el programa de estudios (2014) consigna la evaluación como el elemento integrador que relaciona y da sentido a todos los componentes del currículo (perfil, objetivos, contenidos y metodología).

Dado que este modelo tiene como base el enfoque comunicativo, la evaluación es considerada como un proceso continuo, que permite adecuar progresivamente los componentes programáticos, con base en las características particulares de cada grupo de estudiantes. El concepto de evaluación que se adopta, comprende dos puntos de vista complementarios y no excluyentes: por una parte, los profesores emiten juicios sobre el grado en que los estudiantes han alcanzado los objetivos previamente establecidos (evaluación sumativa o de productos); de manera específica en el CEPE, esta evaluación se realiza principalmente mediante exámenes de aprovechamiento que miden las cuatro habilidades de la lengua por separado, además de la competencia lingüística y otros elementos que se consideran en la calificación final de los alumnos. Por otro lado, el segundo punto de vista alude a los procedimientos adicionales que posibilitan el análisis de los juicios emitidos, así como la retroalimentación y la toma de decisiones (evaluación formativa o de proceso). En la práctica, dentro del CEPE, este tipo de evaluación corresponde a las actividades que los profesores realizan en el salón de clase, mediante las cuales se evalúa el progreso de cada estudiante.

1.2 Instrumentos de evaluación

Como se mencionó en el apartado anterior, para lograr una evaluación consistente, es imprescindible obtener información objetiva, válida y confiable con el fin de emitir juicios y tomar ciertas decisiones con respecto a lo que se está evaluando; para esto, se requiere contar con los instrumentos pertinentes. En consecuencia, quienes realicen la evaluación deben tener claro qué son los instrumentos de evaluación; los diversos tipos que existen y que se pueden emplear o diseñar, en función del objeto de la evaluación; así como los criterios técnicos que deben cumplir para que sus resultados sean confiables.

Los instrumentos de evaluación, de acuerdo con el Centro Nacional de Evaluación para la Educación Superior, A.C. (CENEVAL, 2000), son herramientas que sirven para medir el nivel de ejecución de un individuo o de un grupo, en un dominio o constructo dado (lo que mide o deberá medir); es decir, la cantidad de conocimientos, habilidades, destrezas, capacidades, etc. que posee un individuo. En ese sentido, es posible afirmar que un instrumento de evaluación es un procedimiento estandarizado para tomar una muestra de conducta y describirla con categorías o puntuaciones. Vale la pena mencionar que, los instrumentos son sumamente variados en sus formatos y aplicaciones.

En el ámbito educativo, de acuerdo con Quesada y Gallego (2013), los instrumentos que más se utilizan en la evaluación del aprendizaje son listas de control, escalas de valoración y rúbricas, siendo estas últimas el elemento más complejo y el que se asocia con un mayor potencial para el aprendizaje de los alumnos. Por su parte, Lugo (2007), refiere que los instrumentos más utilizados en la evaluación psicológica son las pruebas objetivas, por sus características de construcción y aplicación. La entrevista que se desarrolla cara a cara entre dos personas, implica mucho más que una charla: constituye el portafolio de evidencias que en las últimas décadas ha tenido un amplio uso en muchos campos, incluyendo el de la educación, por supuesto.

Sin importar el tipo de instrumento que se utilice, cualquiera se puede definir con respecto a su calidad técnica o psicométrica. En este punto, basta decir que un instrumento mide lo que pretende medir en forma consistente, y que si dos pruebas pretenden medir el mismo constructo exacto (definido en forma idéntica), aquel que posea mayor precisión, será considerado el mejor instrumento. Es decir, el técnicamente superior, o más sólido desde el punto de vista de la psicometría, como se explicará más adelante.

En ese orden de ideas, según Flotts (s.f.), un instrumento de evaluación puede ser entendido como un medio para obtener información, evidencia o muestras sobre el objeto que se desea evaluar. Por esta razón, un

instrumento debe ser cuidadosamente elaborado y ceñirse a tres criterios esenciales: validez, confiabilidad y objetividad.

En la enseñanza de segundas lenguas, como queda dicho, dado que la evaluación ha cobrado un papel trascendente, es necesario utilizar diversos instrumentos de evaluación del aprendizaje y para el aprendizaje. Tal es el caso del CEPE, donde se han diseñado y se aplican diversos instrumentos de esta índole; ya sea para la colocación de estudiantes en los cursos de español para extranjeros, ya para registrar el nivel de aprovechamiento al término de los cursos, o bien para la certificación del dominio de la lengua.

Es subrayable que los instrumentos de evaluación deben cumplir con determinados criterios para ser considerados *de calidad* y, por lo tanto, poder confiar en sus resultados. Por ello, es relevante que quienes los utilicen o los vayan a diseñar cuenten con las nociones básicas al respecto. A continuación, se presentan algunos conceptos básicos y sustantivos para la construcción de instrumentos de evaluación, así como sus implicaciones.

Uno de los criterios técnicos fundamentales es la objetividad, que junto con la validez y la confiabilidad son cualidades necesarias para una adecuada evaluación. Desafortunada y generalmente, la mayoría de las veces, la preocupación por construir un instrumento de calidad se centra en garantizar la confiabilidad, mientras que se relega a un segundo plano la validez, y ni siquiera se considera la objetividad como un atributo indispensable.

La objetividad, según Frade (2013), es el atributo que define en qué medida los instrumentos de evaluación permiten identificar en términos reales lo que existe, lo que en verdad es, prescindiendo de criterios subjetivos, inherentes a juicios personales. Es decir, la objetividad está presente en tanto que se reducen al mínimo los factores que puedan afectar la medición, la cual a menudo recae en opiniones individuales, tales como el esfuerzo del alumno para desarrollar una actividad o una tarea.

Del igual suerte, Tristan y Pedraza (2017) explican el valor que tiene la objetividad para el desarrollo de instrumentos de evaluación, al emplearse como herramienta de vigilancia que garantiza la neutralidad en los estímulos presentados a los sustentantes. Para estos autores, la objetividad tiene cinco propiedades: especificidad, neutralidad, independencia, imparcialidad e impersonalidad. Cada una resulta fundamental para interpretar los resultados, eliminar o reducir los sesgos inducidos por la influencia de estereotipos y preferencias en el diseño del instrumento o en la apreciación de jueces, entre otros factores que pudieran afectar el uso ético de los resultados de los instrumentos. Con ello, se muestra que la objetividad es el primer atributo que debe definirse en un instrumento estandarizado, distinguiendo las cualidades que le son propias, para evitar asociarlas incorrectamente con la validez o la confiabilidad.

En la evaluación de lenguas (y específicamente en la evaluación de las habilidades de expresión escrita y oral, como se verá más adelante), se requiere de la participación de expertos en la lengua, que juzguen la calidad de las producciones de los examinados, por lo que la subjetividad se hará presente y será necesario reducirla o controlarla.

El segundo criterio necesario a considerar en todo instrumento de evaluación es la validez, que suele ser compleja de definir; sin embargo, en la historia de la psicología el término "validez", según Ventura-León (2016), fue introducido por la APA y la mayoría de las definiciones coinciden en que un instrumento es válido cuando mide lo que se pretende medir.

De acuerdo con Nunnally (1987), luego de seleccionar el modelo para elaborar los instrumentos de evaluación y después de haberlos construido, el paso siguiente es demostrar su utilidad, es decir, determinar la validez de dichos instrumentos. El término validez tiene algunas connotaciones erróneas, pero está lo suficientemente enraizado en la bibliografía como para establecer otros términos que denoten de modo más adecuado el proceso en cuestión; en un sentido muy general, se puede plantear que un instrumento de medición es válido si cumple satisfactoriamente el propósito para el que fue diseñado.

Esto se traduce en que un índice de validez indica el grado en que una prueba mide lo que se propone medir cuando se le compara con un criterio aceptado. Así, la selección de los criterios de validez satisfactorios y la demostración de un grado apropiado de validez son los aspectos fundamentales en la evaluación psicológica y educativa. La primera condición necesaria para que una prueba sea válida es que posea un grado adecuado de confiabilidad. La validez refleja un juicio en evolución, basado en la investigación, acerca de qué tan adecuadamente una prueba es capaz de medir aquel atributo para el cual se diseñó.

La definición de validez aportada por Gregory (2001), indica que un instrumento es válido en la medida en que las inferencias que se realicen a partir de éste, son apropiadas, significativas y útiles. En consecuencia, este atributo no se puede captar únicamente a través de resúmenes estadísticos. Generalmente, las diferentes maneras de acumular evidencia de la validez de un instrumento, se han agrupado en tres categorías: validez de contenido, validez relacionada con el criterio y validez de constructo. Cabe mencionar que el uso de estas etiquetas, no implica que existan tipos precisos de validez, o que un procedimiento específico de validación sea mejor que otro.

Por su parte, Oliden (2003) considera que la validez es esencial en la medición psicológica y que ésta se relaciona de alguna manera con el sesgo, porque puede alterar los resultados. También menciona que la validez se

asocia con la investigación del significado teórico de las puntuaciones que se obtienen a través de los instrumentos, lo cual demuestra hasta qué punto están relacionadas las respuestas observables de los sustentantes, con las interferencias en función de los resultados del instrumento.

A la luz de lo anterior, surge la pregunta de cómo puede garantizarse la validez de un instrumento o, en otras palabras, cómo se puede demostrar que realmente se está midiendo lo que se pretende medir. Para responder esta pregunta, a continuación se explicarán y ejemplificarán los diferentes tipos de validez.

a) Validez de contenido

Este tipo de validez, de acuerdo con Gregory (2001), se determina a partir del grado en que las preguntas, tareas o reactivos del instrumento constituyen una muestra representativa del universo de conductas para el cual es diseñado. En ese sentido, los estímulos que se construyen para un determinado instrumento pueden considerarse como una muestra que pertenece a una población más amplia de estímulos potenciales, los cuales a su vez definen aquello que realmente se desea medir. Expresado en otros términos, si la muestra de reactivos es representativa, entonces, el instrumento posee validez de contenido. Dicha validez es un concepto muy útil cuando se sabe mucho acerca de la variable que se desea medir, y es más fácil de garantizar cuando la prueba mide el rasgo definido de modo aproximado.

b) Validez relacionada con el criterio

Este tipo de validez, nuevamente de acuerdo con Gregory (2001), se comprueba cuando es posible afirmar que un instrumento es eficaz para estimar el desempeño de los examinados, a través de alguna medida de resultado. En este caso, la variable de interés principal es la medida de resultado, conocida como *criterio*. En síntesis, la puntuación de una prueba es útil cuando proporciona información suficiente para la predicción precisa del criterio.

Cabe destacar que existen dos fuentes que pertenecen a la validez relacionada con el criterio: la validez concurrente y la validez predictiva. La primera, es observable cuando se aplican dos instrumentos que miden el mismo constructo respecto a un mismo sujeto y se obtienen resultados idénticos o muy similares. Por ejemplo: en la evaluación para la certificación del dominio de la lengua española, como se sabe, existen diferentes instrumentos que miden los niveles de desempeño establecidos en el MCER; por lo tanto, existe evidencia de validez concurrente, si un mismo candidato se evalúa con dos exámenes diferentes y obtiene el mismo resultado. Por su parte, la validez predictiva se obtiene tiempo después de que se aplicó el instrumento, como en el caso de los resultados de un examen de colocación y el desempeño de los estudiantes en un curso de español.

Dentro de la evaluación de lenguas, en un examen que se utilice para colocar a un estudiante en un curso específico, tal como se acaba de explicar, es posible comprobar la validez de criterio cuando se observa que el desempeño del estudiante en dicho curso, efectivamente corresponde con el resultado que obtuvo en el examen de la colocación. También, se puede verificar la validez de criterio, cuando hay concordancia entre el resultado que alcanzó el alumno en el examen aplicado al término de un curso y su desempeño en el siguiente nivel.

c) Validez de constructo

Para explicar este tipo de validez, según Gregory (2001), hay que decir que un constructo es una cualidad o rasgo teórico, intangible, en el cual difieren individuos. Algunos ejemplos de constructos son capacidad de liderazgo, hostilidad sobrecontrolada, depresión e inteligencia. Obsérvese que en cada uno de estos ejemplos, los constructos se infieren a partir de la conducta, pero son mucho más que sólo ésta. En general, a un nivel teórico se considera que los constructos poseen cierta forma de existencia y ejercen influencias amplias, pero en cierto grado son posibles de predecir. Un instrumento que se diseñe para medir un constructo, debe estimar la existencia de una característica inferida subyacente (p. ej., capacidad de liderazgo), con base en una muestra de conducta.

Ahora bien, según Bunge (1976), el término *constructo* deriva de la dicotomía cosa-constructo y se entiende que a diferencia de las cosas, los constructos sólo comprenden los aspectos que no pueden tocarse. Con base en esta primera aportación, se propone que a pesar de que el constructo es algo intangible, puede definirse para hacerlo observable y cuantificable; por eso este concepto es fundamental para tomar decisiones en la evaluación. Por su parte, Thorndike (1989) define este concepto, a partir de la psicología, como algo que no puede ser observable, pero que está relacionado con la conducta humana; de igual manera, dicho atributo sirve para explicar y teorizar las muestras de conducta, con la finalidad de poder entenderlas.

En un sentido más amplio y específico, Arias (2014) expone que un constructo puede definirse como el concepto o característica que se pretende medir con el instrumento. El constructo en términos de un instrumento es parte medular, porque a través de éste, se define qué conductas desean observar. Para ponerlo en términos operacionales, debe concebirse en relación con otros constructos; de esta manera, será más fácil delimitar qué es lo que se busca medir. Un ejemplo sería la evaluación de gramática: esta área representa un constructo inobservable que requiere ser medido, para lo cual se requiere de evidencia que permita comprobar el grado en que un estudiante posee este conocimiento. Dicha medición, claro está, no podría llevarse a cabo sin antes definir de manera precisa, qué se entiende por gramática, desde una perspectiva teórica y su posterior definición operacional.

Para proponer una definición de validez empírica en este capítulo, es preciso aclarar que se incluyen tanto la validez criterial como la predictiva. De acuerdo con Thorndike (1989), la validez predictiva se centra en observar las correlaciones entre la medida denominada como criterio de éxito y su puntuación. De tal forma, tanto la validez criterial como la predictiva y la empírica tienen evidencia en los hechos.

La validez empírica, a diferencia de los tipos anteriores mencionados (contenido y constructo), no está relacionada con la teoría del constructo que se desea medir; por el contrario, se refiere a confirmar mediante el instrumento que la habilidad de la persona es real, o que es factible observarla a través de hechos verificables en su desempeño. En este tipo de validez, se puede comprobar cuando se presenta una correlación entre los criterios alcanzados en una prueba y el desempeño de los examinados en situaciones reales; esto nos confirma que existe una mayor efectividad en el instrumento, hay validez si los resultados en el examen (ya sean buenos o malos) se correlacionan con el desempeño en la vida real, en el trabajo o en otros exámenes.

El último criterio esencial que debe cumplir todo instrumento de evaluación es la confiabilidad. Para la *American Educational Research Association*, *American Psychological Association* (APA) y la *National Council on Measurement in Education* (2014), este atributo se refiere a la consistencia entre las puntuaciones del instrumento, aplicado a los mismos individuos en diversas ocasiones. Al respecto, se puede añadir que es la exactitud en la medición o la ausencia de errores en las puntuaciones del instrumento. Esto da lugar a que los resultados sean constantes a pesar de que las circunstancias cambien. En caso de que haya alguna falta de consistencia, es necesario diferenciar qué tipo de cambios intervienen en el rasgo que se pretende medir, puesto que algunos representan un cambio real, mientras que otros influyen en las puntuaciones, debido a circunstancias personales que pueden afectar la evaluación; por ejemplo, el nivel de estrés de los examinados, factores ambientales, entre muchos otros.

En consonancia con Tornimbeni (2008), el atributo medido puede tener cambios reales sin que se ponga en duda la confiabilidad; por el contrario, todo cambio provocado por factores externos puede afectar la confiabilidad de las puntuaciones. En la confiabilidad se integran tres dimensiones: la estabilidad, la consistencia interna y la confiabilidad inter-examinadores, con las que se puede relacionar un tipo distinto de error. El primer factor está relacionado con la puntuación libre de errores de medición, causados por motivos personales. El segundo, se refiere al grado en que un reactivo mide el mismo constructo y, finalmente, el tercero alude a que la medición debe ser independiente a la subjetividad del evaluador. En resumen, la confiabilidad se refiere a la consistencia en la medición; para su verificación, se desarrollaron métodos analíticos que serán explicados a detalle en el apartado de análisis psicométrico.

Como se ha podido ver a lo largo de este primer capítulo, la objetividad, la validez y la confiabilidad son criterios que se consideran necesarios para asegurar la calidad de un instrumento de evaluación. Es importante señalar que un instrumento puede contar con un alto grado de confiabilidad y, sin embargo, no ser válido; es decir, un instrumento puede medir de forma confiable, y no estar midiendo lo que se pretende medir. Por ejemplo, si quisiera medirse la temperatura corporal y para tal propósito se utiliza un flexómetro, siempre se obtendrá una medida; no obstante, está claro que el flexómetro nunca aportará información acerca de los grados de temperatura que presente una persona.

Referencias

American Educational Research Association, American Psychological Association (APA), National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC.

Arias, M., Lloreda, M. V., y Lloreda, M. J. (2014). *Psicometría*. Alianza Editorial.

Bordón, T. (2008). Panorama histórico del desarrollo de algunas de las cuestiones fundamentales en la evaluación de segundas lenguas. En *marcoELE. Revista de Didáctica Español Lengua Extranjera*, (7), 4-25.

Bunge, M. (1976). *La investigación científica: su estrategia y su filosofía*. Barcelona: Ariel.

CENEVAL (2000). *Estándares de calidad para instrumentos de evaluación educativa*, México: INEE.

CEPE (2014). *Programa de Estudios de los cursos de Español y Cultura Mexicana*. México: UNAM.

Díaz Barriga, F. y Hernández, G. (2002). *Estrategias docentes para un aprendizaje significativo* (Vol. 2). México: McGraw-Hill.

Flotts, M, Rodríguez, M. (s.f.). *Cuadernillo técnico de evaluación educativa. Nociones básicas en medición y evaluación en el contexto educativo*. INEE. [Recuperado de: <https://www.inee.edu.mx/wp-content/uploads/2019/08/P2A251.pdf>]

Frade, L. (2013). *La evaluación por competencias: Tomando en cuenta los últimos planes y programas y las modificaciones realizadas en la evaluación*. México: Inteligencia Educativa.

Gallardo, K., Valdés, D., y Álvarez, N. (2015). Las prácticas de evaluación del aprendizaje en relación con los estándares internacionales: un estudio exploratorio. En *Innovación educativa*. 15 (68). 117-133.

Gregory, R., y Ruiz, V. (2001). *Evaluación psicológica: historia, principios y aplicaciones*. México: Editorial El Manual Moderno.

Lugo, I. y Cuellar, Y. (2007). *Construcción de una prueba general de conocimiento de evaluación de competencia laboral* (Tesis de pregrado). México: UNAM.

Nunnally, J. (1987). *Teoría psicométrica*, México: Trillas.

Oliden, P. (2003). Sobre la validez de los tests. En *Psicothema*, 15 (2), 315-321.

Quesada, V y Gallego, B. (2013). *Los instrumentos de evaluación*. [Recuperado de https://nuestroscursos.net/pluginfile.php/5172/mod_resource/content/2/UF4_Los_Instrumentos_de_Evaluacion.pdf]

Samperio, L. (2004). *Determinación del constructo*. Manuscrito no publicado, México: CENEVAL.

Stufflebeam, D., y Shinkfield, A. J. (1987). *Evaluación sistemática: guía teórica y práctica*. Barcelona: Paidós.

Thorndike, R., Hagen, E. y González, E. (1989). *Medición y evaluación en psicología y educación*. México: Trillas.

Tornimbeni, S., Pérez, E., Olaz, F., de Kohan, N., Fernández, A., y Cupani, M. (2008). *Introducción a la psicometría*. Buenos Aires: Paidós.

Tristán, A. y Pedraza, N. (2017). La Objetividad en las Pruebas Estandarizadas. En *Revista Iberoamericana De Evaluación Educativa*, 10 (1).

Ventura-León, J. (2016). Breve historia del concepto validez en psicometría. En *Revista Peruana de Historia de la Psicología*, 2 (89).

Bibliografía recomendada

Abarzúa, A. (s.f). *Cuadernillo técnico de evaluación educativa Confiabilidad, validez e imparcialidad en evaluación educativa*. INEE.

Alderson, J. C., Clapham, C., y Wall, D. (1998). *Exámenes de idiomas: Elaboración y evaluación*. Cambridge University Press.

CAPÍTULO 2

PROCESO METODOLÓGICO PARA EL DISEÑO Y CONSTRUCCIÓN DE INSTRUMENTOS DE EVALUACIÓN DE LA EIE Y LA EIO

Ileana Betzabeth **Lugo Martínez**

Jaime Osvaldo **Islas Farías**

1. Etapas para el diseño y construcción de instrumentos de evaluación

Si bien es cierto que por mucho tiempo los instrumentos de evaluación han sido duramente criticados, durante los últimos años su uso en el contexto educativo, y particularmente en la enseñanza de lenguas, se ha incrementado notablemente. Por ello, se hace especialmente necesario ampliar el conocimiento y despertar la consciencia de quienes los desarrollan y aplican, con el fin de que sean lo más objetivos, válidos y confiables. En ese sentido, a continuación se presentan los procedimientos y pasos deseables que se recomienda seguir para diseñar, elaborar y validar un instrumento de evaluación, para así asegurar que éste sea de calidad y de conformidad con lo que establece la teoría psicométrica, expresada en diversos estándares nacionales e internacionales.

1.1. Instrumentos de evaluación para el aprendizaje de ELE

Para comenzar, es importante destacar que un instrumento de evaluación es una herramienta de medición que permite obtener información para tomar decisiones. Por ello, es necesario conocer qué tan bueno es el instrumento para cumplir con el propósito que se plantea. Es decir, se requiere que quienes intervienen en la confección del instrumento apliquen los principios básicos de la teoría psicométrica.

La creación de un instrumento de evaluación psicológica o educativa es un proceso complejo. Cualquier herramienta de este tipo debe caracterizarse por su valor técnico, esto es, que ha de considerar los principales criterios psicométricos, como la objetividad, la validez y la confiabilidad (conceptos que ya han sido explicados en el capítulo anterior), de tal forma que se asegure la veracidad de los resultados. Además de prestar especial atención a sus características técnicas, se deben considerar las necesidades y derechos de las personas que son evaluadas; así como tener en cuenta a la sociedad que confía en la información obtenida de dicha aplicación. En la evaluación de una lengua, el desarrollo de los instrumentos no puede ni debe ser diferente a cualquier otro campo disciplinar u objeto de medida.

Para Aiken (2003), una de las mayores preocupaciones que expresan constantemente los profesionales de la educación, es que los instrumentos de evaluación psicológica y pedagógica midan efectivamente lo que sus autores sostienen que miden. Por tal motivo, la Asociación Americana de Psicología (APA) y el Consejo Sobre Medición en Educación (NCME) desarrollaron y publicaron un conjunto de estándares y normas, basado en los principios de la psicometría, en el que recomiendan y abordan a detalle, criterios para diseñar, elaborar, aplicar e incluso evaluar a su vez los propios instrumentos de evaluación y los efectos de su uso. Dichos estándares han sido la base del trabajo realizado por diversos organismos especializados, como la ALTE (en Europa) y el SICELE (a nivel internacional) para la certificación de lenguas; o el CENEVAL en México, que se dedica al diseño y aplicación de instrumentos de evaluación a gran escala para la educación media superior y superior del país.

Aunque en la literatura especializada no existe un consenso respecto a las etapas que se deben seguir para diseñar y construir un instrumento de evaluación, es claro que se deben seguir determinados pasos para asegurar que éste sea objetivo, válido y confiable. De acuerdo con el CENEVAL (2000), el proceso de construcción se constituye por tres grandes etapas secuenciadas e interactivas entre sí: el diseño, la elaboración y la validación del instrumento. En cada etapa se debe seguir una serie de procedimientos y pasos, cuya ejecución asegure la validez y la confiabilidad. Es importante aclarar que si bien los procedimientos pueden variar en función del tipo de instrumento que se desee construir, siempre será necesario aplicar un conjunto de principios técnicos y metodológicos que les asegure a los usuarios que el instrumento medirá el constructo deseado.

Por su parte, Brown (2000) refiere que los procedimientos que se empleen para construir un instrumento de evaluación, deben tener como propósito asegurar que éste logre su finalidad; lo cual sólo se consigue al aplicar determinados principios de la medición. Afirma que, independientemente de las etapas que se sigan (aun cuando algunas se realicen en forma paralela), se debe incluir la especificación de la finalidad del instrumento, la construcción y la comprobación de los reactivos, así como el ensamble de la versión definitiva, con el análisis y el estándar de la calificación y los resultados.

Desde el punto de vista de Tristán y Vidal (2006), cuando se diseña un instrumento de evaluación educativa, es imperativo contar con estándares de calidad que puedan servir como referencia y brindar un sustento teórico razonable, para garantizar así la validez, la confiabilidad y otros atributos psicométricos que brinden certeza de que se dispone de un instrumento óptimo.

Concretamente, en la evaluación de lenguas, la ALTE (2011) señala que el proceso de desarrollo de un instrumento consta de tres fases esenciales:

planeación, diseño y validación; si bien considera una cuarta fase, que contempla informar a los tomadores de decisiones precisiones que pueden ser necesarias en algunos contextos. Esto último, porque la difusión de los resultados no contribuye a la producción de las especificaciones, como otras fases, pero tiene por objetivo informar a otros acerca del nuevo instrumento.

Como se habrá podido apreciar, si bien no existe una sola metodología, los diversos autores que se han citado, coinciden en que: se debe planear, construir y validar el instrumento. A la luz de lo anterior, a continuación se presenta una propuesta metodológica que consta de tres etapas: diseño, elaboración y validación. De modo general, se explicará en qué consiste cada una y, de manera específica, se hará referencia al desarrollo de instrumentos para evaluar el aprovechamiento de la expresión e interacción escritas (EIE) y la expresión e interacción orales (EIO), en el aprendizaje de español como lengua extranjera (ELE).

1.1.1. Diseño del instrumento

a) Propósitos del instrumento

Cuando se construye un instrumento, se requiere planear su diseño y definir cuál es su propósito y qué uso se le dará a la información obtenida de su aplicación. El diseño debe surgir de la propuesta de un grupo de expertos en el contenido de la evaluación; este grupo, al que generalmente se le conoce como Comité Técnico (CT), es el encargado de definir el propósito de la evaluación, el cual debe plantearse a partir de la información que espera obtenerse y de la utilidad que se le dará. El propósito debe redactarse, obligadamente, de manera que contenga un *qué*, un *cómo* y un *para qué*.

b) Definición teórica o de constructo

Una vez definido el propósito, el CT debe determinar los marcos teóricos o conceptuales que servirán como referentes, para decidir el significado teórico o constitutivo particular que se dará a éstos con fines de medición; es decir, el constructo que se medirá con el instrumento. Estos marcos constituyen las fuentes de validez teórica del mismo, las cuales deben quedar debidamente documentadas.

Hay que tener en cuenta que el constructo es un concepto difícil de definir, debido a que se trata de algo intangible, que aun así se sabe que existe en alguna cantidad, pero cuya definición necesita ser discutida e inferida, al no haber un consenso entre un grupo acotado de individuos, aunque compartan el mismo campo disciplinario; por ejemplo, la inteligencia. Por esta razón, cuando se diseña un instrumento para medir un determinado constructo, se debe estimar la existencia de algunas características subyacentes que se pondrán en evidencia con base en una muestra de

conducta, ya que a través de un proceso de categorización se convierte en una variable que puede ser observada y, por lo tanto, medida.

c) Definición operacional

Teniendo claro el propósito del instrumento y cuando el CT cuente con los marcos conceptuales, lo siguiente es determinar la definición operacional de lo expresado en la definición teórica del constructo que se pretende medir. Esto consiste en la descripción, en términos mensurables, de cada uno de los elementos de carácter cognitivo identificados en la definición teórica. La definición operacional obliga a que las personas encargadas de la elaboración de los reactivos o tareas, tengan la menor discrecionalidad para interpretar lo que se quiere evaluar, así como la mayor claridad en cuanto a lo que se desea preguntar y el tipo de reactivos y tareas a elaborarse. Por tanto, la definición operacional se debe plasmar en la tabla de especificaciones, que es una de las partes fundamentales del diseño del instrumento, la cual delimita el alcance de lo que se va a medir. En otras palabras, este concepto se traduce como la expresión, en términos mensurables, del objeto de medida.

d) Validación teórica o de contenido

La tabla de especificaciones debe ser validada por un panel de expertos distinto al que la elabora. La validación de contenido se basa principalmente en las opiniones de éstos; así pues, tal proceso implica el desarrollo de aquellas acciones que garanticen que los componentes de la tabla son pertinentes y claros: que están ubicados adecuadamente, que han sido desagregados correctamente y que no falta o sobra ninguna parte del constructo. En ese sentido, la validación de la tabla es un proceso de aseguramiento de la eficacia del contenido y, en su momento, del constructo del instrumento. Una vez concluida la validación, los resultados de esta tarea deben presentarse al equipo que se encarga de elaborar la tabla de especificaciones; este grupo juzgará la pertinencia de las propuestas y realizará los ajustes que considere necesarios. Por lo tanto, la etapa de diseño del instrumento culmina con los ajustes de la tabla de especificaciones, una vez que es validada.

Haciendo una recapitulación de los pasos que se acaban de presentar en cuanto al diseño del instrumento (específicamente en la evaluación del aprovechamiento de la EIE y la EIO), se puede decir que el propósito es evidente, pues se requiere contar con un instrumento que permita conocer el nivel de aprovechamiento de los estudiantes, tras haber tomado un curso de español, para poder promoverlos de nivel. En este caso, los diseñadores del instrumento deben consultar el programa de estudios, los libros de los cursos y aquellos referentes externos que ayuden a definir, por ejemplo, qué se entiende por: aprovechamiento, expresión e interacción escrita y oral. Por más obvio que parezca, los diseñadores deberán

referir por escrito la definición teórica de estos constructos; después, deberán elaborar la tabla de especificaciones en las que se defina operacionalmente el constructo, para finalmente validarla.

1.2. Elaboración del instrumento

a) Elaboración de reactivos y tareas

La elaboración del instrumento comienza propiamente con la selección de los integrantes de dos paneles de expertos. El primero, integrado por académicos competentes en los temas contenidos en el examen, quienes deben estar previamente capacitados para llevar a cabo esta tarea. Este primer grupo es el responsable de elaborar los reactivos o las tareas que conformarán el instrumento, según la cantidad que se determine en la tabla de especificaciones, así como el número de versiones del instrumento que se requieran elaborar. El segundo panel es el encargado de revisar y validar los reactivos para comprobar que se apegan y cumplen con las especificaciones requeridas en la etapa de diseño.

Los reactivos o tareas han de elaborarse siguiendo determinadas reglas, tanto de redacción como de estructura. Para ello, se recomienda elaborar lineamientos de creación y formatos para la elaboración de los reactivos o las tareas. Los formatos ayudarán a que los redactores lleven un control de los reactivos o las tareas que están realizando, en relación con las especificaciones del instrumento. Se recomienda elaborar el número de reactivos necesarios para ensamblar, por lo menos, dos o tres versiones equivalentes a las que se les denominará versiones piloto. Éstas deben ser probadas con sujetos que tengan las características de la población a la que está dirigido el instrumento, y es muy probable que en el pilotaje se encuentren algunos reactivos o tareas que en definitiva sea necesario eliminar.

b) Validación teórica y técnica de reactivos y tareas

Si bien los diseñadores de reactivos y tareas se apegan a los componentes de las tablas, es necesario hacer una revisión para corroborar que cumplen con lo especificado en los componentes, pero que cumplen a la vez con las reglas y los lineamientos para la construcción de reactivos y tareas. Los revisores deben ser expertos en el contenido del instrumento y utilizar una ficha de revisión, o lista de cotejo de validación, para emitir su dictamen, que puede ser: aprobado, aprobado con recomendaciones o rechazado. Los reactivos que se aprueben con recomendaciones deberán regresar al diseñador para que realice los ajustes necesarios.

En general, en los reactivos se revisa el contenido y la redacción. En cuanto al contenido, se debe revisar principalmente la correspondencia con las

especificaciones; además: que no haya más de una respuesta correcta; que la respuesta resuelva satisfactoriamente el reactivo; que no existan errores de contenido en la base, que puedan confundir a los examinados y, finalmente, que la base incluya toda la información estrictamente necesaria para la comprensión del reactivo. Por lo que se refiere a la redacción, se debe revisar que sea clara y adecuada para la población a la que se dirige el instrumento; asimismo, que exista coherencia gramatical entre la base y las opciones de respuesta, de manera que se evite dar claves para identificar fácilmente la respuesta correcta.

1.3. Validación empírica del instrumento

c) Pilotaje de reactivos y tareas

El segundo momento de validación consiste en pilotear o poner a prueba los reactivos y las tareas en condiciones lo más parecidas a las de la aplicación real, para estimar los tiempos de ejecución y los índices estadísticos de los mismos. Cuando un instrumento no se ha aplicado con anterioridad y se encuentra en fase de primer diseño, se debe utilizar el pilotaje denominado "preliminar", que señala que la mayoría o la totalidad de los reactivos tienen propiedades desconocidas y la validez del instrumento no ha sido probada. En el pilotaje preliminar se construyen instrumentos de prueba con dimensiones y características que pueden parecerse en su totalidad, o sólo parcialmente, a las de un instrumento real. En este tipo de pilotaje debe participar una muestra de la población objetivo. Generalmente, es necesario pilotear el doble o triple de reactivos necesarios para integrar una primera versión; acto seguido, se selecciona, desde este primer banco, el material de la primera aplicación y se ensambla la versión definitiva con aquellos reactivos y tareas que obtengan los mejores resultados estadísticos derivados del análisis.

d) Análisis psicométrico

Para el análisis de reactivos y tareas, existen diversos modelos matemáticos basados en la Teoría Clásica de los Tests (TCT), en la Teoría de Respuesta al Ítem (TRI) y en el análisis de Rasch. A partir de dichos modelos, es posible determinar las características psicométricas de los reactivos de manera individual, del instrumento en general, o por áreas o secciones, dependiendo de la información que se quiera analizar.

Debido a que existen diversas variables a considerar al hacer el diseño, es necesario tomar decisiones ante incidentes inesperados, las cuales deben procurar la máxima precisión posible, y esto se logra a partir del uso apropiado de la estadística.

En la evaluación del aprovechamiento de la EIE y de la EIO, necesariamente se deben elaborar tareas que permitan poner en evidencia lo que los alumnos son capaces de hacer para comunicarse en la lengua meta. Por lo tanto, es necesario que produzcan textos orales y escritos que deben ser calificados por profesores expertos en dicha lengua, para determinar si cumplen o no con lo que se espera de ellos. Para validar las tareas, es recomendable que varios profesores califiquen la misma muestra de actuación: las puntuaciones deberán ser analizadas mediante los mismos modelos psicométricos. En este caso se analizarán las puntuaciones emitidas por cada calificador.

2. Elaboración de la tabla de especificaciones

La tabla de especificaciones es una matriz que presenta la organización lógica de los componentes definidos operacionalmente, en la que se debe señalar el número de reactivos o tareas que requieren elaborarse, lo cual nos indicará la longitud del instrumento; es decir, el total de elementos que lo integrarán, así como su distribución por área. Para realizar las especificaciones, lo primero que se requiere definir son las áreas del instrumento (las agrupaciones de los componentes bajo las cuales se reportarán los resultados), por lo que se deben consignar nombres descriptivos que comuniquen un resultado con precisión.

La definición de la longitud del instrumento puede pasar por dos momentos: uno derivado del diseño, y otro de la validación. El primer momento se determina a partir de la aplicación de modelos matemáticos, así como de los elementos que los expertos en la lengua consideren pertinentes para demostrar que el alumno, concluido el curso, posee los conocimientos o la habilidad que se espera de él, para esto, es necesario siempre tomar en cuenta el tiempo que éste tardará en responder y realizar cada tarea. Una vez establecida la longitud del instrumento, se precisa validarla, mediante la siguiente pregunta: ¿los examinados contarán con el tiempo establecido en el diseño del instrumento para su resolución? Si la respuesta es afirmativa, la longitud puede quedar como se expresó en el diseño y se plasmó en la tabla de especificaciones. En caso de que la respuesta sea negativa, debido a que se cuenta con un tiempo mayor, la longitud permanece como se expresó en el diseño del instrumento; pero si es negativa debido a que el tiempo es menor al establecido en el diseño, se han de realizar los ajustes necesarios.

Para elaborar la tabla de especificaciones, existen diferentes formatos; sin embargo, cualquiera que se decida utilizar, deberá siempre tenerse en cuenta que la tabla contenga mínimamente la siguiente información:

- área
- subárea (cuando aplique)
- acción a evaluar o capacidad a evaluar
- condición en la que se realizará la acción, o en la que se pondrá en evidencia la capacidad a evaluar.
- suficiencia en la que se defina la cantidad con la que se asume que el sustentante posee el conocimiento, la habilidad o la competencia evaluada.
- nivel taxonómico
- número de reactivos o tareas por cada componente a evaluar

3. Análisis psicométrico

Los análisis psicométricos son la metodología compuesta por un conjunto de herramientas estadísticas que permiten aportar evidencia de la validez y confiabilidad de un instrumento de evaluación. Esto se logra mediante el cálculo de indicadores psicométricos, a partir de datos empíricos obtenidos en la aplicación de los exámenes, o bien de la opinión de los expertos.

Diseñar un instrumento de evaluación es un proceso que requiere seguir una metodología rigurosa. No obstante, incluso si se acatan todas las recomendaciones y se designan las mejores prácticas para el diseño de instrumentos de evaluación, al aplicar un examen por primera vez con estudiantes reales, muy probablemente se detectarán comportamientos inesperados y aspectos mejorables en los reactivos, los cuales no pudieron preverse durante el proceso de diseño. Por este motivo, siempre es recomendable pilotear¹ exámenes antes de su implementación definitiva, donde el resultado del examen puede tener un impacto real.

Como ya se mencionó anteriormente, existen dos grandes familias de modelos que se utilizan para realizar análisis psicométricos: la Teoría Clásica de los Tests (TCT) y la Teoría de Respuesta al Ítem (TRI). En general, la TCT es más fácil de implementar que la TRI, ya que es posible calcular todos sus índices² utilizando únicamente una hoja de cálculo; mientras que tasar los índices de la TRI es mucho más complicado sin el uso de un software especializado (Winsteps, BILOG-MG, R, etc.).

Sin embargo, la TCT sufre de una limitación muy importante: todos sus índices son dependientes de la muestra de sustentantes que responde el examen; esto es, los índices de la TCT serán distintos cada vez que el

1. Pilotaje: probar un instrumento de evaluación en un ambiente lo más parecido a una aplicación real.

2. Los índices, o indicadores, son valores que resumen características estadísticas de los exámenes y sus ítems o criterios de calificación, por lo que permiten tomar decisiones respecto a su pertinencia; esto aporta evidencia de su validez y confiabilidad.

examen se aplique a un grupo diferente de personas. En contraste, los índices de la TRI son independientes de la muestra.

La TRI es una teoría que engloba muchos modelos con múltiples características y propósitos. Si bien el más conocido y potente es el modelo logístico de tres parámetros, éste requiere de muestras de gran tamaño (alrededor de 1,000 examinados), que en la práctica puede ser muy difícil reunir, particularmente cuando se trabaja con un constructo como la evaluación de lengua extranjera, donde la población objetivo es muy específica. Una alternativa que puede implementarse con tamaños de muestra más reducidos (100 examinados) es el modelo logístico de un parámetro, también conocido como Modelo de Rasch (Rasch, 1993).

El Modelo de Escalas de Calificación (Andrich, 1988) es una adaptación sencilla del modelo de Rasch, que está incorporado en la mayoría de softwares especializados para el análisis de la TRI. Se utiliza cuando el examen no está compuesto por reactivos dicotómicos³ sino por escalas politómicas⁴, como es el caso de pruebas productivas que son puntuadas con criterios de calificación.

La Medición de Rasch de Múltiples Facetas (Linacre, 1989) es otra derivación del Modelo de Rasch, que se puede implementar con el software especializado FACETS. Esta adaptación considera que el puntaje observado en una prueba productiva es el resultado de la interacción de múltiples variables o facetas, por lo que permite discernir qué cantidad de un puntaje obtenido en un criterio de calificación se debe a la dificultad propia del criterio; a su vez, qué proporción deriva de la dificultad global de la tarea, cuánta de la severidad del calificador, y cuánta de la habilidad del examinado. Con lo cual, este método permite evaluaciones más justas, además de facilitar la detección de sesgos o problemas en tareas o calificadores. Para que el modelo sea capaz de diferenciar entre todos los factores que se combinan al producir un puntaje, es necesario que la producción de un mismo examinado sea evaluada por más de un calificador, y que todos los que cumplen este rol arbitren las producciones de múltiples examinados.

No se pretende aquí exponer a profundidad los marcos teóricos, las justificaciones y los supuestos que sustentan estos modelos⁵, sino ofrecer una guía general y práctica de su uso en el diseño de instrumentos de evaluación. A continuación, se presentan los índices psicométricos más empleados para el análisis de pruebas productivas y cómo se interpretan sus valores.

3. Aquellos reactivos en los que sólo se observan dos resultados posibles: acierto o error.

4. Reactivos o escalas con más de dos niveles de logro, por ejemplo, un criterio de calificación que se evalúa sobre una escala de 5 puntos.

5. Para una revisión de estos temas, se recomienda, además de las referencias especificadas, consultar a Martínez, Hernández & Hernández (2006).

3.1. Teoría Clásica de los Test

3.1.1 Dificultad del criterio

Este concepto aplica para pruebas productivas puntuadas con criterios de calificación, y se refiere al puntaje promedio obtenido por todos los examinados bajo ese criterio, con respecto al puntaje máximo que es posible obtener en él. El valor de la dificultad siempre se encuentra entre 0 y 1. Aunque puede parecer contraintuitivo en un principio, un índice de dificultad cercano a 0 indica que el reactivo es difícil, mientras que si está más cercano a 1, señala que es fácil.

La función de los reactivos o criterios de un instrumento de evaluación es aportar información sobre el nivel de habilidad con el que cuentan los examinados. Tanto los criterios que son tan fáciles que casi todos los sustentantes pueden cumplirlos, como los que son tan difíciles que ninguno logra acreditar, no aportan información valiosa sobre el nivel de habilidad de los examinados, por lo que se consideran inadecuados.

Un índice de dificultad aceptable debe estar, al menos, por arriba de 0.10 y por debajo de 0.90. Para mediciones más rigurosas, también se utiliza 0.20 como límite inferior y 0.80 como superior.

3.1.2 Correlación del criterio

Este atributo, como su nombre lo indica, se refiere a la correlación⁶ que existe entre el puntaje de los examinados en el criterio y su puntaje en todo el examen.

Mientras más cercano a 1 sea el índice, mejor es el comportamiento del criterio, porque esto indica que los sustentantes que mostraron un buen desempeño en este criterio, también tuvieron un buen desempeño en el examen; por su parte, aquéllos con un mal desempeño en el criterio, también presentan un menor desempeño general. En contraste, cuando el índice es negativo o registra un valor cercano a 0, es señal de que los sustentantes con un buen desempeño en el criterio, no tienen un buen desempeño en el examen, y quienes exhiben un mal desempeño en el criterio, cuentan con un buen desempeño en el resto del examen. Esto es señal de que el criterio no está midiendo el mismo constructo que el resto del examen. Para que el índice de correlación de un reactivo se considere aceptable, debe estar por lo menos por arriba de 0.15.

6. Correlación: relación o asociación estadística entre dos variables. Una correlación alta indica que cambios en una variable en cierta dirección, corresponden con cambios en la otra variable, asimismo en una dirección específica.

Como ya se dijo, los índices de la TCT dependen de la muestra, por lo que un criterio que mostró una dificultad adecuada en una aplicación, puede presentar una inadecuada en otra con una muestra diferente. Si únicamente se implementan análisis de la TCT, la mejor forma de atenuar esta limitación es piloteando el examen con muestras lo más grandes posibles, cuyo nivel de habilidad sea lo más cercano o representativo de la población real a la que se le aplicará el examen.

3.2. Teoría de Respuesta al Ítem (Modelo de Escalas de Calificación y Medición de Rasch de Múltiples Facetas)

3.2.1 Dificultad del criterio y de la tarea; severidad del calificador; habilidad del examinado

De acuerdo con los modelos de la TRI, todas las facetas que interactúan en una prueba productiva para obtener el puntaje observado, se encuentran en una misma dimensión, por lo tanto, en la misma escala, y son directamente comparables. El software especializado en TRI reporta estos valores en una unidad conocida como *logits*⁷. Por convención, a la dificultad promedio de la prueba se le asigna un valor de dificultad de 0 *logits*, de modo que valores de dificultad, severidad o habilidad que están por encima de 0 (positivos), se encuentran arriba del promedio, y valores inferiores a 0 (negativos), se sitúan por debajo del promedio.

Tras la primera aplicación de un instrumento, una vez que se estiman los índices psicométricos de los reactivos, estos pueden anclarse⁸, lo que los vuelve independientes de la muestra, pues los índices del examen no cambian, aunque la muestra de examinados sea diferente.

3.2.2 Error estándar de la medida del criterio/tarea/calificador/examinado

Se refiere a la precisión con la que se estimaron los índices psicométricos. Si el error es pequeño, se puede decir que la precisión es alta: por lo tanto, es deseable que este valor se encuentre lo más cercano posible a 0.

7. Del inglés "log odds units" (unidad de logaritmo de momios), debido a que se calcula a partir de una transformación logarítmica que convierte valores probabilísticos (momios) en puntajes de una escala continua.

8. Anclaje: procedimiento en el que los índices estimados de ciertos reactivos, tareas, criterios o calificadores se "fijan" de modo que no se vuelven a estimar, y cuando se crean nuevas tareas o se incluyen a nuevos calificadores, sus índices psicométricos se estiman a partir de los índices anclados que sirven como punto de referencia.

3.2.3 Ajuste de criterio/tarea/calificador/examinado

Se refiere al grado en que los puntajes observados en el examen se comportan de acuerdo a lo previsto por el modelo de Rasch; por ejemplo, que los examinados con el mayor nivel de habilidad obtengan los puntajes más altos en los criterios de mayor dificultad. Cuando se observan comportamientos inesperados, como que sustentantes con un bajo nivel de habilidad logren resolver las tareas más complejas o, por el contrario, que los de alto nivel obtengan puntajes bajos en los criterios más fáciles, se dice que hay un desajuste.

Un índice de ajuste ideal se encuentra muy cercano a 1.00; mientras que cuando está por arriba de 1.50, se considera que hay un desajuste leve, y si es mayor a 2.00, es considerado como un desajuste severo. Este fenómeno puede deberse a que la tarea o el criterio no están midiendo el mismo constructo que el resto del examen, o bien a que el calificador está interpretando incorrectamente los criterios de calificación; también cabe la posibilidad de que los examinados más hábiles hayan respondido sin dedicar el esfuerzo suficiente, o que los menos diestros hicieran trampa.

3.2.4 Confiabilidad de separación del criterio/tarea

Es un índice que señala, para el examen completo, el grado en que el puntaje obtenido por los examinados refleja su verdadero nivel de habilidad. Mientras más se acerca el valor de este índice a 1, mayor es la confiabilidad del examen. Idealmente, su valor no debe estar por debajo de 0.70.

3.2.5 Contraste de diferencias entre calificadores

Se emplea para determinar si hay diferencias importantes en la severidad de los calificadores o, si en general, todos ejercen un rigor similar. Para poder decir que los calificadores cuentan con un nivel de exigencia afín, el *valor p*⁹ debe ser mayor a 0.05.

3.2.6 Porcentaje de acuerdo entre calificadores

Alude a la medida de la confiabilidad entre los calificadores. Indica el porcentaje de veces que un calificador asigna el mismo puntaje que sus pares, en idénticas condiciones (mismo examinado, tarea y criterio). Se compara el porcentaje observado con el porcentaje que se esperaría observar debido al azar. A mayor porcentaje de acuerdo, mayor confiabilidad entre los

9. Este valor se implementa en pruebas de hipótesis y se refiere a la probabilidad de que los resultados observados se hayan debido al azar. Mientras más pequeño es este valor, es más probable que los resultados observados sean estadísticamente significativos.

calificadores; idealmente, el porcentaje observado nunca debe ser menor que el esperado por el azar.

3.2.7. Correlación del calificador

Es una correlación entre los puntajes asignados por un calificador y sus pares. Un índice de correlación bajo (menor a 0.15) indica que el calificador es inconsistente con respecto a la evaluación de los demás.

4. Definición de puntos de corte

Además de determinar la adecuación de las tareas, así como de los criterios y calificadores que forman parte de los instrumentos de evaluación, los análisis psicométricos también permiten definir los puntos de corte de un instrumento.

De acuerdo con el CENEVAL (2000), un *punto de corte* es el puntaje mínimo o criterio a superar para considerar que el desempeño de una persona es satisfactorio y distinguirlo de otro que no lo es.

Definir puntos de corte permite utilizar el puntaje que obtienen los sustentantes en un examen, para determinar si pertenecen a una u otra categoría o nivel de logro, de acuerdo con su desempeño (Jornet y Backhoff, 2008). Estas categorías nos permiten tomar decisiones posteriores respecto a los examinados.

Un examen puede tener un sólo punto de corte que separe dos categorías; por ejemplo: las categorías Aprobado y No Aprobado, dentro de un examen de aprovechamiento. También es posible que haya más de un punto de corte, si es necesario separar múltiples categorías. Si se trata de un examen para determinar en qué nivel de curso debe ser colocado el estudiante, debe haber un punto de corte para cada uno de dichos niveles.

El punto de corte, acotado a la evaluación de ELE, será determinado por la cantidad de puntos que el examinado necesita obtener en las categorías de calificación establecidas en el instrumento. Esto, específicamente en pruebas que evalúan las habilidades productivas, como la expresión oral y escrita, ya que estos rubros son valorados por un calificador, mediante el uso de criterios de calificación.

En pruebas criterioles¹⁰, la definición de puntos de corte se realiza mediante un proceso a juicio de expertos, en el que deben participar personas con conocimiento y experiencia en el constructo que medirá el examen; dichos integrantes reciben un entrenamiento previo para participar en esta actividad. Cada experto realiza su juicio de manera independiente y sin conocer los juicios de los demás para prevenir que haya sesgos.

10. Pruebas criterioles: se establece un criterio externo que los examinados deben alcanzar.

4.1. Consideraciones generales para el proceso de definir puntos de corte

Pueden haber tantos métodos para definir puntos de corte, como exámenes y propósitos para evaluar existen (APA, 2014); pero independientemente del método utilizado, ciertas consideraciones metodológicas deben tomarse en cuenta para realizar este proceso (Cizek, 2005). A saber:

4.1.1. Identificar el propósito de los puntos de corte

Aun cuando haya exámenes que evalúan contenidos similares, pueden tener puntos de corte muy diferentes, dependiendo de cuál es su propósito. Un examen de colocación, por ejemplo, busca determinar cuál es el nivel de curso más apropiado para los examinados, de acuerdo con sus habilidades y conocimientos actuales; mientras que un examen de aprovechamiento pretende evaluar las habilidades y conocimientos adquiridos durante el proceso educativo.

El propósito del examen y sus puntos de corte, las posibles consecuencias de las decisiones que se tomen mediante su uso y, en general, todo el contexto que rodea al instrumento de evaluación, siempre deberán ser explicados a los expertos que participan en el proceso de definir dichos puntos; así, ellos podrán tomar decisiones informadas, por lo que toda esta información ha de definirse durante las primeras etapas de desarrollo del instrumento.

4.1.2. Planificar la actividad para definir puntos de corte

Es necesario decidir cuándo es el momento más conveniente para establecer los puntos de corte dentro del proceso de desarrollo del instrumento. Esta actividad siempre se llevará a cabo después de construir, por lo menos, una versión de examen; pero puede realizarse antes o después de contar con evidencia empírica del funcionamiento del examen.

La robustez de algunos métodos para definir puntos de corte radica en que toman en cuenta evidencia del comportamiento real del examen, mientras que con otros métodos se pueden establecer los puntos de corte incluso antes de pilotear el examen. La desventaja de estos métodos es que los reactivos y las tareas del examen podrían desestimarse durante el proceso de pilotaje.

4.1.3. Crear referentes para los puntos de corte

Este rubro apunta a definir los referentes de desempeño o niveles de logro, es decir, los nombres de las categorías que se van a utilizar y los descriptores de

los niveles de desempeño de cada una. Para esta tarea se debe considerar el propósito del examen, la definición del constructo y la tabla de especificaciones.

En el caso de la enseñanza y certificación de ELE, es común adoptar los niveles de logro del MCER, o un referente similar, pues se trata de un marco de referencia con descriptores bien definidos y ampliamente utilizado (Consejo de Europa, 2002). Los niveles de logro deben estar bien acotados y definidos, con el fin de determinar con precisión el grado mínimo de habilidad requerido para alcanzar estos niveles.

4.1.4. Selección y entrenamiento de participantes para el juicio de expertos

Para este aspecto, se requiere elegir a un grupo de expertos que cumplan con las características necesarias para participar en el proceso de establecer puntos de corte. En general, aunque depende del método utilizado, se recomienda la valoración de 10 expertos, y nunca menos de 5. Los jueces deben recibir un entrenamiento en el que se les explique cuál es el objetivo del examen, y que además se les capacite sobre el método implementado para definir los puntos de corte.

4.1.5. Retroalimentación a los participantes

Debido a la subjetividad que conlleva el juicio de los expertos, inevitablemente habrá discrepancias en la opinión de los participantes, aun cuando todos sean muy competentes en el constructo que se busca medir.

Una buena práctica, si bien no indispensable, es realizar el juicio de expertos dos veces. Para esto, después de que los expertos realizan su primera labor, se les muestran los juicios de los demás participantes y son invitados a discutir entre ellos para justificar sus decisiones, antes de realizar la actividad por segunda vez. El objetivo de la retroalimentación no es uniformar la opinión de los expertos, sino garantizar que todos tengan el mismo entendimiento de los niveles de logro que se están utilizando en el examen, así como enriquecer sus perspectivas individuales.

4.1.6. Evaluar el proceso de definir puntos de corte

Desde que se comienza el proceso de definir puntos de corte, se deben producir y recopilar evidencias de cada paso del mismo. Esto sirve no sólo como evidencia de la validez del instrumento, sino que también permite evaluar y retroalimentar el proceso de diseño del instrumento.

4.2. Métodos para establecer puntos de corte

Si bien, como queda dicho, es imposible enumerar todos los métodos con los que se pueden definir puntos de corte, un concepto fundamental que se le enseña a los participantes en el juicio de expertos (durante su entrenamiento para la tarea, y que es común en la gran mayoría de métodos utilizados) es el concepto de la “persona apenas aceptable” o “persona mínimamente aceptable”. Esto es, imaginar a una persona con el nivel de habilidades y capacidades apenas suficientes, que permitan decir que “entra rozando” o “de panzazo” dentro de un nivel de logro o categoría de desempeño. Por ejemplo, si tenemos que pensar en una persona apenas aceptable para el nivel B2 del MCER, debemos considerar qué habilidades son necesarias para estar dentro de esta categoría, tomando en cuenta también que la más pequeña disminución en estas habilidades, bastaría para que la persona ya no alcanzara a entrar en este nivel.

Durante el entrenamiento, los jueces deben comprender este concepto, así como ser capaces de imaginar a las personas apenas aceptables para cada uno de los niveles de logro o categorías que se van a evaluar. Es importante garantizar durante el entrenamiento que todos los jueces entienden estos conceptos, y que están de acuerdo respecto a qué habilidades debe poseer la persona mínimamente aceptable en cada nivel de logro; debe ser así pues, aunque sus juicios son independientes, deben partir de este concepto compartido.

Uno de los métodos más conocidos y utilizados para definir puntos de corte, por su facilidad de implementación, es el Método de Angoff, que se ha implementado varias ocasiones en instrumentos desarrollados en el CEPE. Para pruebas productivas, en este método se les presentan a los expertos las tareas que componen el examen, junto con sus criterios de calificación, y se les pide que imaginen que 100 personas apenas aceptables responden el examen. Los expertos deben indicar cuántos puntos, en promedio (de acuerdo con la escala utilizada), creen que obtendría dicho grupo en cada rúbrica o criterio de calificación de cada tarea. La forma más fácil de determinar este promedio es indicando cuántas de las 100 personas apenas aceptables obtendrán cada calificación posible; acto seguido, promediar los valores resultantes. Los jueces deberán repetir este proceso por cada nivel que la prueba pretenda medir.

Finalmente, se suman los puntajes promedio de todos los criterios de cada experto, para obtener su respectivo punto de corte, y todos estos se promedian para obtener un punto de corte definitivo. Para una revisión muy completa de los métodos más utilizados para definir puntos de corte, se puede revisar a Cizek (2005), así como el Council of Europe (2009).

5. Lineamientos técnicos y reglas para la elaboración de tareas para la evaluación de la EIE y la EIO

La evaluación de lenguas extranjeras, cuyo constructo se sustenta en el enfoque comunicativo, necesariamente debe basarse en la realización de tareas mediante las que los aprendientes demuestren la competencia comunicativa que han alcanzado en la lengua meta. Evaluar con base en la realización de tareas que representan la realidad es la mejor y más contundente forma para medir la competencia lingüística adquirida.

Una evaluación ciertamente significativa debe basarse en lo que los individuos requieren saber para comunicarse en el mundo real. Las tareas bien definidas son la mejor estrategia para poner en evidencia lo que aprendieron, o lo que son capaces de hacer, las personas para comunicarse en la lengua meta; así como para desenvolverse en determinadas situaciones comunicativas y diferentes ámbitos de la vida real, según su nivel de habilidad.

De acuerdo con el *Marco Común Europeo de Referencia para las lenguas* (Consejo De Europa, 2002), las tareas son actividades que se realizan en la vida cotidiana, en los diferentes ámbitos donde se desempeñan los individuos. La realización de una tarea requiere de la activación estratégica de competencias específicas en un ámbito determinado, con un objetivo perfectamente definido y un resultado puntual. Las tareas pueden tener un carácter enormemente variado y, en mayor o menor medida, comprender actividades de la lengua. En síntesis, es posible afirmar que una tarea permite resolver un problema concreto, como se haría en situaciones comunicativas reales.

De manera concreta, una tarea se define como un planteamiento que demanda una producción por parte del examinado y tiene como finalidad conocer la competencia comunicativa alcanzada por éste; consta de unas instrucciones y de un texto de entrada. Una tarea puede incluir uno o más reactivos cuando se evalúa la comprensión auditiva o de lectura. Así pues, una tarea se conforma de dos partes:

- Las instrucciones, que son las pautas que se le proporcionan al examinado para producir el texto de salida.
- Un texto de entrada, que consiste en un estímulo que provocará el texto de salida.

Según el MCER (Consejo de Europa, 2002), para evaluar las habilidades de comprensión se pueden diseñar tareas con un mismo material de entrada (input), en tanto se puedan prever distintos resultados, ya sea de forma cuantitativa (de respuesta abierta), o de forma cualitativa (de respuesta cerrada). El texto de entrada puede contener diferente información, variar

en grado de complejidad cognitiva o cantidad de apoyo a disposición de los examinados; por ejemplo: material visual, palabras clave, pistas, gráficos, diagramas, entre otros. Es posible escuchar o leer cada texto las veces que se consideren necesarias, o bien se pueden establecer límites si así se desea. En el caso de la interacción y la expresión, las condiciones para realizar la tarea son factibles de manipular, con el fin de que resulte más o menos compleja; por ejemplo, variando el tiempo permitido para planificar su realización, la duración de la interacción o la expresión, el carácter predecible o impredecible, la cantidad y tipo de apoyo proporcionado, entre otros elementos. Todo esto dependerá, desde luego, del grado de competencia que se vaya a evaluar.

Para elaborar tareas enfocadas a calificar la expresión oral y escrita, además de lo que ya se ha mencionado, se deben seguir determinados lineamientos técnicos y reglas de construcción, de tal forma que todos los examinados cuenten con las mismas condiciones y oportunidades para demostrar su competencia y producir lo que se espera de ellos.

Como primer paso, se recomienda que quienes diseñen las tareas seleccionen textos que sean auténticos y realicen las adaptaciones necesarias para que los examinados produzcan lo que se les requiere, en apego a lo establecido en la tablas de especificaciones. Dada la diversidad cultural de quienes realizarán las tareas, es altamente conveniente evitar el uso de textos que aporten contenidos con fuerte carga ideológica, o que induzcan algún tipo de sesgo cultural que pueda alterar, perjudicar o desfavorecer el desempeño de los sustentantes. Es importante considerar que todos los materiales que se utilicen, presenten un contenido lo más neutro posible, de tal forma que no resulten ofensivos para nadie; asimismo, se debe cuidar que los materiales elegidos provengan de fuentes confiables y reconocidas, que dispongan de normas de estilo y que garanticen sus trabajos de edición y corrección.

Se recomienda evitar:

- Contenidos de carácter ideológico, político, religioso y sexual de cualquier tipo.
- Nombres o logotipos de marcas comerciales. En caso de que sea imprescindible la utilización de alguna, lo más conveniente será ajustarla o cambiarla por una ficticia que, por su puesto, resulte creíble.
- Temas estereotipados sobre la sociedad, la historia, la religión, el arte o la política, que puedan ser agresivos o provocativos para determinadas culturas.

Debido a que se trata de tareas que se utilizarán para evaluar la competencia comunicativa de español, también es conveniente evitar el uso de textos que hayan sido previamente traducidos de otras lenguas.

Referencias

Aiken, L. (2003). *Tests psicológicos y evaluación*. México: Pearson Educación.

ALTE (2011). *Manual para el diseño y desarrollo de pruebas de certificación de idiomas*. Estrasburgo, Francia: Language Policy Unit of the Council of Europe.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014) *Standards for educational and psychological testing*. Washington, D.C.

Andrich, D.A. (1988). *Rasch models for measurement*. Newbury Park, CA: Sage Publications.

Brown, D. (2000). *Principles of language learning & teaching*. New York: Longman.

CENEVAL (2000). *Estándares de calidad para instrumentos de evaluación educativa*. México: INEE.

Cizek, G. (2005) Standard Setting. En Downing, S. y Haladyna, T. (Eds.) *Handbook of Test Development* (pp. 225-260). Mahwah, New Jersey: Taylor & Francis.

Consejo de Europa (2002). *Marco Común de Referencia para las Lenguas*. Ministerio de Educación, Cultura y Deporte & grupo ANAYA. Recuperado de <http://cvc.cervantes/obref/marco>

Council of Europe (2009). *Relating Language Examinations to the 'Common European Framework of Reference for Languages: Learning, Teaching, Assessment'* (CEFR).

Jornet, J, y Backhoff, E. (2008). *Modelo para la determinación de Niveles de Logro y Puntos de Corte de los Exámenes de la Calidad y el Logro Educativos (Excale)*. México: INEE.

Linacre, J. (1989). *Many-facet Rasch measurement*. Chicago: MESA Press.

Martínez, M., Hernández, M.J. & Hernández, M.V. (2006). *Psicometría*. Madrid: Alianza Editorial.

Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests*. Chicago: MESA Press.

Tristán, A. y Vidal, R. (2006). *Estándares de calidad para pruebas objetivas*. Colombia: Aula Abierta.

Bibliografía recomendada

Cronbach, L. J. (1951). *Coefficient alpha and the internal structure of tests*. *Psychometrika*, 16 (3), 297-334.

Linacre, J. (2008). The Expected Value of a Point-Biserial (or Similar) Correlation. *Rasch Measurement Transactions*, 2008, 22:1 p. 1154.

SECCIÓN II.

LA EXPRESIÓN ORAL

CAPÍTULO 3

LA HABILIDAD ORAL Y SU EVALUACIÓN

Klaudia Yvonne **Buzo Hachec**

Moisés **López Olea**

1. Introducción

Como primer paso para desarrollar instrumentos de evaluación adecuados, es necesario definir qué se va a valorar a través de estos. En este capítulo, nos enfocaremos en describir los aspectos a observar y medir de la oralidad en el ámbito del aprovechamiento; es decir, qué evaluar al final de un curso de lengua con respecto a los logros obtenidos y los aprendizajes esperados de la oralidad. En otras palabras, el objetivo de este capítulo es delimitar teóricamente los constructos de la habilidad oral que se buscan evaluar al final de un curso. Con esto en mente, definiremos los constructos pertinentes: la oralidad, la expresión, la interacción y la conversación; posteriormente, comentaremos sobre la subjetividad, los criterios y los parámetros para diseñar una evaluación oral válida y confiable. En términos generales, este capítulo presenta la definición teórica de los constructos que se pueden medir en la evaluación oral para asegurar la validez de los instrumentos.

2. Conceptos generales de la oralidad

La oralidad es un rasgo intrínseco al ser humano y lo distingue de otros seres animados, es el medio primordial para el desenvolvimiento social y la generación de conocimiento (Martín Leralta, 2019). En esta sección se precisará qué es la oralidad y cuáles son sus características principales. Posteriormente, se definirán los constructos de 'expresión', 'interacción' y 'conversación', como el primer paso para diseñar un instrumento de evaluación oral de español como segunda lengua. De esta manera, buscamos exponer cuáles son los aspectos a evaluar en el ámbito de la oralidad.

2.1 La oralidad

Hablar es una de las actividades humanas más básicas y fundamentales. Tradicionalmente, se concibe la *expresión oral* como un término técnico para referirse a una de las cuatro destrezas lingüísticas básicas, o actividades comunicativas de la lengua, que se corresponde con la capacidad

de producir un texto oral (Consejo de Europa, 2001; Cortés-Moreno, 2015; Nation, 2011).

La oralidad es central en el aprendizaje de una segunda lengua (Nation, 2011), ya que normalmente las personas van a clases de lengua para hablar o conversar con otros; típicamente preguntamos: ¿hablas español?, ¿hablas inglés? Considerando que la oralidad es central en la enseñanza de segundas lenguas, podemos concluir que también lo es en la evaluación. Hay ciertos rasgos de la oralidad que hay que tomar en cuenta para evaluarla; a continuación, ahondaremos al respecto.

2.1.1 Rasgos fundamentales de la oralidad

Aunque la oralidad y la escritura se realicen en la misma lengua, ambas actividades son diferentes y autónomas; representan dos modos de comunicación distintos, con sus características, entornos y funciones propias (Baralo, 2000). Por ello, conviene subrayar los rasgos fundamentales de la oralidad para caracterizarla en un ámbito de evaluación.

Primero, la oralidad se realiza a través de sonidos, por lo que un rasgo fundamental será la pronunciación, así como el volumen y la entonación (Pinilla-Gómez, 2004). En segundo lugar, la oralidad se desarrolla de manera natural en una primera lengua durante varios años. De acuerdo con Bordón Martínez (2004), esto no es así en una segunda lengua, debido a que las destrezas orales suelen ser las más difíciles de controlar para los estudiantes de lengua. De acuerdo con la autora, estas dificultades no se deben obviar, ni en el aula ni en su evaluación. En tercer lugar, la oralidad no suele ser planeada, aunque hay géneros orales que tienen mayor planeación. No obstante, siempre que una persona hable, lo hará más o menos de manera automática; lo cual le permite detenerse y cambiar lo que estaba diciendo, o dejar enunciados sin terminar, lo que se conoce como interrupciones en el discurso. En cuarto lugar, la oralidad suele ser inmediata, por lo que está unida con la espontaneidad del habla. Hablar suele dejar poco tiempo para pensar o planear qué decir; asimismo, la retroalimentación comúnmente se da en la inmediatez (Pinilla Gómez, 2004). Por ello, en la oralidad hay pausas y silencios.

Otra característica fundamental de la oralidad son las palabras, frases o estrategias para ganar tiempo y pensar, hacer pausas o reaccionar al interlocutor (Luoma, 2004). Entre éstas, podemos pensar en marcadores del discurso como *bueno, vale, hombre*; interjecciones como *ah, eh, y oh*; además de muletillas como *mmh, ahm y este*, entre muchas otras. Es natural que aparezcan estas expresiones lingüísticas cuando cualquier persona habla en cualquier situación comunicativa.

Para cerrar, Luoma (2004) destaca que la oralidad crea la imagen de un hablante para los otros, ya que, de manera consciente o inconsciente, cuando éste habla, usa la velocidad, las pausas, la variación en la entonación y el volumen para apoyar y realzar lo que se está diciendo. Además, estos atributos muestran a los otros la identidad, origen, nacionalidad, entre otros rasgos, del hablante.

Todos estos rasgos fundamentales deben ser tomados en cuenta a la hora de evaluar la oralidad en una segunda lengua. Por ejemplo, no se puede castigar en la evaluación el uso de muletillas, ya que son naturales a la misma oralidad; tampoco el hecho de que un hablante no termine sus enunciados o haga pausas para pensar. Será necesario tomar en cuenta que en una situación de evaluación, probablemente los estudiantes requieran cierto tiempo para planear su discurso. En resumen, no recomendamos penalizar en las calificaciones aquello que es natural al habla, incluso en hablantes nativos.

2.2. Los constructos de la oralidad

En la evaluación oral, como en cualquier otra, los examinadores y calificadores asignan puntajes e interpretan calificaciones, con una referencia implícita o explícita al constructo en el cual se basa el examen o la prueba (Young, 2000). Un problema en que la definición de constructos sea implícita, o se dé por sentada, es que los calificadores pueden tener distintas ideas de lo que hay que medir en una evaluación; por ejemplo, en la oralidad, qué es una fluidez suficiente para obtener puntajes altos y cuál obtendría bajos; o se califica con menos puntos a quien tiene más errores gramaticales, aunque no se trate de un examen de gramática, sino de oralidad; o bien, se otorgan más puntos a quien hable más, aunque hay culturas cuyas tradiciones discursivas no permiten robar turnos de habla ni extenderse en su discurso. Estas inconsistencias originan injusticias en la evaluación oral, por ello es fundamental definir clara y *explícitamente* el constructo sobre el cual se basa la prueba o el examen. Los examinadores deben estar formados debidamente para elicitar la oralidad necesaria para evaluar a los estudiantes, y los calificadores, para asignar puntajes adecuados. Tomando en cuenta lo anterior, en las siguientes secciones, se discuten y se definen los constructos de 'expresión', 'interacción' y 'conversación'¹.

1. De manera común, se hace una distinción únicamente entre expresión e interacción en la enseñanza y evaluación de segundas lenguas. Recientemente, Martín Leralta (2019) destaca la importancia sobre aprender a conversar. Por razones teóricas, conviene distinguir la conversación como un constructo independiente, ya que tiene características propias que la distinguen de otras interacciones, como se verá en la sección 2.2.3.

2.2.1 Expresión oral

Una persona puede hablar consigo misma o con otros; en realidad, expresarse no requiere, necesariamente, de interlocutores ya que hablar puede ser una actividad individual (Luoma, 2004). Esta perspectiva sobre la expresión oral es útil sobre todo en el ámbito educativo (Luoma, 2004). Como constructo, la expresión oral puede entenderse como la producción de enunciados y mensajes en la modalidad oral, asociada a géneros monologados o a actividades individuales (Luoma, 2004; Consejo de Europa, 2001). En un monólogo, una persona habla, sin necesariamente esperar una respuesta; esto tiene implicaciones para el diseño de un examen oral, ya que si se habla de expresión oral, en las pruebas y tareas del examen habrá solamente un estudiante. En términos de definición de constructo, podemos pensar la expresión oral como la producción de mensajes en monólogo.

Los géneros orales asociados a la expresión oral monologada, siguiendo a Muñoz-Basols y Gironzetti (2019), son: una anécdota para dar a conocer un suceso; un relato para contar una historia; una exposición o presentación para hablar de un tema ante el grupo; una conferencia para divulgar conocimientos. A estos géneros orales monologados, podríamos añadir, por ejemplo, un audio que se envíe en aplicaciones de mensajería instantánea, o un podcast que se grabe de manera individual para tratar diversos temas.

En este tipo de evaluación, la expresión oral permitirá elicitar muestras orales de un estudiante individual, para escucharlas y evaluarlas. Antes de la evaluación del aprendizaje, en el aula debería haberse trabajado el *constructo* de expresión, prestando atención a parámetros como la fluidez y la cohesión; además de esto, será importante preparar a los estudiantes para los géneros orales monologados: sería injusto que en un examen oral se le pida al estudiante una presentación individual, cuando en la clase nunca se trabajó ni se mejoró este género. En síntesis, en el aula se puede enseñar propiamente el uso de la lengua oral, aunado a los géneros orales monologados; una vez que se enseña, entonces es justo evaluarlo.

2.2.2 Interacción oral

A diferencia de la expresión monologada, en la interacción oral intervienen dos o más participantes, además de contar con otras cualidades específicas. Cortés-Moreno (2015) propone incluir la interacción y la mediación a las cuatro destrezas básicas: la expresión oral y escrita, así como la comprensión auditiva y la lectora. Autores como Sun (2014) afirman que se requiere un cambio de perspectiva en la enseñanza de lenguas, que consistiría en efectuar una transición de la competencia comunicativa a la *competencia interaccional*. Esta concepción surge debido a que hay todavía un predominio de la competencia lingüística actualmente en las clases de

lengua, factor que repercute en el hecho de que los estudiantes posean un amplio dominio gramatical, pero una escasa aptitud para desenvolverse en contacto directo con hablantes nativos de la lengua meta, en esquemas interaccionales propios de la cultura meta.

Si se brinda una enseñanza bien fundamentada de cómo interactuar en la lengua meta, y con esto se desarrolla la competencia interaccional, entonces también es posible evaluarla. Para ello, es necesario definir el constructo de *interacción*, como se hizo anteriormente con el de *expresión*. La interacción implica negociar significados intencionales en una co-construcción en la que el hablante ajusta su oralidad al efecto que busca causar en su interlocutor; asimismo, entra en juego la capacidad de anticiparse a la respuesta del interlocutor, para evitar posibles malos entendidos y esclarecer los propósitos comunicativos y las intenciones de uno mismo y del interlocutor (He y Young, 1998; Nation, 2011; Sun, 2014; Jacoby y Ochs, 1995; Young, 1999, 2000). En resumen, la interacción implica un hablante y al menos un interlocutor, que buscan comunicarse en los términos simples de una transacción; por ejemplo: obtener información sobre una cosa o lugar, pedir informes sobre trámites, hacer una cita, ir con el médico, entre otras.

En estas interacciones, los participantes tienen un estatus asimétrico, es decir, se encuentran en una situación dispar, ya que uno ostenta una situación de poder, mientras que el otro no. Con la frase "situación de poder", nos referimos a que hay un hablante que proveerá lo que se busca en la interacción, y otro que lo recibirá. Por ejemplo, en una cita con el médico, éste posee más conocimiento sobre los síntomas y las condiciones que el paciente manifiesta; o bien, para pedir informes, uno de los participantes cuenta con la información que el otro requiere; o si le pides permiso a tu jefe para tener vacaciones, este participante lo otorga o no. En cada una de estas interacciones, hay ciertos comportamientos y ciertas respuestas esperadas, de acuerdo con el estatus de cada participante. Otros ejemplos podrían ser: pedir comida en un restaurante, hacer un trámite para obtener el pasaporte, inscribirse a un curso, invitar a salir a alguien, etc.

Sun (2014) propone analizar esquemáticamente este tipo de interacciones, para así poder brindarle al estudiante una visión completa y holística al respecto. Por ejemplo, en la Figura 1, podemos ver un esquema que puede emplearse para pedir disculpas:

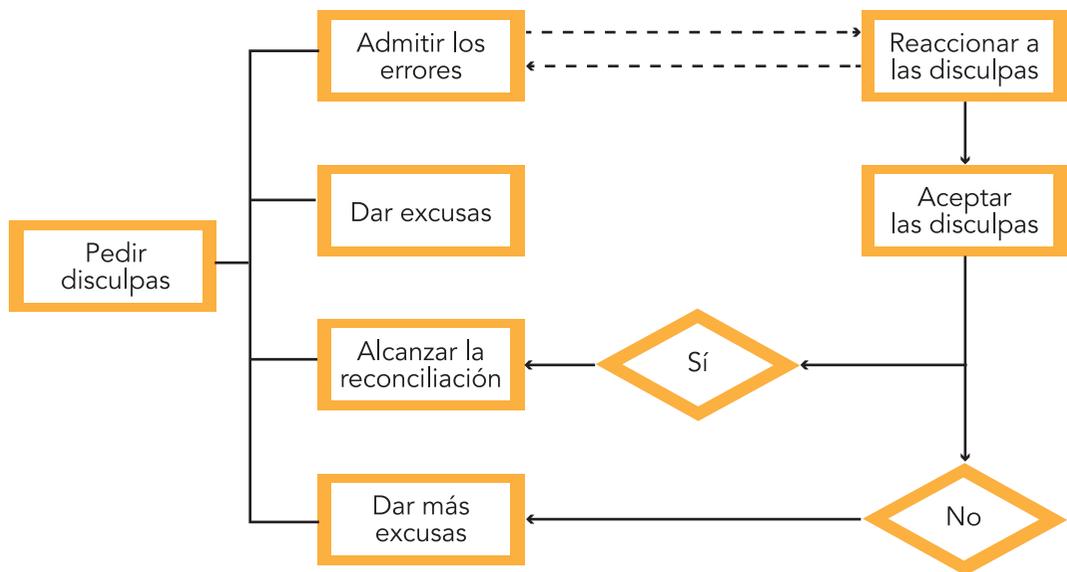


Figura 1. Esquema de la práctica interaccional "pedir disculpas"
(tomado de Sun, 2014)

En la Figura 1, es posible observar que esta interacción no se limita a simplemente decir "discúlpame", para que la otra persona responda en automático: "sí, no hay problema". A veces, se entabla otra dinámica, pues la disculpa podría no ser aceptada por el oyente, y el hablante podría manifestar más excusas e insistir en la disculpa, hasta que se llega a una reconciliación. En otras palabras, el significado se construye cooperativamente en una interacción y hay esquemas de comportamiento esperados en los interlocutores.

Enseñar este tipo de estrategias comunicativas sería parte de la competencia interaccional. Siguiendo con este ejemplo, podrían abordarse los distintos casos hipotéticos (ya que cada cual demandaría y llevaría implícita una dinámica de interacción muy diferente) de cómo pedirle disculpas a tu jefe, a tu amigo, a tu madre y, a su vez, la manera de cómo ofrecer excusas e insistir si es necesario, hasta alcanzar la reconciliación y también viceversa: cómo aceptar o rechazar una disculpa. Una vez que esto se ha enseñado en clase, sería posible evaluarlo en un examen de aprovechamiento.

En resumen, la interacción implica un intercambio de dos o más participantes, donde se opera una transacción (de información, compra-venta, de servicios, de disculpas, etc.), durante la cual el estatus de los participantes es asimétrico y su comportamiento se rige por ciertos esquemas preestablecidos cultural y socialmente, así como por su cercanía o distancia social.

2.2.3 Conversación

Definitivamente, la conversación es un tipo de interacción, en la cual también participan dos o más hablantes. Sin embargo, se distingue de otras prácticas de interacción por su carácter natural y espontáneo; es decir, no hay respuestas esperadas ni esquemas predefinidos, en otras palabras, puede decirse que los participantes en una conversación actúan libremente. Por lo tanto, la conversación se entiende como una interacción comunicativa libre, natural y espontánea, que asume la forma de una secuencia de intervenciones orales. Durante éstas, se involucran interlocutores en un mismo estatus equitativo, sin planear ni organizar de manera previa el intercambio, ninguno de ellos lo controla y su objetivo es la comunicación en sí misma, simplemente platicar entre sí para compartir ideas, como medio de convivencia (Cestero Mancera, 2000, 2005, 2017; Gallardo Páuls, 1996). Muñoz-Basols y Gironzetti (2019) definen la conversación como un género oral dialogado cuyo fin es intercambiar información y reforzar lazos de amistad.

Aunque la participación en una conversación es libre y espontánea, el intercambio no es del todo arbitrario. Se ha establecido que este tipo de interacción presenta ciertos patrones y una estructura reconocible (Sidnell y Stivers, 2013). Por ejemplo, Cestero Mancera (2017) distingue dos unidades primarias: el turno y la alternancia de turnos. Cuando un hablante interviene, se considera como un turno; al tomar la palabra otra persona, se considera como *alternancia*. Los turnos pueden ser turnos de habla, es decir, la intervención de un hablante, o bien turnos de apoyo, que son interjecciones o frases que manifiestan el interés, la sorpresa, la aprobación o desaprobación del interlocutor. Los turnos pueden agruparse en intercambios, y estos a su vez, en *secuencias*. Un conjunto de secuencias de intercambios es lo que se define estructuralmente como una conversación (Cestero Mancera, 2017). La adquisición y consciencia de esta toma de turnos, turnos de apoyo, intercambios y secuencias serían aspectos a desarrollar en la competencia conversacional (Cestero Mancera, 2007).

Dicha competencia no es un tema menor ni sencillo. Así pues, en un examen para evaluar la competencia conversacional, sería injusto evaluar negativamente a un estudiante de español que participó poco, cuando en su cultura, por ejemplo, los turnos de habla son cortos, o quizá no se acostumbra robar turnos (es decir, interrumpir al emisor); más aún si en la clase nunca se le enseñó a conversar.

Valdría la pena considerar que en una conversación típica dentro de la cultura mexicana, es algo común el robo de turnos, a tal grado que muchos temas o historias se interrumpen y nunca llegan a su conclusión. No obstante, hay ciertas maneras de robar turnos que son groseras, y otras que son típicas y frecuentes. Sería injusto, también, evaluar negativamente

una conversación en términos de coherencia, cuando en la práctica cotidiana se puede cambiar de tema y retomarlo más tarde durante toda la secuencia de intercambio. Este tipo de aspectos son los que hay que enseñar en cuanto a la conversación, sin perder de vista y siempre en función de que puedan ser evaluados más adelante. Es erróneo pensar que la competencia conversacional se desarrollará de manera natural en una segunda lengua, puesto que los estudiantes de español como lengua adicional ya han adquirido patrones y estructuras sobre cómo conversar en su idioma y su cultura.

Aunque la conversación es, por su propia naturaleza, libre y espontánea, es posible enseñar las pautas en que se comporta una segunda lengua, para después evaluar estos parámetros. Lo más importante en este punto será definir clara y concretamente el constructo que se busca medir y, llegado el momento de la evaluación, determinar adecuadamente los puntos clave a considerar, para cuantificar y valorar en su justa dimensión el progreso de cada estudiante.

En resumen, el primer paso para elaborar un instrumento de evaluación oral es definir qué se va a evaluar. En el ámbito de la oralidad, se pueden medir la *expresión*, la *interacción* o la *conversación*. Después, se delimita el propósito del examen, el cual puede ser la colocación, el aprovechamiento, la certificación, entre otros. El cómo evaluar tiene que ver con que las tareas a diseñar sean funcionales para elicitación de la expresión, la interacción o la conversación orales, así como los parámetros que se observarán para su justa ponderación. Definitivamente, no es posible diseñar de manera adecuada un examen de lengua oral, hasta no haber definido el constructo que se pretende observar y lo aprendido por los alumnos sobre éstos.

3. Evaluación de la habilidad oral

Típicamente, se habla de una evaluación subjetiva cuando ésta es realizada por un examinador. En contraste con las pruebas de comprensión o de gramática, que pueden resolverse con *ítems*, o reactivos de opción múltiple, que contendrán una opción correcta y otras erróneas, aquí nos encontramos con pruebas de respuesta abierta; es decir, no existe una respuesta única o concreta, así que el evaluador no puede ser reemplazado por una hoja de respuestas o por una computadora, sino que la evaluación depende de su juicio profesional. Con la finalidad de reducir al máximo el papel de la subjetividad del evaluador, para este tipo de pruebas se crean rúbricas de puntuación, de acuerdo con descriptores previamente establecidos, mismos que deben ser claros y bien delimitados, ya que no deben prestarse a múltiples interpretaciones por parte de los evaluadores, sino apuntar hacia una evaluación lo menos dispar posible.

Asimismo, para lograr una reducción en su subjetividad, es necesario que el grupo de expertos que se encargará de la evaluación, pase primero por un curso de formación, de tal manera que la comprensión en la implementación de las rúbricas sea claro y esté estandarizado para todos los participantes. Este tipo de cursos o talleres se llevan a cabo con un grupo de profesores expertos, a partir de las rúbricas y exámenes reales de sustentantes que han participado en una experimentación o prueba piloto. El procedimiento consiste en que una vez explicados y comprendidos los descriptores de las rúbricas, los expertos evalúan las pruebas tal como lo harían en una evaluación real, comparten los resultados y discuten el porqué de un puntaje y otro hasta que todos llegan a un consenso. Es posible apreciar en este proceso que a mayor número de pruebas evaluadas y discutidas, se logrará mayor consenso y, por lo tanto, se conseguirá una estandarización en la evaluación más imparcial y homogénea.

A pesar de lo anteriormente expuesto, hay factores humanos difíciles de controlar (sino es que imposibles), los cuales influyen en la evaluación de un examinador ya preparado y estandarizado. Por ejemplo, un examinador puede presentar agotamiento físico y mental si por alguna razón se encuentra evaluando muchas pruebas en un corto periodo de tiempo o sin descanso; del mismo modo, el examinador que ha evaluado muchas pruebas, puede llegar a sentir un nivel excesivo de confianza en sus habilidades, por lo que, sin darse cuenta, seguramente habrá ocasiones en las que otorgue notas más estrictas o más laxas. Por ello, resulta indispensable que se realicen cursos de estandarización periódicos con la finalidad de que los expertos constaten y equilibren su nivel en cuanto al entendimiento y aplicación de las rúbricas.

Las medidas para contrarrestar la subjetividad pueden ser todavía más estrictas. Una forma de hacerlo es por medio de la *doble corrección*, según el *método de triangulación* propuesto por Allwright (1991), el cual consiste en que dos expertos evalúen la misma prueba, para después realizar precisamente una triangulación sacando la media de las evaluaciones. Por supuesto, también se parte del presupuesto de que las evaluaciones de ambos expertos no diferirían mucho una de la otra. En caso de que así sucediera, se tendría que recurrir a la evaluación de un tercer experto y triangular las dos más próximas, aunque esto se trataría de una situación inusual. De cualquier manera, es recomendable siempre guardar un respaldo de la prueba, es decir, un video o grabación de voz. El banco de pruebas tiene diversos fines académicos y prácticos. Por un lado, gracias a este recurso es posible realizar estudios e investigaciones en diversas ramas del conocimiento; por otro, sirve como *seguro*, tanto para el evaluador, como para quien presenta la prueba: el primero puede escuchar la grabación varias veces hasta estar seguro de que el desempeño del sustentante efectivamente corresponde con la evaluación

emitida; por otro lado, el segundo podría solicitar una revisión, en caso de no estar conforme con su puntaje o nota. Así pues, el respaldo de la prueba sirve como un seguro para ambas partes.

Es importante mencionar que el estado físico y mental del estudiante o candidato, al momento de realizar la prueba, también podría tener un peso en la subjetividad del evaluador. El primer paso al realizar una prueba oral frente a frente, es tranquilizar al sustentante, para motivarlo a que se sienta lo menos presionado posible, y que así su actuación sea lo más cercana a una situación de interacción real y cotidiana, donde se sintiera cómodo y seguro. Asimismo, es destacable que todo experto debe mantener su profesionalismo y no dejarse influir por el estado anímico de la persona que está examinando; aun así, esto no exime a algunos expertos de considerar que, en ciertos casos, el sustentante se encontraba nervioso o en un estado que afectaría su desempeño.

En ese sentido, otra recomendación sería que el examinador no haya sido previamente profesor del sustentante, o que hayan tenido algún contacto anterior; sin embargo, esto podría ser considerado por algunos como una medida innecesaria, ya que ante todo, se debe confiar en el profesionalismo del experto. Bajo esta perspectiva, nuevamente se considera ventajoso el contar con grabaciones o videos del desempeño del sustentante durante la prueba. En cambio, el estado de ánimo de éste al actuar frente a una máquina, una grabadora o una computadora, tiende a ser más neutral que al hacerlo ante una persona, ya que no esperan contar con empatía o respuesta alguna por parte de un objeto; además, parece que se intimidan menos con un objeto que frente a una persona.

Por último (y se menciona ahora justamente por ser, en cierto sentido, lo principal a tomar en cuenta), tenemos las tareas a realizar. Con el propósito de lograr minimizar al máximo la subjetividad en la evaluación, es necesario contar con tareas de evaluación muy objetivas, cuyas pautas a desarrollar sean completamente claras tanto para el sustentante como para el evaluador. De tal suerte, los puntos a desarrollar o las preguntas a responder no deben resultar ambiguos en absoluto. Como ya se ha dicho, los objetivos de la misma (el qué, el cómo y el para qué) deben estar redactados de tal forma que no haya manera de que se preste a diferentes interpretaciones. Así, el evaluador sabrá exactamente qué tipo de respuesta esperar y se verá ceñido por la objetividad de lo esperado. Como se verá más adelante en el capítulo 4, la eficaz elaboración de la tarea cobrará un papel definitivo en el desempeño de quien presente la prueba.

3.2 Los parámetros de evaluación de la habilidad oral

Los parámetros son elementos sustantivos en el proceso de hacer lo más objetiva posible una prueba subjetiva como la que nos ocupa en este momento. En palabras de Bordón (2004: 999):

La única manera de asignar resultados fiables a pruebas abiertas que precisan calificación subjetiva, es que los evaluadores tengan claro el alcance de cada uno de los parámetros de actuación establecidos en las escalas que se adopten. Un estándar uniforme garantiza la objetividad de la calificación y, consecuentemente, la fiabilidad de la misma. No se contará con una herramienta fiable si no se forma a evaluadores que sepan utilizarla, y eso sólo se consigue con la práctica que lleve a la homogeneización de criterios.

Gutiérrez (2009) expone que es necesario hacer una revisión de las competencias descritas en el MCER que debe desarrollar un aprendiente de lengua para comunicarse adecuadamente en el idioma extranjero. Con base en ello, retoma el siguiente cuadro organizado por Vázquez (2000: 17) sobre las competencias comunicativas aplicadas a la expresión oral.

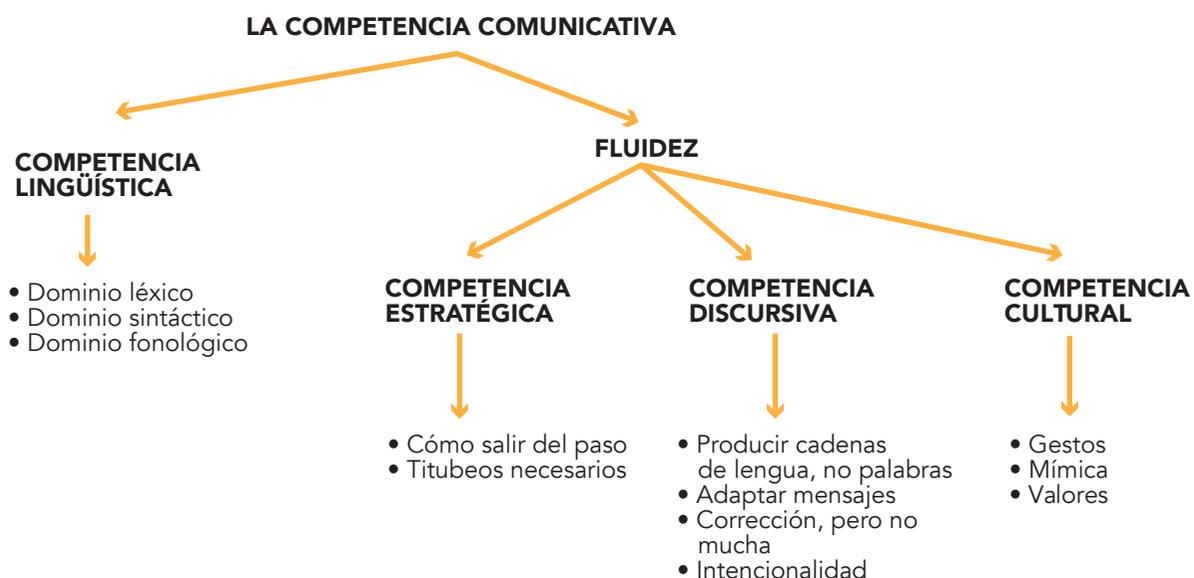


Figura 2. Un modelo de competencia comunicativa (Vázquez, 2000)

Resulta obvio que el desempeño de una prueba de aprovechamiento, depende del dominio del léxico y de la competencia cultural del sustentante, determinados con base en los planteamientos del programa del curso. No obstante, hay aspectos interesantes en este cuadro que podrían considerarse un tanto complicados de evaluar como parámetros fijos en una evaluación, tales como la competencia estratégica o la competencia cultural. Por esto, dentro del ya aludido trabajo de Gutiérrez (2009: 29), se nos deja en claro que, dentro de la enseñanza, no solamente se han de seguir las directrices marcadas por el MCER, sino que es necesario considerar un marco pedagógico, como el de las teorías lingüísticas de Canale (1983), Bachman (1990), o Bachman y Palmer (1996). Entonces, ¿qué parámetros es necesario evaluar como parte de la habilidad oral en una prueba de aprovechamiento para hacer de éste un proceso válido, confiable y viable?

En todo programa de estudios, se consideran aspectos relacionados con el sistema de la lengua y aspectos relacionados con las habilidades comunicativas. Es decir, por parte del sistema de la lengua se distinguen parámetros como gramática, sintaxis y léxico; mientras que en el sistema comunicativo, se ponderan: fluidez, entonación, coherencia, cohesión y dicción. Por otro lado, también se valora lo esperable del sustentante, en cuanto a la resolución de la tarea; en algunas rúbricas, se incluye un parámetro que se denomina como evaluación holística, en otras palabras, la evaluación del conjunto de parámetros en general.

Gramática

En este sentido, no hay mucho que aclarar acerca de uno de los parámetros más populares en la evaluación del aprendizaje de una lengua extranjera. No solamente es considerada en cada una de las actuaciones comunicativas, sino que en muchas ocasiones, esta competencia es evaluada con un examen único. En todo caso, este parámetro es sometido exclusivamente a los contenidos del curso. En el campo de la habilidad oral, en las pautas de la resolución de la tarea no puede haber una función comunicativa como *Contrarrestar una opinión*, si nuestro estudiante es un alumno de nivel A1 del MCER; pero sí se evaluarán las competencias gramaticales necesarias en el cumplimiento de la función comunicativa, como *Presentarse*, por ejemplo. De igual manera, el examinador de la prueba deberá tener presente tanto los contenidos del programa como las rúbricas, que deberán ser lo suficientemente claras como para no permitir que un alumno que domine estructuras de niveles más avanzados, obtenga una evaluación más favorable que alguien cuyo dominio se limita a los temas del programa que están siendo evaluados en esa prueba. Esto llega a ocurrir en pruebas que evalúan el nivel A2 del MCER, donde algunas funciones comunicativas pueden resolverse mediante el uso del presente

de indicativo, pero también con estructuras del presente de subjuntivo. En ocasiones, un candidato que resuelve las pruebas empleando el subjuntivo, llega a tener una evaluación mayor que aquel que responde echando mano de estructuras que no lo requieren. En este caso, la validez de la evaluación no estaría siendo justa para el estudiante que está obteniendo un puntaje menor, si en el programa de estudio no se ha estudiado ese modo verbal que pertenece al nivel B1 del MCER.

Sintaxis

Este parámetro no suele considerarse en muchas evaluaciones, posiblemente debido a las complicaciones para describirlo en la rúbrica. La lengua española se caracteriza por ser un tanto flexible en cuanto al orden de sus oraciones y, en realidad, llega a causar dificultades cuando este orden cambia el sentido de la frase, o cuando ésta llega a ser completamente agramatical; por ello, es considerado en ocasiones dentro de los descriptores de la rúbrica dentro del mismo parámetro de la gramática, o de manera independiente en las pruebas de expresión escrita.

Léxico

Este parámetro encuentra similitudes con el de la gramática, en el sentido de que se encuentra muy sometido a los objetivos del curso que se está evaluando. No obstante, cuenta con una dificultad mayor, ya que el nivel léxico no es medible tan claramente como el gramatical. Siguiendo el ejemplo anteriormente descrito en el punto de gramática, sobre el uso del modo subjuntivo presente, en el léxico no tenemos un bloque exacto de qué términos pertenecen a un nivel o a otro. Actualmente, se cuenta con el referente del Plan Curricular del Instituto Cervantes, el cual es considerado en muchos programas de estudio, principalmente para estandarizar el lenguaje utilizado en los textos escritos y auditivos y que, por tanto, se asume que serán comprendidos y aprendidos de manera explícita o implícita por los estudiantes de un nivel específico del MCER. Sin embargo, al no haber un control exacto de cada uno de esos términos, cabría preguntarse, ¿qué tan justo sería no evaluar de manera más favorable al candidato que hace uso de un léxico considerado por encima del nivel evaluado en la prueba de aprovechamiento? Así es como volvemos a la importancia de la claridad de las rúbricas. Lo más probable es que un caso como éste se resuelva explicitando el grado de dominio, mediante los descriptores: usa términos que pertenecen a otras lenguas; recurre a constantes repeticiones; suficiente para la resolución de la tarea; uso variado y preciso; uso de términos especializados, etc.

Fluidez

La fluidez es uno de los parámetros más claros en la competencia comunicativa oral, ya que en el dominio de esta habilidad, lo que se busca es

que el discurso del candidato sea lo más natural posible. En ese sentido, es cierto que la naturalidad en la velocidad del habla puede variar de una cultura a otra; por lo tanto, aquí es importante señalar que en la comprensión del mensaje hablado, no es necesario que el hablante se desempeñe con una soltura idéntica a la del hablante nativo, sino que basta con que su mensaje pueda ser comprendido sin necesidad de poner en práctica la paciencia del oyente. Es decir, que las pausas al hablar deben ser justificadas y naturales, no constantes y largas, porque es precisamente en esta diferencia donde sale a relucir la falta, ya sea del vocabulario o de las estructuras gramaticales necesarias para desempeñar la función comunicativa solicitada.

Entonación

La entonación es otro parámetro que puede encontrarse envuelto en cierta polémica, debido principalmente a que está estrechamente ligado con la pragmática, otra destreza que rara vez es evaluada en las pruebas de comunicación oral. ¿Qué ocurre cuando un candidato cumple con la tarea solicitada en los demás parámetros sin mucha dificultad, pero la entonación es idéntica en todas sus frases? ¿Afecta esto a su competencia comunicativa en la lengua extranjera? Dicha cuestión deberá ser respondida justificadamente antes de pensar en integrar este parámetro en la evaluación. En nuestro caso, no ha sido considerado para los exámenes de aprovechamiento del CEPE.

Pronunciación

Quizá este parámetro sea más conocido como "acento". Aquí se ha preferido nombrarlo pronunciación, ya que es lo que realmente se estaría evaluando. Al llamarlo "acento", podría ahora ser entendido como qué tan identificable es la lengua materna del aprendiente, de acuerdo con el modo en que se pronuncia el español. Este aspecto ha sido evaluado en muchas academias, incluido el CEPE, como si fuera parte del dominio de la habilidad oral. No obstante, con el paso de los años ha carecido de importancia que la pronunciación del alumno no se acerque a la nativa, y ha sido mucho más relevante el hecho de que todas las palabras pronunciadas sean comprendidas por el oyente. Finalmente, para lograr una comunicación asertiva, lo importante es la comprensión del mensaje y no si el hablante conserva el acento de su lengua materna; además, para muchas personas, mantener su acento significa conservar su identidad nacional o cultural.

Cohesión y coherencia

¿Les ha ocurrido alguna vez que hablan con algún aprendiente de español, y él o ella está efectivamente hablando en esta lengua, pero no se le entiende nada? A eso mismo se refieren estos parámetros, que se relacionan con

el orden de las ideas. En el lenguaje escrito, los ordenadores del discurso se encuentran presentes de manera más explícita y por lo tanto, identificable; pero no se debe olvidar que en la lengua oral también se encuentran presentes, aunque varias veces pasen desapercibidos. En el caso de la habilidad oral, los creadores de las rúbricas deben tener presente que en el ámbito de la cohesión, no se puede obligar a un candidato a hacer uso de ordenadores discursivos muy formales en funciones comunicativas que no lo requieren. En otras palabras, algunas rúbricas otorgan los puntajes más altos al hecho de que los candidatos hagan uso de organizadores más complejos o formales dentro del desempeño de una tarea, como la de *hablar de su rutina diaria*. Si bien en este caso el candidato tendría que hacer uso de conectores y organizadores del tipo: en la mañana/tarde/noche; en primer lugar, después, algunas veces, etc., no será necesario que emplee algunos más formales, tales como 'por un lado', 'no obstante', 'en consecuencia', 'por lo que se refiere a'. Esto le estaría beneficiando al candidato que desarrolle la tarea de manera artificial, con el fin de conseguir un mayor puntaje, y le restaría puntos al candidato que se desempeñe de manera natural. Una vez más, queda clara la necesidad de contar con rúbricas justas, cuyo objetivo se encuentre claramente especificado en cada uno de sus puntos: el qué, cómo y para qué, deben ser los primeros parámetros a considerar en toda evaluación. Tampoco será justo penalizar la ausencia de marcadores en tareas orales en las cuales los hablantes nativos tampoco los usen. Asimismo, será necesario considerar el desarrollo de conectores y marcadores en una segunda lengua; se ha visto que en niveles iniciales los estudiantes usan más marcadores textuales, mientras que en niveles avanzados aparecen marcadores de control de contacto.

Por último, se tendrá que estimar que el hecho de que la tarea sea individual o de interacción, agregará o modificará algunos parámetros. Al llevar a cabo una exposición individual, el parámetro de cohesión y coherencia podría ser más estricto, ya que en todo monólogo hablado se hace un mayor uso de los organizadores, conectores, y las ideas al surgir espontáneamente, también deben ser claras. Por su parte, en una interacción podrían considerarse, además de los parámetros ya descritos, algunos otros como la *interacción en sí*; es decir, que la toma de turnos se efectúe y que además se aporte un estímulo para continuar con el desarrollo de la tarea, de manera que no solamente un candidato se dedique a proponer y el otro a responder positiva o negativamente a dichas proposiciones.

Resolución de la tarea

Las cuestiones a considerar dentro de esta categoría son aquellas que indican el grado en que el sustentante cumple con todas las pautas indicadas en la tarea. Sucede en muchas ocasiones que durante el desarrollo de la misma, tienden a concentrarse mucho en algunos puntos y a omitir otros, ya sea

porque los olvidaron o porque no los entendieron; aquí la comprensión de lo propuesto cobra un aspecto muy relevante. Lo recomendable es asegurarse de que antes de iniciar con la tarea, al sustentante le quede claro lo que debe hacer. En el caso de las pruebas que se realizan por medios digitales, la claridad en la redacción de las instrucciones y de las pautas es lo que asegurará que los sustentantes puedan llevarlas a cabo con éxito y que, tanto si lo consiguen como si no, la evaluación sea justa para todos.

4. Conclusiones

En este capítulo, nos hemos referido a tres grandes temas sobre la evaluación oral: primero, el concepto de la oralidad y sus rasgos generales; segundo, la definición de los constructos asociados a ésta en la evaluación y, tercero, los parámetros más comunes para evaluar la habilidad oral. Se puede decir que hemos establecido la caracterización de lo que es viable tomar en cuenta para evaluar el habla en exámenes de aprovechamiento, por ejemplo. Éste es el primer paso en toda elaboración de instrumentos de evaluación: además de precisar el propósito, se debe definir *explícitamente* los constructos de lo que se va a medir o valorar en la evaluación. Aunque parezca cosa sencilla decir que se evaluará la oralidad, es necesario que todos los examinadores o calificadores tengan claro de manera común desde el mismo concepto de oralidad hasta los parámetros para calificarla. Este capítulo pretende ser una guía en el objetivo de convencionalizar, por decirlo de alguna manera, lo que se entiende por oralidad y sus constructos en un proceso de elaboración y aplicación de instrumentos para evaluar la producción oral.

Con respecto a los constructos de la oralidad, hemos caracterizado la expresión, la interacción y la conversación. Esto es necesario puesto que las tareas o pruebas que se diseñen deben orientarse hacia el constructo que se quiere medir, de tal manera que los elaboradores de los instrumentos de evaluación se avoquen a la creación con base en los constructos, las tablas de evaluación y las directrices del diseño; los profesores se aseguren de incluir estos contenidos en clase y los estudiantes puedan prepararse adecuadamente para el examen. No sería justo evaluar después de un curso habilidades de interacción oral cuando en las clases sólo se ha practicado con presentaciones orales. La definición de los constructos es vital desde las primeras etapas del diseño de los instrumentos.

Finalmente, hemos señalado que las pruebas orales, al ser instrumentos de respuesta abierta, se enfrentan a la subjetividad de los calificadores en su valoración. Para reducir esta subjetividad, es necesario, como mencionamos previamente, estandarizar el juicio del calificador, en otras palabras, todos los calificadores deben tener en común los parámetros

que se evalúan en la oralidad, por ejemplo, la gramática, el léxico, la fluidez, la entonación, la pronunciación, la cohesión y la coherencia y la resolución de la tarea.

Evaluar el aprendizaje es una labor compleja. No se puede garantizar la validez, confiabilidad, justicia y ética de un instrumento de evaluación si en su proceso de elaboración se parte de intuiciones. Es necesario seguir una ruta crítica, empezando por qué se va a evaluar. En el caso de la oralidad, se requiere de este paso fundamental: evaluar la expresión, la interacción o la conversación. Y en cada uno de esos constructos que se puede valorar, qué parámetros son necesarios, con qué profundidad y qué cantidad. Sobre todo, en la evaluación del aprendizaje, es fundamental que los contenidos de la evaluación estén alineados con los contenidos de la enseñanza. Si enseñamos a nuestros estudiantes a expresarse en español, a interactuar con otros y a conversar, entonces lo podemos evaluar en un examen de aprovechamiento.

Referencias

Allwright, D. y Bailey, K. (1991). *Focus on the Language Learner*. Cambridge: Cambridge University Press.

Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, L y Palmer, A. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

Baralo, M. (2000). El desarrollo de la expresión oral en el aula de E/LE. En *Carabela*, 47, s5-36.

Bordón, T. (2004). La evaluación de la expresión oral y de la comprensión auditiva. En Sánchez Lobato, J. & Santos Gargallo, I. (Eds.). *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2) / lengua extranjera (LE)* (pp. 983-1003). Sociedad General Española de Librería.

Canale, M. (1983). From communicative competence to communicative language pedagogy. En Richards, J. C. & Schmidt, R. W. (Eds.). *Language and Communication* (pp. 2-27). Londres: Longman.

Cestero Mancera, A. M. (2000). *El intercambio de turnos de habla en la conversación. Análisis sociolingüístico*. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá.

Cestero Mancera, A. M. (2005). *Conversación y enseñanza de lenguas extranjeras*. Madrid: Arco Libros.

Cestero Mancera, A. M. (2007). La conversación en los programas de enseñanza de lenguas extranjeras. En *Frecuencia L. Revista de Didáctica de Español Lengua Extranjera*, 33, 3-8.

Cestero Mancera, A. M. (2017). La conversación. Bases teóricas y metodológicas para su investigación y su enseñanza en ELE. En *Linred. Lingüística en la Red*, 14. Recuperado de http://www.linred.es/monograficos_pdf/LR_monografico14-articulo1.pdf

Consejo de Europa (2001). *Marco Común de Referencia para las Lenguas*. Ministerio de Educación, Cultura y Deporte & grupo ANAYA. Recuperado de <http://cvc.cervantes/obref/marco>

Cortés-Moreno, M. (2015). Reflexiones en torno al clásico modelo de las cuatro destrezas lingüísticas: propuesta de un nuevo modelo. En *Decires. Revista del Centro de Enseñanza para Extranjeros*, 5 (8), 45-62.

Gallardo Páuls, B. (1996). *Análisis conversacional y pragmática del receptor*. Valencia: Episteme.

Gutiérrez Acosta, R. (2009). *La objetivación de los descriptores de evaluación de la expresión oral en la enseñanza de E/LE a alumnos adultos en inmersión lingüística*. Universidad Nebrija.

He, A. & Young, R. (1998). Language proficiency interviews: a discourse approach. En Young, R. & He, W. (Eds.). *Talking and testing: discourse approaches to the assessment of oral proficiency* (pp. 1-24). John Benjamins.

Jacoby, S. & Ochs, E. (1995). Co-construction: an introduction. En *Research on language and social interaction*, 28 (3), 171-183.

Luoma, S. (2004). *Assessing speaking*. Cambridge: University Press.

Martín Leralta, S. (2019). Destrezas I: comprensión y expresión orales. En Jiménez Calderón, F. & Rufat Sánchez, A. (Eds.). *Manual de formación para profesores de ELE* (pp. 99-120). Sociedad General Española de Librería.

Muñoz-Basols, J. & Gironzetti, E. (2019). Expresión oral. En Muñoz-Basols, J., Gironzetti, E. & Lacorte, M. (Eds.). *The Routledge handbook of Spanish language teaching. Metodologías, contextos y recursos para la enseñanza del español L2* (pp. 198-212). Oxon: Routledge.

Nation, I. S. P. (2011). Second language speaking. En Hinkel, E. (Ed.) *Handbook of research in second language teaching and learning*. Volume II (pp. 444-454). Routledge.

Pinilla Gómez, R. (2004). La expresión oral. En Sánchez Lobato, J. & Santos Gargallo, I. (Eds.). *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2) / lengua extranjera (LE)* (pp. 879-897). Sociedad General Española de Librería.

Sidnell, J. & Stivers, T. (2013). *The handbook of conversation analysis*. Malden: Wiley Blackwell.

Sun, D. (2014). From communicative competence to interactional competence: a new outlook to the teaching of spoken English. En *Journal of Language Teaching and Research*, 5 (5), 1062-1070.

Young, R. (1999). Sociolinguistic approaches to Second Language Acquisition. En *Annual Review of Applied Linguistics*, 19, 105-132.

Young, R. (2000). Interactional competence: challenges for validity. En *Symposium on Interdisciplinary interfaces with language testing*. Simposio llevado a cabo en el Language Testing Research Colloquium, Vancouver, Canadá.

Young, R. F. (2011). Interactional competence in language learning, teaching and testing. En Hinkel, E. (Ed.) *Handbook of research in second language teaching and learning*. Volume II. (pp. 426-443). Routledge.

CAPÍTULO 4

LINEAMIENTOS PARA LA CREACIÓN DE INSTRUMENTOS DE LA EVALUACIÓN ORAL

Klaudia Yvonne **Buzo Hachec**
Moisés **López Olea**

1. Introducción

Hoy en día, existen obras de referencia que ayudan a los profesores en la elaboración de instrumentos para evaluar el aprendizaje de manera adecuada. De acuerdo con los objetivos de la evaluación, los docentes podrían recurrir a instrumentos y estrategias tanto cuantitativas como cualitativas, o una a combinación de ambas. Por ejemplo: exámenes objetivo, exámenes tipo *quiz*, listas de cotejo, rúbricas, portafolios, etc. (cf. Sánchez Mendiola y Martínez González, 2020). Todo profesor de educación básica, media superior y superior podría implementar estos materiales para elaborar instrumentos de evaluación adecuados.

El ámbito de la enseñanza de español como lengua extranjera, por supuesto, no es la excepción. Como parte del proceso de elaboración de instrumentos de evaluación oral en ELE, hay que considerar varios aspectos para garantizar la validez, confiabilidad y justicia de los mismos. En primer lugar, es necesario definir el propósito de la evaluación y los constructos que serán valorados; en segundo lugar, la tabla de especificaciones y, finalmente, la rúbrica de calificación. El diseño y la creación de instrumentos de evaluación oral no difiere de esta ruta general para desarrollar evaluaciones del aprendizaje.

En el capítulo anterior (Buzo Hachec y López Olea, este volumen), presentamos reflexiones de carácter teórico sobre la oralidad, sus rasgos principales y los constructos que pueden ser evaluados en el contexto de la enseñanza del español a extranjeros. También comentamos acerca de los parámetros que comúnmente se califican en muestras orales de lengua. En el presente capítulo haremos un recorrido práctico, paso por paso, para seguir los lineamientos fundamentales que permitan construir instrumentos de evaluación oral adecuados, válidos, confiables y justos, orientados en todo momento a la creación de instrumentos de evaluación oral en español como lengua extranjera. Particularmente, orientaremos un ejemplo de prueba oral hacia la evaluación del aprendizaje o del aprovechamiento,

que consiste en evaluar el aprendizaje al final de un curso (De Santiago y Fernández, 2017).

En primer lugar, trataremos acerca de las características generales de los instrumentos de evaluación oral. Después, paso por paso, describiremos y ejemplificaremos los lineamientos para desarrollar un instrumento de esta índole: desde la planeación y los objetivos, la definición y selección de los constructos que se medirán en el instrumento en cuestión; así como también, el modo en que se elabora una tabla de especificaciones y, a partir de ellas, cómo se realizan pruebas, tareas y rúbricas de calificación. Finalmente, expondremos sobre la experimentación de los instrumentos diseñados. En todos los casos, mostraremos un ejemplo para ilustrar estos lineamientos. Consideramos que seguir estos pasos permite garantizar que la evaluación oral en cuestión sea válida, confiable, justa y ética.

2. Características de los instrumentos para la evaluación del aprovechamiento de la oralidad

En esta sección, enlistaremos una serie de características que, consideramos, tienen en común todos los instrumentos de evaluación oral. Principalmente, ponderamos la evaluación del aprendizaje, es decir, aquella que se lleva a cabo al final de un curso de ELE, o como es conocida: la evaluación del aprovechamiento.

- Los exámenes orales de aprovechamiento pueden realizarse de manera presencial, cara a cara, con o sin grabación, pero también es factible llevarlos a cabo mediante una computadora.
- Dichas pruebas pueden ser audio o videograbadas. Aunque esto no es obligatorio, se trata de una característica a considerar; de hecho, se recomienda implementarla, dado que beneficia la calificación y también la posibilidad de revisar el examen la veces que sean necesarias.
- Las tareas a desarrollar en el examen deben estar basadas en situaciones que se llevaron a cabo durante el curso a evaluar. En otras palabras, no es conveniente evaluar a los alumnos con tareas que no se abordaron como parte de su práctica en el salón de clases. Por ejemplo, si se piensa evaluar la interacción con juegos de rol, será necesario que en el curso se hayan realizado este tipo de actividades.
- Las pruebas podrán llevarse a cabo de manera individual o bien en parejas. Esto dependerá de las necesidades, circunstancias y tiempo de evaluación, así como de los estudiantes y del centro o instituto de lenguas. La decisión debe estar fundamentada en el constructo que se evaluará y en la tabla de especificaciones. Si se evalúa únicamente la expresión, con la presencia individual del estudiante será suficiente; pero si se evalúan otros constructos, tales como

interacción o conversación, podría ser pertinente que hubiera más de un estudiante en la situación de examen.

- Las situaciones a desarrollar deberán ser lo más cercanas posible a las que se espera que el alumno enfrente en la vida real, siempre y cuando hayan sido parte del contenido de la enseñanza en el aula.
- Al ser la oralidad el objeto de medida, hay que considerar que sus rasgos pueden estar presentes en las muestras de lengua de los estudiantes: pausas, muletillas, repeticiones, reformulaciones, reparaciones, entre otros aspectos comunes en la oralidad. Al ser rasgos propios de esta modalidad de la lengua, no tendrían por qué ser penalizados o calificados negativamente ante su expresión natural.

Una vez que tenemos en mente todas estas características, podemos proceder a la elaboración, paso a paso, de un instrumento de evaluación oral. Para esto, en la siguiente sección, describiremos el proceso de elaboración de pruebas orales a través de cinco lineamientos que seguimos.

3. Lineamientos para elaborar un instrumento de evaluación oral

Para elaborar un instrumento de evaluación oral, proponemos observar los siguientes lineamientos: planeación y objetivos, definición de constructos, elaboración de tareas y pruebas, diseño de la rúbrica de calificación y, por último, validación y experimentación. A continuación, desarrollamos y ejemplificamos cada uno de estos pasos, con una prueba oral ficticia de final de curso.

3.1. Planeación y objetivos

Bordón (2000) enlista una serie de requerimientos a tomar en cuenta para confeccionar un examen de expresión oral de alta calidad. De acuerdo a la autora, los pasos a seguir son los siguientes: determinar el tipo de examen, especificar los objetivos, definir los contenidos (áreas temáticas; nociones y funciones pragmáticas; exponentes lingüísticos, subdivididos en gramática y vocabulario). Asimismo, establecer los procedimientos: distribución de las partes, tipos de pruebas, duración; enseguida, definir criterios de evaluación y experimentar el examen; para después comprobar la validez de las pruebas y la fiabilidad de los resultados; y finalmente, recapacitar para obtener retroalimentación.

En este capítulo, como se ha mencionado, en el primer paso nos centraremos en exámenes de aprovechamiento, para determinar el tipo de prueba necesaria. Para avanzar con el segundo paso, se tendrán que definir los objetivos: qué se necesita evaluar, cómo y para qué. Estos tres

parámetros serán nuestras bases para más tarde desarrollar la definición de los constructos del examen.

En primer lugar, el “qué” debe responder a los contenidos del curso a evaluar. No obstante, cabe señalar que no se retomará cada uno de ellos, sino sólo sus contenidos mínimos; es decir, no se evalúa la totalidad de temas y contenidos de un curso, sino sólo aquellos elementos que los expertos en el tema consideran necesarios e imprescindibles para desenvolverse de manera adecuada en el siguiente curso. Los contenidos de la evaluación deberán ser claramente definidos, de manera que no haya posibilidad de malas interpretaciones de los mismos, por parte de los examinadores o de los sustentantes del examen. Además, deberán responder no solamente a los contenidos del programa escolar, sino también a los objetivos y procedimientos adoptados en el aula (Bordón, 2000, 159). Aquí se muestra un ejemplo del proceso de definición de los objetivos, tomando como muestra el nivel B1, de acuerdo al MCER.

Ejemplo de planeación y objetivos

- **Constructo a evaluar:** expresión oral
- **Propósito:** aprovechamiento
- **¿Por qué?** Porque se ha practicado la expresión oral en presentaciones durante el curso.
- **Nivel:** B1 de acuerdo al MCER
- **Objetivo:** Evaluar la expresión oral en una presentación académica sobre un tema familiar, de interés personal, para describir experiencias, acontecimientos, deseos y aspiraciones y justificar brevemente opiniones o explicar planes.

3.2. Definición de constructos

Una vez que se haya determinado cuál es el objetivo de la evaluación, habrá que decidir cuál será el objeto de medida que se va a evaluar. En otras palabras, es necesario definir qué aspecto de la oralidad se va a medir en un contexto dado (Luoma, 2004; Bordón, 2000). No es adecuado pedir que los estudiantes *hablen por hablar*, con el fin de evaluarlos. Esta etapa tiene que ver con la definición de constructos que ya se discutieron en el capítulo 3 (Buzo Hachec y López Olea, este volumen); es decir, habrá que tomar la decisión sobre evaluar la *expresión*, la *interacción* o la *conversación*.

Para tomar las decisiones con respecto a qué evaluar de la oralidad, habrá que apegarse a documentos de referencia. Para el caso de la evaluación del aprovechamiento, este documento tiene que ser el programa de un curso, porque los contenidos de la evaluación deben estar alineados con los contenidos de la enseñanza. Si durante el curso se practicó la

expresión, por ejemplo, mediante presentaciones orales, la evaluación podría enfocarse en este constructo; o bien, si se enseñó la interacción o se practicó la conversación (su estructura y sus partes), cualquiera de ambos será el constructo a evaluar.

La definición del constructo es importante, porque determina en gran medida el diseño de las tareas del examen de aprovechamiento. Por ejemplo, si durante un curso sólo se practicaron géneros monológicos, el examen oral no necesitará de dos estudiantes, sino solamente de uno. En cambio, si se desarrollaron la competencia interaccional o conversacional, en el diseño del examen podríamos pensar que sería conveniente tener dos estudiantes en la misma situación de examen, para que interactúen. No es suficiente con incluir actividades en las que los estudiantes hablen entre sí: es necesario enseñarles a interactuar y a conversar para que realmente se pueda justificar la elección de estos constructos. Si se enseña a conversar, por ejemplo, es factible evaluar acciones como mantener un turno o robarlo, usar marcas de control de contacto, entre otras estrategias.

En resumen, la definición de los constructos para un examen oral que evalúa el aprovechamiento, debe relacionarse directamente con aquello que fue enseñado en un curso. Se puede optar por evaluar la *expresión* de los estudiantes, la *interacción* o la *conversación*. En todo momento, el examen debería medir qué aprendieron los alumnos sobre estos contenidos durante el curso.

Para ejemplificar la elección de un constructo para evaluar, continuaremos con el desarrollo de un instrumento de evaluación oral. En la sección anterior, se definió que el propósito de la evaluación es determinar el grado de aprovechamiento, es decir, la evaluación del aprendizaje después de un curso; éste sería, hipotéticamente, de nivel B1 y han habido presentaciones orales de temas diversos ante el grupo, con sus respectivas evaluaciones.

Dado que se trata de un contenido del curso, se optará por evaluar presentaciones orales en el examen final. En este sentido, una presentación oral es un género más monológico. Aunque al final de cada presentación haya preguntas, en realidad se trata de la preparación de un tema para su exposición durante un tiempo determinado. En este caso, el constructo que se va a evaluar no es *interacción* ni *conversación*, sino *expresión*.

Actualmente, la presentación oral se considera un género discursivo de carácter académico y ha sido descrito de manera adecuada por diversos autores en el contexto de la enseñanza del español como lengua extranjera (Pastor Cesteros y Ferreira Cabrera, 2018; Montoya Magno y Pérez Díaz, 2020; Villar, 2011, 2014; Vergara Padilla y Ainciburu, 2016). Villar

(2011) comenta los rasgos generales de las presentaciones orales, también llamadas ponencias o exposiciones:

- a) La situación comunicativa es institucional, académica o universitaria. Para los efectos de la evaluación que estamos desarrollando es compatible, primero porque ha sido contenido de la enseñanza, y segundo, porque el ámbito es propiamente el salón de clase.
- b) El emisor es el estudiante. Esto es útil para la evaluación del aprendizaje, puesto que es una tarea que permite elicitar lengua oral de los estudiantes que vamos a evaluar.
- c) La audiencia es el profesor y sus compañeros de clase. Durante el curso se han llevado a cabo presentaciones orales frente a grupo; este puede ser el caso para la presentación final, objeto de evaluación del instrumento que estamos desarrollando; o bien, una situación de examen individual ante el profesor.
- d) Los roles sociales son asimétricos. Hay un evaluador y un evaluado, un profesor que califica y estudiantes que presentan la prueba. Esto es compatible también para el instrumento oral con el que estamos ejemplificando.
- e) El mensaje es un contenido disciplinar, académico o de divulgación. Con respecto al tema, para B1 ya es posible optar por temas de este tipo.
- f) El carácter es planificado y monológico, pudiendo dialogar con la audiencia al final. Durante el curso, los estudiantes han tenido tiempo previo de preparar su presentación. Para la evaluación del aprendizaje se puede optar por dos opciones: una exposición planeada ante grupo o en situación de examen individual ante el profesor. En cualquier caso, habrá que dar tiempo de planeación, ya que es característico de este género. Por otro lado, se recomienda determinar un tiempo mínimo y uno máximo para la exposición oral, de tal forma que el monólogo esté delimitado por dicho plazo.
- g) El registro es formal y el léxico preciso. Para el nivel B1 del curso, y los contenidos de la enseñanza, esto se puede adecuar, pero durante el curso se recomienda enseñar a presentar y manejar un registro formal sin perder de vista el nivel de dominio.
- h) Procedimientos: exponer, informar, argumentar, persuadir. Estas actividades de la lengua han de considerarse en las pautas de las instrucciones de la tarea. Para el nivel B1 del curso, y de acuerdo con los contenidos de la enseñanza, pueden integrarse o adaptarse.
- i) La función curricular es evaluativa en el marco de la asignatura o programa del curso. Esta característica de las presentaciones orales es altamente compatible con el instrumento de ejemplo que estamos desarrollando. Al haber enseñado cómo exponer en español, y practicado varias veces durante el curso, el último

examen oral no será ajeno a los estudiantes, ya que habrán recibido retroalimentación y esto reduciría el nerviosismo y estrés característicos de todo examen de aprovechamiento.

Como se ha visto, la elección de una presentación oral para la evaluación del aprendizaje es adecuada. Primero, porque ha sido contenido de la enseñanza, y segundo, porque es un constructo bien definido como género. Es innegable que la exposición oral es una estrategia de evaluación bastante común y utilizada, tanto para la educación básica, media y superior en México (Montoya Magno y Pérez Díaz, 2020) como en el ámbito de la adquisición de español como segunda lengua.

En resumen, la elección del constructo para nuestro instrumento de ejemplo de la evaluación del aprendizaje es, de manera general, la expresión oral, y, de manera particular, la presentación académica oral. Esto implica cierto diseño en las tablas de especificaciones, así como determinado tipo de instrucciones y pautas en la prueba. Generalmente, cuando se ha establecido el propósito de la evaluación, es necesario seleccionar el objeto de medida, es decir, el constructo de la oralidad que se va a evaluar: pueden ser géneros monológicos, dialógicos o conversacionales. Lo importante es definir también ese objeto de medida, con respecto a qué es y cuáles son sus características, para determinar de manera precisa qué aspectos se van a valorar de éste; lo anterior ha quedado ejemplificado con el caso de nuestro instrumento de evaluación oral. En la siguiente sección, se abordará cómo elaborar una tabla de especificaciones que permita a cualquier profesor del mismo curso construir un instrumento de evaluación similar para todos sus estudiantes del mismo nivel.

3.3. Elaboración de tablas de especificaciones

Una vez delimitados los objetivos y el constructo del examen, se deberá responder al cómo y al para qué evaluar esos objetivos. Estas tres preguntas serán las que conformen la tabla de especificaciones, con el fin de crear tareas con pautas claras y específicas.

En la tabla de especificaciones, se deberán considerar, además de los objetivos y los constructos: el tipo de tarea, las pautas y los medios o *inputs* (imágenes, textos de entrada como anuncios, noticias, textos cortos...; audios, etc.). De igual manera, “el cómo” tendrá que ser justificado, y lo suficientemente claro para que a todos los sustentantes se les asignen tareas que evalúen las mismas muestras de lengua y que, además, contengan instrucciones claras; de tal manera que no tengan necesidad de preguntar al examinador qué deben hacer exactamente, debido a que los lineamientos se presten a diferentes interpretaciones.

Tanto los creadores de tareas, como los examinadores y los examinados, deben poder comprender qué es lo que los sustentantes deben hacer; qué se espera de su desempeño y cómo deben realizarlo.

Con la finalidad de asegurar que nuestros objetivos y el cómo evaluar están bien justificados, se responde a la pregunta *para qué*; es decir, qué muestra de lengua esperamos obtener de nuestros candidatos, ya que dicha muestra será la que los examinadores deberán evaluar, con base en rúbricas específicas. En la enseñanza de ELE, el *para qué* tiende a responder a las funciones comunicativas que se espera que el candidato lleve a cabo, con base en los objetivos y los medios de evaluación. Una vez que se tienen resueltas estas interrogantes, la tabla de especificaciones será la guía a partir de la cual se desarrolle una variedad de tareas distintas, pero de las cuales se puedan obtener muestras de lengua similares, para que todos los examinados sean evaluados de igual manera.

Ejemplo de elaboración de tablas de especificaciones

Capacidad que se evalúa	Tarea que se realiza	Características del estímulo
Presentar oralmente información sobre un tema familiar o de interés personal.	Desarrollar un monólogo de mínimo 7 y máximo diez minutos de un tema del ámbito personal, público o académico; se deben describir experiencias, acontecimientos, deseos y aspiraciones, así como justificar brevemente sus opiniones y explicar sus planes.	Una instrucción general con seis pautas: dos fijas para introducir y concluir, y cuatro para, respectivamente, describir, comparar, opinar y explicar; todo en una presentación oral ante el grupo.

3.4. Elaboración de pruebas y tareas

El uso de la lengua oral varía de acuerdo con el contexto y el propósito, por ello, el diseño de las tareas de un examen es una etapa muy importante en el desarrollo de evaluaciones orales (Luoma, 2004). Para diseñar una tarea o una prueba, es necesario contar con las tablas de especificaciones. De lo contrario, podría suceder que cada profesor diseñe y evalúe aspectos distintos de la oralidad para el mismo curso, o que cada uno elabore diferentes pruebas, con distintas instrucciones, lo que resultaría en la obtención de muestras de habla dispares, a partir de las cuales

sería difícil calificar de manera objetiva. En nuestro ejemplo, es necesario contar con un instrumento de evaluación estándar u homologado, puesto que se trata de la evaluación del aprendizaje; así que todos los estudiantes deben ser evaluados de manera similar, para así determinar si aprueban o no el curso, y qué tanto han aprendido de la oralidad en español.

En el ejemplo del instrumento de evaluación que estamos elaborando, ya hemos definido el propósito, que es ponderar el aprovechamiento al final de un curso de nivel B1; también, el constructo del objeto de medida de la evaluación, que es la expresión oral de exposiciones académicas; asimismo, presentamos la tabla de especificaciones, que establece la capacidad a evaluar, cómo se evalúa y con qué materiales. En este punto, hemos llegado a la etapa del diseño de la prueba oral. En nuestro ejemplo, se trata sólo de una prueba para evaluar la oralidad, pero en otros cursos o en otras evaluaciones, una sola prueba oral podría contener dos o tres tareas, en un protocolo de aplicación distinto.

Considerando la tabla de especificaciones para nuestro ejemplo, sería adecuado crear la instrucción y las pautas que le permitan al estudiante saber exactamente qué es lo que tiene que hacer durante su evaluación. El material que éste ve para llevar a cabo la prueba, se compone de una instrucción y las pautas determinadas en las especificaciones, así como todas las indicaciones necesarias para que cumpla satisfactoriamente su presentación oral. Con base en el propósito, el constructo y la tabla de especificaciones, las instrucciones y las pautas podrían presentarse como se muestra a continuación.

Ejemplo de instrucciones

Evaluación oral final: presentación oral (7-10 minutos)

Tema: pasatiempos

Instrucción: elige un tema sobre **tus pasatiempos** y elabora una presentación oral con imágenes. En tu presentación tienes que:

- Introducir el tema;
- describir **el pasatiempo de tu elección (cómo es, cómo y con quién lo haces, con qué frecuencia, etc.)**;
- comparar **cómo ha cambiado el pasatiempo en tu vida (cómo era antes y cómo ahora, qué cambios ha tenido, etc.)**;
- opinar **sobre lo positivo y lo negativo del pasatiempo**;
- explicar **cuáles son tus planes en el futuro con este pasatiempo**;
- concluir.

A los elaboradores se les muestran las instrucciones con partes en color rojo. Estas partes pueden variar de prueba a prueba, es decir, el tema

puede cambiar, pero tiene que estar dentro del nivel de B1; del mismo modo, es viable modificar también las instrucciones, pero debe estar presente la descripción, la comparación, la opinión y la explicación, así como dos pautas de introducción y conclusión. De esta manera, hay una correspondencia con la tabla de especificaciones, con el contenido que se desea medir y, aunque la tarea cambie de tema, en esencia seguirá siendo la misma evaluación y la misma prueba para todos los estudiantes que terminen el curso de B1. En otras palabras, se garantiza la ética y la justicia al valorar de la misma forma el aprovechamiento para todos los estudiantes al final del curso.

En resumen, la prueba se diseña en correspondencia con la tabla de especificaciones, en la cual se indica si se elabora una tarea o más. Esto, considerando tanto los contenidos a evaluar, como el tiempo disponible para hacer el examen o la tarea de evaluación.

Las instrucciones y las pautas deben ser claras y contener léxico que no esté por encima del nivel del estudiante o del curso que se quiere evaluar, para asegurar que la realización de la tarea no sea impedida por la falta de comprensión de lo que el sustentante tiene que hacer en el examen (López Olea, 2020; López Olea y Nieto Rodríguez, 2019). Finalmente, debe decidirse el protocolo de aplicación, es decir: si en la situación de evaluación habrá dos profesores examinadores, o sólo uno; asimismo, si participarán al mismo tiempo dos estudiantes o la prueba será de modo individual. En el caso de nuestro ejemplo, se trata de un género monológico, por lo que en la situación de evaluación, puede estar solamente un sustentante.

Cuando se llega a esta etapa, el diseño de la prueba está listo para ser aplicado. Pero todavía no termina el proceso. Una vez que se obtiene la muestra de lengua oral, hay que decidir cómo valorarla, medirla o calificarla. No sería adecuado hacerlo desde las intuiciones e ideas subjetivas de un profesor, sino que se requiere de un instrumento que permita evaluar esa muestra de habla de la manera más objetiva posible. Para ello, se diseñará también una rúbrica de evaluación. En la siguiente sección, abordamos este importante paso.

3.5 Elaboración de rúbricas

Evaluar una muestra de expresión oral no es tarea sencilla. Si bien la tabla de especificaciones y la planeación de las tareas son pasos primordiales, la evaluación siempre se hallará sobre una base subjetiva, ya que la lengua oral es un objeto muy flexible y eso dificulta, tanto la medición de su dominio, como del proceso de aprendizaje de los alumnos. En palabras de Luoma (2004: 59): "ni siquiera es certero saber qué tan posible es encontrar caminos de aprendizaje claros que siguen la mayoría de los

alumnos” (traducción propia). Además, en cuanto al desarrollo de escalas para las rúbricas de la expresión, Luoma (2004: 59) continúa con una cita de Northon (1996), quien menciona que durante esta tarea se trata de describir un fenómeno complejo con un pequeño número de palabras sobre las bases de una teoría incompleta (traducción propia).

No obstante, existen rubros más o menos establecidos para medir la expresión oral en los exámenes de aprovechamiento. Lo primero que se debe plantear es el tipo de rúbrica adecuada para la evaluación: la *holística* proporciona un sólo puntaje basado en la valoración integral del desempeño del alumno, mientras que la *analítica* arroja un puntaje diferente por cada criterio, desglosando los componentes de la rúbrica (González, Sierra y Sosa, 2020). Una vez definido el tipo de rúbrica, se debe cuestionar qué es lo más importante a medir en la expresión oral. Como se recordará, en el capítulo 3 de este libro, *La habilidad oral y su evaluación*, se hace mención de los rubros que más comúnmente se consideran en las rúbricas; a saber: gramática, léxico, fluidez, entonación, pronunciación, cohesión, coherencia y resolución de la tarea. Si bien, estos son los rubros más comunes en este tipo de rúbricas, se debe tener siempre presente que todos ellos cubran con lo practicado durante el curso, ya que medir el desempeño de los alumnos con rubros que no fueron puestos en práctica, demeritaría la justicia de la evaluación.

Una vez determinados los rubros, se deben planear el tipo de descriptores y el número de bandas que habrá en la rúbrica. No es recomendable desarrollar más de cuatro escalas, ya que para el momento de llevar a cabo la evaluación, los examinadores ya deberán haber estudiado las rúbricas y tendrían que conocerlas y entenderlas. Aun cuando los examinadores cuenten con la rúbrica para su consulta inmediata, se debe saber a detalle qué es lo que buscan evaluar en los alumnos para centrar su atención, tanto en esos rubros, como en el nivel de medición.

Otra forma de facilitar la comprensión de los rubros y de la escala de medición para los examinadores y alumnos (se considera importante dar a conocer también a los alumnos exactamente qué se espera de ellos y cómo serán evaluados), es la de agrupar los diferentes parámetros en un sólo rubro; por ejemplo, se puede desarrollar un campo de *Uso de la lengua*, en donde se midan todos los aspectos relacionados con la misma: gramática, léxico, cohesión, etc. Y en la escala de medición se describiría el grado de desempeño alcanzado en la muestra de lengua. Asimismo, se podría considerar un segundo rubro, el de *Holística o Resolución de la tarea*, y describir también el grado de cumplimiento del mismo en la escala de medición, tal como se aprecia en el ejemplo de rúbrica mostrado a continuación. Esto facilita el uso de las rúbricas y el análisis de la muestra, por parte de los examinadores y de los alumnos.

Ejemplo de elaboración de rúbricas

Rúbrica para evaluar una presentación oral al final de un curso de B1				
Rubro	Poco 0	Suficiente 1	Bueno 2	Experto 3
Uso de la lengua	Puede presentar con una pronunciación y entonación apenas comprensibles; usa gramática para describir, opinar y explicar con muchos errores que dificultan la comprensión; emplea léxico escaso, sin conectores y/o frases memorizadas durante su presentación.	Puede presentar con una pronunciación y entonación que interfieren poco en la comprensión; usa gramática para describir, opinar y explicar con varios errores que podrían afectar la comprensión; emplea poco léxico y conectores básicos con repeticiones durante su presentación.	Presenta con una pronunciación y entonación adecuadas; usa gramática adecuada para describir, opinar y explicar con pocos errores que podrían afectar la comprensión; emplea una variedad limitada de léxico y conectores durante su presentación.	Presenta con una buena pronunciación y entonación; usa gramática adecuada para describir, opinar y explicar con errores mínimos; emplea una amplia variedad de léxico y conectores durante su presentación.
Cumplimiento de la tarea	Cumple con una o dos de las pautas principales; pueden faltar la introducción y la conclusión en su presentación oral, ofreciendo información escasa.	Cumple con algunas de las pautas principales; pueden faltar la introducción y la conclusión en su presentación oral, ofreciendo información apenas adecuada.	Cumple con la mayoría de las pautas; pueden faltar la introducción, la conclusión y una de las pautas principales en su presentación oral, ofreciendo información adecuada.	Cumple con todas las pautas en su presentación oral, ofreciendo información extensa con ejemplos.
Formato de la presentación	Faltan elementos visuales en una presentación con un formato poco ordenado y visualmente deficiente.	Apenas incluye elementos visuales en su presentación; el formato puede ser ordenado y visualmente, poco amigable.	Incluye elementos visuales adecuados, en una presentación con un formato ordenado y visualmente amigable.	Incluye una amplia variedad de elementos visuales, en una presentación con un formato ordenado y visualmente amigable.

Este ejemplo de rúbrica puede ajustarse de acuerdo con las necesidades del curso, de los estudiantes y de cada profesor. Otro ejemplo de rúbrica para evaluar la oralidad en una segunda lengua, se puede encontrar en Hidri (2018: 43), en el cual se incluye el logro de la tarea, la comunicación interactiva, la fluidez, la gramática, el vocabulario y la fonología. La inclusión independiente de los rubros o su integración en una categoría como “uso de la lengua” depende de las necesidades de evaluación, del constructo a evaluar, de las tablas de especificaciones y de las decisiones de los expertos que elaboran los materiales de evaluación. En el proceso de validación, es posible considerar si es práctico y útil dejar los rubros separados, o incluirlos en una sola categoría. De cualquier manera, la rúbrica elaborada debe ser la misma para la evaluación de la oralidad de todos los estudiantes. En todo caso, también la eficacia de las rúbricas debe ser sometida a validación por parte de un grupo de expertos, así como a experimentación, la cual puede realizarse por medio de una lista de cotejo, como se muestra en el ejemplo de la siguiente sección.

3.6. Validación y experimentación

Una vez que se han elaborado las tablas de especificaciones, las tareas y las rúbricas, es necesario llevar a cabo una prueba de validación y otra de experimentación, antes de dar por concluido el proceso de elaboración del examen final del curso. Es vital verificar que el constructo y el diseño de la tabla de especificaciones son adecuados, y que las tareas efectivamente eliciten las muestras evaluables para medir el aprovechamiento de los estudiantes; asimismo, que las rúbricas ayudan a calificar de manera adecuada. Chappelle (2011) relaciona el proceso de validación de pruebas de evaluación con la obtención de evidencia sobre el significado de los puntajes de la evaluación, y esto a su vez, con la adecuación de la toma de decisiones en el proceso de elaboración del examen. En otras palabras, habrá que verificar que el examen o la prueba mida efectivamente lo que dice medir y que lo califique de manera adecuada.

En primer lugar, el equipo de elaboradores de los documentos técnicos (como la tabla de especificaciones, las rúbricas y las tareas) debe ser diferente al equipo de validación. Un grupo de expertos, en este caso de enseñanza de español como segunda lengua, revisa los documentos con base en listas de cotejo y hace observaciones sobre ellos. Sería ideal que su área de especialidad sea la habilidad oral. Estas observaciones se envían a los elaboradores y se hacen las adecuaciones pertinentes. Es recomendable que este proceso se realice justamente después de terminar con la elaboración de las tablas, las pruebas y tareas, así como las rúbricas.

En segundo lugar, la fase de experimentación se refiere a uno o varios pilotajes. Este podría hacerse con estudiantes extranjeros voluntarios y también con hablantes nativos. Es necesario aplicar las tareas y asimismo calificarlas, de tal manera que pueda observarse que las tareas diseñadas están elicitando propiamente las muestras de lengua que se busca evaluar, así como los contenidos de la evaluación; además, que el proceso de calificación refleja el aprovechamiento de los estudiantes. De igual modo, es posible analizar si la calificación de diferentes examinadores varía mucho; es decir, que uno califique muy bajo y otro muy alto, por ejemplo, o si se alcanzan puntajes similares con la misma prueba.

El pilotaje con nativos es útil, ya que en los estudios sobre interlengua, suele ser común la metodología de contrastar el discurso nativo con el no nativo (Barrón, 2003; Fung y Carter, 2007; López Olea, 2017). Esto no quiere decir que se espere que los estudiantes extranjeros alcancen la competencia nativa, sino que es útil para contar con un referente de aquello que resulta esperable en el nivel nativo. Por ejemplo, si en una tarea se quisiera evaluar la capacidad de los estudiantes para hacer peticiones, pero ni siquiera el grupo de nativos hace peticiones al resolver la tarea, quiere decir que este criterio debe modificarse. Otro ejemplo sería el uso de marcadores discursivos; si al resolver la tarea los estudiantes extranjeros usan pocos marcadores, y por esa razón se califica más bajo, tendría que haber el referente de que, cuando menos, un nativo efectivamente usó más marcadores en esa tarea. En resumen, pilotear o experimentar las tareas orales en grupos de nativos, puede darle a la fase de experimentación un referente de aquello que podría esperarse en la tarea, con respecto a un estándar nativo. Esto no debe afectar negativamente la calificación de estudiantes con niveles básicos o intermedios de español, más bien debería ayudar a determinar cómo valorar el uso (o falta del mismo) de aspectos lingüísticos que ayuden a calificar de manera más justa.

Para llevar a cabo el proceso de validación de los documentos creados para la evaluación oral, se podrían hacer análisis psicométricos complejos, tal como propone Hidri (2018), para determinar qué tan objetivos son los calificadores al usar una rúbrica de evaluación oral. Si no se contara con los recursos y las posibilidades para hacer un análisis tan complejo, al menos podría emplearse una lista de cotejo como la siguiente.

Ejemplo de validación

Lista de cotejo para validar la tabla de especificaciones, prueba oral y rúbrica			
	NUNCA	A VECES	SIEMPRE
La tabla de especificaciones es clara y permite elaborar distintos modelos de tarea que evalúan el mismo contenido y la misma capacidad oral del estudiante.			
Las instrucciones son claras para todos los estudiantes y elicitan una muestra de lengua evaluable que los guía para elaborar la tarea asignada.			
La rúbrica es práctica de usar y clara, además evalúa el desempeño oral de los estudiantes en el nivel B1; algunos alcanzan el descriptor experto.			
La evaluación oral es adecuada de acuerdo con los profesores y alumnos que la emplean.			

En otras palabras, en esta etapa del proceso se evalúa la eficacia de la evaluación misma. Un grupo de expertos distinto a los elaboradores valora si la tabla de especificaciones cumple con ser clara, establecer la capacidad que se pondera y que a partir de ella se puedan elaborar tareas o pruebas; también, si la tarea o las pruebas son adecuadas, si la rúbrica es práctica, clara y útil para calificar, y si la calificación se estandariza a partir de ésta. Se pueden hacer comentarios en la lista de cotejo que sugieran o recomienden modificaciones. Los elaboradores reciben esta evaluación y, a partir de ella, hacen los ajustes pertinentes.

Además de probar las tareas y las rúbricas en el proceso de experimentación, también se pone a prueba el protocolo de aplicación: cómo se recibe al o a los estudiantes que van a hacer el examen; la disposición del espacio; la interacción entre el o los examinadores con el estudiante; el tiempo de duración del examen; el cierre y la despedida, así como la comunicación de los resultados al estudiante; y, en dado caso, el proceso de revisión de examen, si el estudiante así lo solicita.

Después de la validación, se hacen ajustes a los materiales y documentos. Posteriormente, una vez que se ha experimentado el examen, nuevamente se aplican las modificaciones necesarias para que la versión final alcance un resultado óptimo. Sólo entonces es momento de aplicarlo al final de un curso de español para calificar y evaluar el aprovechamiento.

En síntesis, el proceso de elaboración de una evaluación oral no es tarea sencilla. Se requiere de expertos en el área, que elaboren las tablas, las

tareas y las rúbricas; además, de otro grupo de expertos que revisen y validen el diseño y la elaboración de todos los materiales. Es necesario ajustar aquello que sea pertinente para garantizar la validez y confiabilidad del instrumento de evaluación. Es importante que este proceso se siga de manera rigurosa, ya que si no hay control sobre él, pueden aparecer inconsistencias o errores que únicamente afectarían al estudiante. El equipo de expertos revisores en la validación no puede hacer las correcciones directamente sobre los materiales, sino que envían sus observaciones y los elaboradores son quienes corrigen. De esta manera, el proceso será riguroso y claro. Como puede verse, se requieren de muchas manos en este proceso de elaboración de instrumentos de evaluación oral.

En algunos contextos institucionales, es posible que no haya un grupo grande de profesores que puedan diseñar y elaborar las pruebas, ni tampoco que pueda validarlas. Probablemente sea más común que un profesor de ELE se encuentre solo ante la tarea de evaluar oralmente a su grupo al final de su curso. En este caso, este profesor solitario podría seguir los lineamientos comentados y ejemplificados en este capítulo y buscar quizá a algún colega que pueda revisar su tabla de especificaciones, sus pruebas y tareas y su rúbrica, antes de aplicar la evaluación oral. Esto, en aras de la comunidad de prácticas que puede conformarse en la tarea evaluativa del español como lengua extranjera.

4. Conclusiones

De manera general, en la evaluación de la oralidad, como en toda tarea, lo principal es tener claro qué se evaluará y para qué. Teniendo bien identificados los objetivos, se deberá buscar la estrategia más adecuada para obtener la muestra de lengua que se necesita, así conseguiremos la respuesta del cómo evaluar, lo cual nos llevará a la planificación del instrumento con las pautas adecuadas. Como siguiente paso, emprendemos la elaboración de las rúbricas de las tareas. Después, como en todo proceso, se deberá comprobar que tanto nuestros objetivos como las tareas y las rúbricas están siendo justos, válidos y confiables, mediante instrumentos que traduzcan los resultados en pruebas objetivas; es decir, mediante datos psicométricos que nos guiarán en los ajustes específicos que deban realizarse o, en dado caso, en la modificación integral de nuestro instrumento, hasta dar con el que se ajuste mejor a nuestros objetivos. Éste, además, debe ser justo con la evaluación de los alumnos, siempre tomando en cuenta el tipo de actividades y habilidades que fueron puestas en práctica durante las clases del curso.

En este capítulo, hemos presentado una serie de lineamientos que pueden seguirse para elaborar un instrumento de evaluación oral: desde la planeación y los objetivos, hasta la validación y experimentación de la prueba.

Esta forma de elaborar instrumentos de evaluación oral, puede ayudar a que haya validez, confiabilidad y justicia en los exámenes que presenten nuestros estudiantes de español como segunda lengua. También es viable seguir estas pautas de acuerdo con las necesidades y las circunstancias de cada centro o instituto de lenguas, siempre y cuando se establezcan los documentos más importantes: la definición de constructos, la tabla de especificaciones y la rúbrica. Una recomendación final importante es documentar todo el proceso: es posible acompañar el examen o la prueba de un manual técnico que recabe todos los datos al respecto: quiénes participaron en él y cuáles fueron los fundamentos para las decisiones tomadas.

Evaluar el aprendizaje no es tarea sencilla. Tampoco lo es evaluar la oralidad. Pero si nos formamos en el proceso de elaboración de instrumentos y aprendemos la manera de cuidar la validez, la confiabilidad y la justicia, cada día nos acercaremos más al verdadero propósito de toda evaluación: obtener información veraz para medir o valorar el aprendizaje y tomar decisiones para la mejora de éste.

Referencias

- Barrón, A.** (2003). *Acquisition in interlanguage pragmatics*. John Benjamins.
- Bordón, T.** (2000). *La evaluación de la expresión oral en el aula de E/LE*. Arco Libros
- Chapelle, C.** (2011). Validation in language assessment. En Hinkel, E. (Ed.). *Handbook of research in second language teaching and learning*. Volume II (pp. 717-730). Routledge.
- De Santiago, J.** y Fernández, J. (2017). *Fundamentos para la enseñanza del español como 2/L*. Arco Libros.
- Fung, L.** y Carter, R. (2007). Discourse markers and spoken English: native and learner use in pedagogic settings. En *Applied Linguistics*, 28 (3), 410-439.
- González Garibay, V.;** Sierra González, R. & Sosa Ramírez, K. P. (2020). Rúbrica. En Sánchez Mendiola, M. & Martínez González, A. (Eds.). *Evaluación del y para el aprendizaje: instrumentos y estrategias* (pp. 109-125). Coordinación de Desarrollo Educativo e Innovación Curricular, UNAM.
- Hidri, S.** (2018). Assessing spoken language ability: a many-facet Rasch analysis. En Hidri, S. (Ed.). *Revisiting the assessment of second language abilities: from theory to practice* (pp. 23-48). Springer.
- López Olea, M.** (2017). *Marcadores del discurso en la interlengua de hablantes estadounidenses de español como segunda lengua: un estudio sobre su orden de adquisición (tesis de maestría)*. Universidad Nacional Autónoma de México, Ciudad de México.
- López Olea, M.** (2020). Cuestiones de léxico en el ámbito de la evaluación de español como L2. En Nieto Rodríguez, M. & Pacheco Martínez, E. R. (Coords.). *Ciclo Internacional de Conferencias sobre Evaluación*. Conferencia llevada a cabo en el Centro de Enseñanza para Extranjeros, UNAM, Ciudad de México.
- López Olea, M.** y Nieto Rodríguez, M. (2020). La relevancia del léxico en la certificación del español. En Carmona Hernández, M. (Presidencia) *11o Simposio Internacional del Centro de Enseñanza para Extranjeros*. Simposio llevado a cabo en el Centro de Enseñanza para Extranjeros, UNAM, Ciudad de México.
- Luoma, S.** (2004). *Assessing speaking*. Cambridge University Press.

Montoya Magno, M. y Pérez Díaz, A. (2020). Exposición oral. En Sánchez Mendiola, M. y Martínez González, A. (Eds.). *Evaluación del y para el aprendizaje: instrumentos y estrategias* (pp. 165-178). México: Coordinación de Desarrollo Educativo e Innovación Curricular, UNAM.

Pastor Cesteros, S. y Ferreira Cabrera, A. (2018). El discurso académico en español como LE/L2: nuevos contextos, nuevas metodologías. En *Journal of Spanish Language Teaching*, 5 (2), 91-101.

Sánchez Mendiola, M. y Martínez González, A. (2020). *Evaluación del y para el aprendizaje: instrumentos y estrategias*. México: Coordinación de Desarrollo Educativo e Innovación Curricular, UNAM.

Vergara Padilla, M. A. y Ainciburu, M. C. (2016). Las presentaciones orales en la evaluación del español académico. Las percepciones previas del docente y la actuación de alumnos universitarios de diferentes nacionalidades. En *La formación y competencias del profesorado de ELE: XxVI Congreso Internacional ASELE*. Asociación para la Enseñanza del Español como Lengua Extranjera.

Villar, C. M. (2011). Las presentaciones académicas orales de los estudiantes alemanes de E/LE. Del discurso monológico al dialógico. En *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, 10, 130-172.

Villar, C. M. (2014). *Las presentaciones académicas orales en E/LE de estudiantes alemanes*. Peter Lang.

CAPÍTULO 5

ANÁLISIS DE LOS ACTUALES PROCEDIMIENTOS E INSTRUMENTOS DEL CEPE PARA LA EVALUACIÓN DE LA EXPRESIÓN ORAL

Marcela **Nieto Rodríguez**
Edgar René **Pacheco Martínez**

Introducción

Los exámenes de aprovechamiento con que se cuenta a la fecha en el CEPE son producto de varios desarrollos, que tuvieron lugar en esta institución desde principios de los años 2000 y 2001. Durante este periodo se actualizaron los programas de estudio, por lo que pasamos de cinco a nueve niveles de lengua. Dicha transformación curricular obedecía también a que, en esa misma época, apareció el MCER en Europa, el cual influyó en el diseño curricular de la enseñanza de lenguas, tanto en el CEPE como en otros centros de enseñanza del planeta.

Asimismo, en el año de 2002 la Universidad de Salamanca y el CEPE firmaron un convenio de colaboración, en pos de homologar sus respectivos contenidos curriculares, con la finalidad de elaborar exámenes de certificación que pudieran ser validados mutuamente (Campos, 2005: 56). La reestructuración curricular que sufrió el CEPE en aquel momento abrió paso a la construcción de una base teórico-pedagógica que serviría para el diseño del Certificado de español como lengua adicional (CELA), que se homologa con los exámenes DELE en sus distintos niveles. El nuevo programa de estudios, tras un largo proceso de diseño e intenso trabajo colegiado, entró en vigor en 2007.

En el marco de la creación de los exámenes del CELA, la experiencia acumulada en el diseño curricular y el trabajo conjunto con instancias internacionales, como la USAL y el Instituto Cervantes, motivaron de igual manera la necesidad de contar con nuevos instrumentos de evaluación en los cursos de lengua. En este mismo impulso, los exámenes de aprovechamiento se adaptaron al nuevo programa y en 2011 se computarizaron las secciones de sistematización lingüística, comprensión de lectura y comprensión auditiva, para facilitar su administración, mientras que se decidió evaluar con pruebas subjetivas las expresiones oral y escrita.

Esto se consideró un parteaguas en la estandarización de los instrumentos y de los criterios de evaluación, pues anteriormente se encargaba a cada profesor la creación de exámenes, su aplicación y calificación. Se partió del trabajo colegiado para la definición de los contenidos de los exámenes, teniendo como base las tablas de especificaciones elaboradas a partir de los perfiles de egreso, determinados a su vez en el nuevo programa de estudios, puesto en vigor en 2007. Dichas tablas de especificaciones fueron el punto de partida para el diseño de las pruebas objetivas, que serían desde entonces aplicadas por computadora; por su parte, para las pruebas subjetivas, se recurrió a rúbricas. Asimismo, la implementación y calificación de estas últimas, de acuerdo con el protocolo de aplicación de los nuevos exámenes, estarían a cargo de los maestros de cada curso, mediante intercambio de pares (Campos, 2011: 125).

2. Los instrumentos de evaluación de EO de los niveles 1, 2, 3 y 4

2.1. Qué se evalúa

En los documentos pertinentes para este análisis que fueron consultados, se encontró que los instrumentos de evaluación de EO de los niveles mencionados evalúan los contenidos establecidos en el nuevo programa de estudios del CEPE que, como se acota en la introducción, está vigente desde el año de 2007. De igual modo, se afirma que los niveles de nuestro programa están homologados con los niveles de referencia del MCER.

Sin embargo, a raíz de esta investigación se ha determinado que, aun cuando las pruebas objetivas correspondientes a la sistematización lingüística y a las comprensiones lectora y auditiva fueron elaboradas a partir de tablas de especificaciones (vinculadas a su vez con los contenidos del programa del CEPE), las pruebas subjetivas no se diseñaron a partir de tablas: sólo se elaboraron rúbricas para orientar a los docentes en aquellos aspectos a evaluar sobre las expresiones oral y escrita (Campos, 2011: 132).

Este dato no es de menor importancia. Por el contrario, nos parece que es causa y origen de algunos aspectos que podrían ser sujetos a mejoramiento, pues hemos identificado que al no contar con tablas de especificaciones para las pruebas subjetivas de expresión oral y escrita, las tareas evidencian considerables variaciones entre sí en cada nivel. Pensamos que establecer tablas de especificaciones, no sólo para las pruebas objetivas, sino también para las subjetivas, ayudaría a darles uniformidad a las tareas. Sin embargo, de ello hablaremos con mayor detalle hacia el final de este capítulo.

Así pues, en los exámenes de EO de los primeros cuatro niveles del programa de español del CEPE, se evalúan los contenidos especificados para cada nivel, determinados a su vez por sus correspondientes perfiles

de egreso. Sin embargo, no se define claramente qué contenidos específicos son los que se tienen que evaluar, lo que deja espacio amplio para que intervenga la subjetividad del profesor que diseña las tareas, así como la de quienes aplican el examen y evalúan el desempeño de los estudiantes; con lo cual, se reduce la validez de las pruebas.

2.2. Descripción de los instrumentos

Aunque más adelante se hará una descripción detallada de cómo se lleva a cabo la aplicación del examen de EO, por lo pronto adelantaremos que consta de tres tareas: la primera consiste en una breve entrevista entre pares. Su finalidad es la de establecer un ambiente de relajación para los estudiantes y romper el hielo entre ellos; la segunda consiste en la producción de cada uno de los estudiantes; y por último, la tercera consiste en la interacción, realizada mediante un juego de roles. De este modo, aquí haremos la descripción de los modelos de exámenes por cada nivel, divididos en tareas de producción y tareas de interacción. La descripción que presentamos se hace de forma general, para salvaguardar los bancos de exámenes del CEPE; sin embargo, a lo largo del capítulo mencionaremos algunos otros detalles específicos que nos han parecido importantes, en cuanto a que presentan áreas de oportunidad o mejoramiento.

Español 1.

Todos los modelos de este nivel, de producción e interacción, tienen fecha de elaboración del año 2013.

Producción.

Modelo 1. Describir física y psicológicamente a una persona.

Modelo 2. Describir un lugar.

Modelo 3. Dar información personal sobre gustos y preferencias.

Interacción.

Modelo 1. Diálogo entre un médico y un paciente.

Modelo 2. Diálogo por teléfono entre el vendedor de una casa y su comprador.

Modelo 3. Diálogo entre un recepcionista de hotel y un turista.

Modelo 4. Diálogo entre dos amigos para decidir el lugar dónde vacacionar: la playa o la montaña.

Modelo 5. Diálogo entre dos amigos para decidir el lugar dónde vacacionar: la ciudad o el campo.

Modelo 6. Diálogo entre dos amigos para decidir el país dónde vacacionar: España o un país de Latinoamérica.

Español 2

Producción

Es importante decir que los modelos de producción se componen de dos grupos, elaborados en diferentes años. El primero de ellos, del modelo 1 al 4, pertenece al 2011; el segundo, del modelo 5 al 14, pertenece al 2017.

Modelo 1. Describir acciones y estados en futuro, imaginando cómo será el mundo en unas décadas.

Modelo 2. Dar información y expresar preferencias al hablar sobre su deporte favorito.

Modelo 3. Dar información y expresar preferencias al hablar sobre su espectáculo favorito.

Modelo 4. Describir acciones y estados en futuro, mediante la exposición de cómo cree el estudiante que será la vida en las ciudades del futuro.

Modelo 5. Describir acciones y estados usuales en pasado, usando el tiempo verbal copretérito, mediante la narración de cómo era la vida escolar del estudiante.

Modelo 6. Dar información, expresar opinión y argumentar sobre los medios de transporte.

Modelos 7, 8, 9, 10, 11 y 13. Describir acciones y estados en presente, usando como guía una fotografía en la que aparecen algunas personas realizando diversas actividades, en un lugar determinado.

Modelo 12 y 14. Describir acciones y estados en presente y en futuro, usando como guía una fotografía en la que aparecen algunas personas realizando diversas actividades, en un lugar determinado.

Interacción

Para la tarea de interacción, en el Departamento de Español se cuenta con cinco modelos, los cuales fueron elaborados en el año de 2011. A diferencia de los modelos de producción de este mismo nivel, que tuvieron un incremento en su número gracias a las adiciones del año 2017, los de interacción no aumentaron.

Modelo 1. Diálogo entre un mesero y un turista, en el cual se expresan saludos y se proporciona información personal; asimismo, se solicita y da información sobre alimentos y bebidas.

Modelo 2. Diálogo entre dos amigos para concertar su asistencia a un partido de fútbol, en el que se expresan saludos y se manifiestan preferencias; además se invita, se acepta una invitación, y se acuerda una cita.

Modelo 3. Diálogo entre dos amigos para organizar la fiesta de fin de cursos, en el que se expresan saludos, se invita y se acepta una invitación; además, se solicita y se proporciona información, se propone un plan y se llega a acuerdos, y, por último, se expresan frases de despedida.

Los dos siguientes modelos tienen como título “Acuerdo con un compañero”. Sin embargo, el modelo 4 tiene la misma temática que el 3: organizar una fiesta de fin de cursos. A diferencia del modelo 3, el 4 no cuenta con instrucciones detalladas, sino más bien demasiado escuetas, lo que no brinda mucho margen de comprensión al estudiante: “haces propuestas y le preguntas (al compañero) si está de acuerdo o no”.

Como recién dijimos, el modelo 5 se titula también “Acuerdo con un compañero”; pero mientras los dos modelos anteriores tratan sobre coordinar una fiesta, en este último se plantea la organización de un viaje entre los dos amigos, que es similar a la temática de modelos de interacción del Español 1, en los que hay un desacuerdo entre destinos de viaje: tienen que argumentar para convencer al otro y, finalmente, llegar a un acuerdo. No obstante, a pesar de que se le podría sacar mucho provecho a la situación planteada, las instrucciones son idénticas a las del modelo 4.

Español 3

Producción

Los modelos de producción de este nivel tienen como fecha de elaboración el año 2011.

Modelo 1. Narrar en el pasado con apoyo gráfico (imágenes de carteles de películas y de personas dando y recibiendo regalos) detalles sobre la historia de una película, o anécdota personal de un regalo que el estudiante no esperaba.

Modelo 2. Narrar en el pasado con apoyo gráfico detalles sobre la historia de una novela o un cuento que el estudiante haya leído, o bien acerca de la experiencia de un robo que haya sufrido, o sobre alguna vez en que haya estado en una situación en la que debía tomar una decisión muy importante para su vida.

Modelo 3. Narrar en el pasado con apoyo gráfico detalles sobre la historia de una película que el estudiante haya visto y le haya gustado, o acerca de un recuerdo muy relevante de su vida escolar en la época de secundaria/preparatoria, o bien sobre una situación en la que al estudiante le faltó valor para realizar algo.

Interacción

Los modelos de interacción de este nivel también tienen fecha de elaboración de 2011.

Modelo 1. En este modelo, se trata de llegar a un acuerdo con el compañero, sobre qué película verán juntos. Los estudiantes se saludan, uno invita y el otro acepta la invitación; se argumenta y se presentan contraargumentos y, al final, llegan a un acuerdo entre los dos y se despiden.

Modelo 2. Igual que en el modelo anterior, los estudiantes también tienen que llegar a un acuerdo sobre qué regalo le darán a una compañera. Los estudiantes se saludan, uno de ellos expone la situación, para enseguida argumentar y presentar contraargumentos; finalmente, ambos llegan a un acuerdo y se despiden.

Modelo 3. Se desarrolla una situación en la que un estudiante marca por teléfono a un empleado de una oficina de gobierno para agendar una cita de renovación de pasaporte. Ambos se saludan, uno de ellos solicita información y el otro la proporciona; al final, acuerdan una fecha y hora para la cita y se despiden.

Modelo 4. La situación consiste en que un estudiante solicita información a un empleado del CEPE sobre los cursos, horarios y fechas. Ambos se saludan, uno solicita la información y el otro la proporciona; para concluir, agradecen y se despiden.

Modelo 5. La situación se desarrolla en el aeropuerto. Los estudiantes se saludan, uno debe solicitar información sobre vuelos y horarios, porque espera a un grupo de amigos y su vuelo se encuentra retrasado, y el otro la proporciona; finalmente agradecen y se despiden.

Modelo 6. Al igual que en el modelo anterior, éste también se desarrolla en el aeropuerto, porque la maleta de un estudiante se ha perdido. Es así como éste explica la situación y le describe la maleta al estudiante que funge como empleado de la oficina de información. Se solicita y se proporciona información personal; al final, agradecen y se despiden.

Español 4

Producción.

Los modelos de producción del 1 al 9 tienen como fecha de elaboración el año 2013. De los modelos 10 y 11, no se cuenta con ningún dato. Por lo general, se les da a los estudiantes dos minutos para la preparación y tres minutos para la realización de la tarea, aunque en las instrucciones de algunos modelos puede haber discrepancias: 30 segundos para preparar, de 3 a 4 minutos para realizar la tarea, etc.

Modelos 1, 2, 3 y 4. Se trata de una serie de 5 frases célebres de personajes históricos o de autores literarios. La función que se les pide a los estudiantes demostrar es la expresión de acuerdo o desacuerdo, con respecto a dos frases que tienen que elegir de entre la lista de cinco; finalmente, deben argumentar sus razones para concordar o estar en desacuerdo. Es importante decir que estos modelos no son operativos.

Modelo 5. Este modelo se compone de dos partes. En la primera, se le pide al estudiante que se presente, que exprese sus intereses y preferencias. En la segunda, tiene que elegir hablar sobre personas, programas o libros que le gusten; o bien sobre un hecho "inusual" o un conflicto que haya experimentado en su vida.

Modelo 6. Este modelo, a diferencia del anterior, sólo consta de una parte. Cuenta con un título que le ayuda al estudiante a darse una idea previa sobre lo que tendrá que desarrollar en su exposición. En seguida, se plantea una interrogante sobre un tema polémico; en este caso, la imagen de la mujer en los anuncios de publicidad. Posteriormente, se le proporciona al estudiante una lista de cuatro aspectos a desarrollar. Finalmente, se le pide que exprese su opinión sobre el tema y que argumente su postura.

Modelo 7. Este modelo es estructuralmente muy parecido al anterior. Presenta el tema desde el inicio, a partir de un título. Las instrucciones, que se enuncian inmediatamente, solicitan que el estudiante hable sobre aspectos positivos y negativos sobre el internet como herramienta en la educación y en la vida cotidiana; de manera que el texto oral que se espera que desarrolle el examinado tendrá una parte argumentativa y otra de contraargumentos. A continuación, se enlistan aspectos positivos y negativos, cuyo objetivo es guiar la producción del estudiante.

Modelo 8. Este modelo sigue estructuralmente a los dos anteriores. En las instrucciones, se le pide al estudiante que hable sobre las ventajas de los teléfonos celulares en la vida cotidiana; sin embargo, no se solicita que mencione desventajas, si bien abajo se enlistan aspectos positivos y negativos, cuyo objetivo es fungir como orientación para el desarrollo del tema.

Modelo 9. Similar a los modelos 6, 7 y 8; se diferencia de éstos en que se le pide al estudiante ya no sólo que exprese los aspectos positivos y negativos del sistema de transporte del “Distrito Federal”, sino que también debe comparar dicho sistema con el de su ciudad de origen.

Modelo 10. Este modelo parece ser una versión mejorada del modelo 7, pues comparten el mismo tema. El título y las instrucciones son idénticas, pero en algunos puntos se aprecia que se trata de un ajuste del modelo anterior. En primer lugar, cuenta con un ejemplo que proporciona un modelo de expresión: “me gusta usar internet cuando...”, el cual pretende orientar al estudiante en la estructura elicitada, aun cuando, estrictamente hablando, ésta no pertenece a los contenidos específicos del Español 4. Y en segundo lugar, los aspectos positivos y los negativos se enuncian como “ideas de apoyo” y ya no son palabras individuales, sino frases que le aportan al estudiante información más detallada para consolidarse como una guía más efectiva.

Modelo 11. A pesar de no contar con las fechas de elaboración de estos modelos, es posible observar cuáles son más recientes, porque presentan muchas mejoras. El decimoprimer modelo es, a todas luces, uno mejor estructurado: cuenta, en primer lugar, con un título que establece un dilema, a partir del que se desarrollarán argumentos y contraargumentos (¿medicina natural o de patente?); enseguida, presenta un texto breve, cuya extensión es de 110 palabras, que funciona como input para contextualizar y activar los conocimientos previos del estudiante; en tercer lugar, se proporcionan ejemplos de estructuras sintácticas y formas gramaticales significativas de acuerdo con el nivel; y, por último, se muestran algunas ideas que funcionan como orientación. Es relevante advertir que las estructuras sintácticas y la mayor parte del léxico de toda la tarea, son adecuados para el nivel de Español 4.

Interacción

En cuanto a los modelos de interacción para este nivel, del 1 al 4 se registran en el 2013. Respecto al modelo 5, no se señala ningún dato. Todos ellos consisten en debates sobre dilemas que implican la toma de decisiones compartidas. Su estructura replica la de los modelos de producción de este mismo nivel: elegir una de las opciones que plantea el dilema, defender la postura propia y explicar sus ventajas y desventajas. Sin embargo, a diferencia de las producciones, estas tareas de interacción contemplan el diálogo entre dos estudiantes y, por lo tanto, la interacción entre ellos.

Modelo 1. En este modelo, los estudiantes tienen que discutir sobre las ventajas y las desventajas de concebir el deporte como un negocio. De este modo, las funciones comunicativas serán expresar su opinión sobre dicho tema, argumentar para apoyar su punto de vista y contraargumentar para debilitar la opinión contraria. Al final, se enlistan algunas

palabras aisladas que conforman, en un primer grupo las ventajas, y en un segundo, las desventajas.

Modelo 2. Este modelo tiene como tema las ventajas y desventajas de ser un deportista famoso. Su estructura es muy parecida a la del modelo 1: instrucciones, título de la tarea, pautas y lista de ventajas y desventajas, enunciadas mediante palabras aisladas.

Los tres siguientes modelos son muy parecidos entre sí, aunque el 3 y el 4 tienen como fecha de elaboración el año 2013, mientras que el 5 no la enuncia.

Modelo 3. Trata el tema del nuevo concepto de belleza. A diferencia de los anteriores, tiene un texto input de 42 palabras de extensión, con una pregunta al final que prepara al estudiante para el debate que debe desarrollar con el compañero. Después, se muestran las pautas y, al final, los aspectos positivos y los negativos, ya no en palabras aisladas, sino en frases que aportan mayor contexto y guía.

Modelo 4. Los estudiantes deben debatir sobre el tema de las playas privadas. Igual que el modelo 3, presenta un texto input de 30 palabras, con una pregunta al final que enmarca la discusión; asimismo, cuenta con las pautas y la lista de frases sobre ventajas y desventajas.

Modelo 5. Es posiblemente el modelo más reciente porque, pese a calcar la estructura de los modelos 3 y 4, añade un elemento extra: solicita del estudiante una narración sobre una experiencia que haya vivido en el pasado. El debate gira en torno a la elección entre “comida sabrosa o comida saludable”. Cuenta con un texto input de 108 palabras, el más extenso de todos, con una pregunta al final que sirve de encuadre. Al final, se presentan las pautas, así como las ventajas y desventajas en forma de frases.

2.3. Cómo y por qué se evalúa

Como se verá más adelante, en el capítulo 8 (donde abordamos el análisis de los procedimientos y de los instrumentos de evaluación de la EE, así como en el presente nos enfocamos en esta misma temática, pero correspondiente a la EO), pormenorizaremos cómo se evalúa y determina la calificación; en parte, por los niveles de dominio del español, o por los objetivos de aprendizaje que se han alcanzado a lo largo de un curso, o de un periodo de tiempo de exposición a la lengua meta. El principal documento que sirve para determinar las fronteras entre cada nivel es el *Programa de estudios de los cursos de español del CEPE*, que especifica los contenidos nociofuncionales y las estructuras gramaticales asociadas, los contenidos para la competencia sociolingüística y aquellos para la competencia estratégica. Otro texto igualmente útil para especificar los

contenidos correspondientes a los diferentes niveles de dominio de una L2, que también se emplea en el CEPE, es el ya ampliamente aludido hasta este punto MCER; a partir de esta referencia, nuestro centro homologó sus niveles con los establecidos ahí. Las correspondencias se muestran en la siguiente figura:

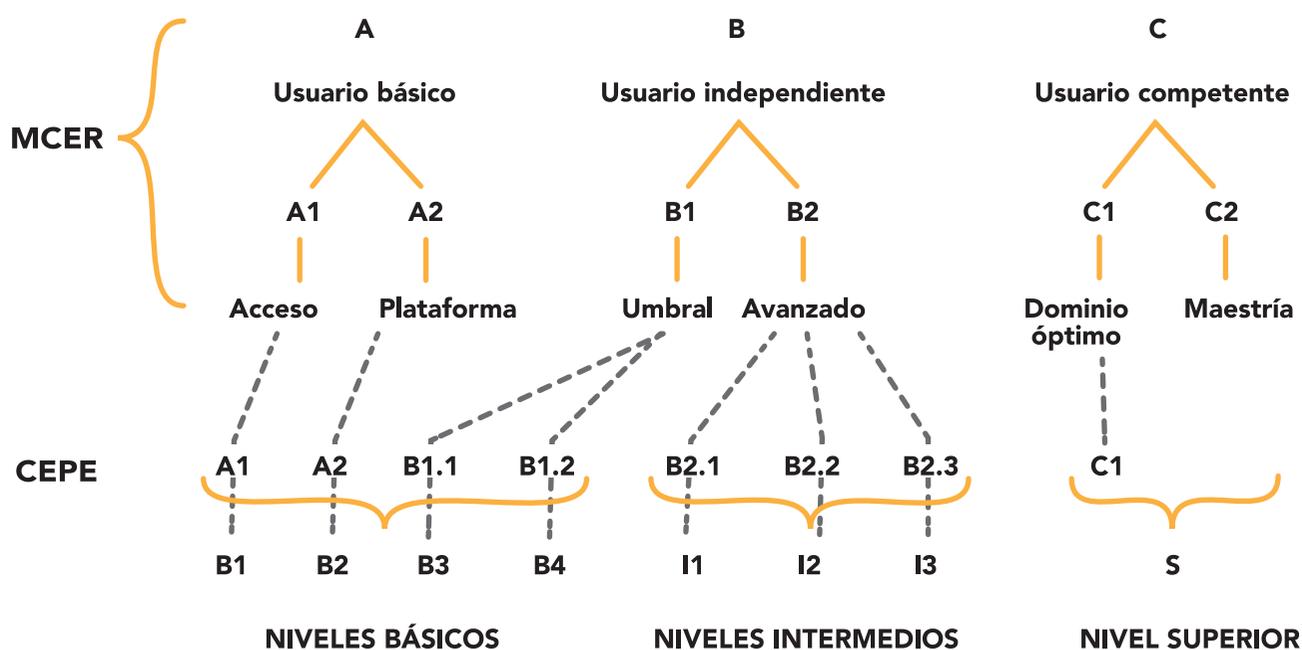


Figura 3. Correspondencia entre los niveles del MCER y los del CEPE (imagen tomada del Programa de estudios, CEPE, 2014, p. 37).

Así pues, los objetivos de aprendizaje y los contenidos de cada nivel de español (recordemos que los analizados aquí son los niveles 1, 2, 3 y 4, es decir, A1, A2 y B1 del MCER) están especificados en nuestro currículo. Tras la última actualización curricular de los cursos de español y cultura del centro, se renovó también toda la instrumentación de evaluación en el área de español. Consecuencia de ello fue la computarización de los exámenes para la evaluación de aprovechamiento de la competencia lingüística, así como las habilidades de comprensión auditiva y lectora, para las cuales se elaboraron tablas de especificaciones que definieron los contenidos y objetivos por alcanzar, al final de cada nivel.

Sin embargo, para el caso de las expresiones oral y escrita, no se elaboró tabla alguna, sino solamente baremos o rúbricas. Este hecho trae como consecuencia algunas problemáticas que permean todo el proceso del desarrollo evaluativo, incluidos los instrumentos, los procedimientos, el entrenamiento de evaluadores expertos, las pruebas y tareas, los análisis

de resultados, la retroalimentación para la institución, la evaluación de la evaluación, etc. Ejemplo de ello es que los modelos de exámenes de EO demuestran una amplia variedad de contenidos, objetivos, diseños, entre otros detalles, que (desde nuestra perspectiva) reducen de manera drástica la validez y la confiabilidad de las evaluaciones en nuestra institución.

No obstante la afirmación anterior, estamos convencidos de que, mediante una metodología rigurosa, es posible reformular las debilidades que hemos identificado a fin de fortalecer lo que estimamos que funciona adecuadamente, sobre lo cual hablaremos con mayor detalle en las secciones posteriores.

2.4. Análisis de los instrumentos

Los modelos de examen de Español 1 para la producción oral tienen en común funciones comunicativas propias del nivel, tales como describir física y psicológicamente a individuos, así como lugares u objetos; o bien, proporcionar información personal básica: datos personales, gustos, aficiones, etcétera. Si bien los contenidos, las funciones y las situaciones se adecuan al nivel respectivo, una de las problemáticas que identificamos en estos modelos es que la redacción de las pautas debe cuidarse más en función de qué estructuras gramaticales y qué léxico se emplea. Por ejemplo, encontramos una oración subordinada relativa sustantiva (“nombra lo que hay en la fotografía...”) y léxico fuera de nivel, lo cual puede simplificarse a fin de ajustar tanto estructuras gramaticales, como vocabulario, para no causar incomprensiones innecesarias y no frustrar al examinado.

Los modelos de interacción de este nivel implican juegos de roles, en los que las parejas de estudiantes tienen que desarrollar diálogos en situaciones comunicativas específicas (tales como una consulta médica, llamadas telefónicas, transacciones comerciales o tomas de decisiones entre familiares y amigos). En estos modelos, identificamos estructuras sintácticas y léxico igualmente más complejos en relación con el nivel que se está evaluando; por ejemplo: construcciones invertidas con verbos de emoción, propios del Español 2; verbos en presente de subjuntivo en oraciones subordinadas; oraciones subordinadas relativas adjetivas, entre otros detalles. Aunque el estudiante podría escuchar estas formas durante su inmersión, no tendría cómo ni por qué dominar las formas ajenas a su nivel en nuestros modelos de examen. Esto bien puede solventarse, si adecuamos estos aspectos a fin de que nuestras instrucciones y pautas sean comprensibles.

Con respecto a los modelos de producción de Español 2, como se mencionó anteriormente, subrayamos que pertenecen a dos grupos: uno elaborado en 2011 y el segundo en 2017. En general, las funciones comunicativas que se busca que produzcan los estudiantes son descripción

en presente, pretérito y futuro, tanto de personas y lugares, como de acciones y estados; proporcionar información personal y expresar preferencias, gustos y aficiones; expresar la opinión propia y argumentar en torno a ella; y, finalmente, expresar secuencias narrativas, en pretérito y copretérito, sobre hechos pasados. Algunos aspectos problemáticos recurrentes son las estructuras sintácticas y el léxico, evidentemente fuera de nivel; por ejemplo: “describe cómo piensas que será la vida en las ciudades del futuro”, podría simplificarse a “¿Cómo será la vida en el futuro?” Entendemos que aparece el verbo “describir” para especificarle al estudiante qué función debe demostrar que domina; sin embargo, también pensamos que la pregunta misma puede detonar una respuesta en la que encontremos dicha función comunicativa, sin necesidad de ser explícitos. Otro punto relevante es que las temáticas de los modelos se repiten de forma constante (como el caso de deporte y futuro), lo que reduce la creatividad de los estudiantes para elaborar respuestas; sin mencionar que algunos estudiantes expresan su desacuerdo en hablar sobre deportes, por lo que solicitan modificar el tema para su producción. ¿La motivación detonada por el tema de conversación, podría influir en el desempeño de un estudiante en el momento de su evaluación? Es una pregunta que nos obliga a reflexionar sobre las temáticas que seleccionamos y en la necesidad de enriquecer nuestros bancos de exámenes, para favorecer la diversidad de intereses. Asimismo, sería pertinente pensar que las pautas, a veces introducidas en forma de preguntas sin orden subyacente, podrían organizarse mejor, a fin de proporcionarle al estudiante una guía certera para desarrollar de forma lógica su participación.

Por su parte, los modelos de interacción de español 2, elaborados todos en el año de 2011, tienen como inconveniente su antigüedad, que contrasta con los modelos de producción añadidos en 2017, elaborados para enriquecer el banco de modelos de este mismo nivel. Los cinco modelos de interacción se presentan como juegos de roles, mediante los cuales, los estudiantes deben elaborar diálogos en situaciones comunicativas más complejas que aquellas planteadas para el nivel de Español 1: diálogos sobre transacciones comerciales y otros en los que está implicada la toma de decisiones. Por otro lado, identificamos estructuras sintácticas y léxico por encima del nivel que se evalúa en las pautas, factor que provoca preguntas para aclarar la comprensión durante el examen. Esto lo sabemos por experiencia propia y también lo pudimos constatar mediante las grabaciones de exámenes orales de aprovechamiento, realizadas para esta investigación.

Los modelos de producción de Español 3 fueron diseñados igualmente en el año de 2011, por lo que cuentan con una década de antigüedad. Las funciones comunicativas solicitan elaborar narraciones sobre experiencias personales ocurridas en el pasado, con la ayuda de elementos

visuales. Claramente, el objetivo de estos modelos es que el examinado demuestre su control y su capacidad de discriminar los usos adecuados de los tiempos pretérito y copretérito.

De igual suerte, todos los modelos de interacción de este nivel pertenecen también al año 2011. En los seis modelos encontramos juegos de roles que proponen situaciones comunicativas en contextos personales o públicos. En dos de ellos, los estudiantes deben tomar decisiones a partir de propuestas en común, mientras que en el resto de los modelos, los estudiantes deben interactuar para solicitar y proporcionar información.

Algo que destaca a los ojos de la época actual (años veinte del siglo XXI), sin duda es el tema de la equidad de género, sobre todo en el campo de enseñanza de ELE (Gómez, 2021). Identificamos que los temas sobre belleza, moda y dietas muestran un marcado sesgo hacia las mujeres, pues en las instrucciones se llega a enunciar “compañera” en lugar de “compañero”; mientras que en los temas relacionados con deportes, como fútbol, mencionan que se tiene que invitar a un “amigo” y no a un “amigo o amiga”. Este tipo de minucias pasaban desapercibidas años atrás, pero hoy en día son fácilmente identificables y podemos ir trabajando en diseñar materiales que demuestren nuestro compromiso con la equidad de género.

En nivel 4, se tiene el mayor número de modelos tanto para producción, como para interacción. Para la primera, se cuentan con 11 modelos, la mayoría elaborados en 2013, por lo que también convendría actualizarlos o renovar el banco de exámenes. En este nivel, identificamos que la tipología y la estructura de los modelos de examen son muy variadas entre sí. De los modelos 1 al 4, se tienen 5 frases célebres que funcionan como texto input para denotar la producción del estudiante. Sin embargo, dichas frases no han pasado por un tamiz, a fin de contextualizarlas correctamente y de adecuar sus estructuras sintácticas y su léxico, para ajustarlas al nivel apropiado. Además, los examinados no cuentan con pautas que les sirvan de orientación para desarrollar, paso a paso, cada parte de su intervención.

Los modelos del 5 al 9 son temática y estructuralmente similares. Todos ellos presentan un tema polémico inicial, a partir del cual el estudiante expresará su perspectiva personal, para después elaborar una argumentación que apoye su opinión propia. Y, a pesar de su similitud estructural, consideramos que las “pequeñas” variaciones son factores de diferenciación que aumentan de manera exponencial en el momento de la evaluación; pues, por ejemplo, en los modelos 5, 6, 7 y 8 se pide presentar su opinión y argumentar a favor de ella, mientras que en el modelo 9, se añade una última sección en la que se deben hacer comparaciones entre dos tipos de transporte de diferentes ciudades. Esto puede parecer, a primera vista, algo sin importancia, pero estamos convencidos de que todos los modelos

deben ser homogéneos (no que unos exijan más que otros) para garantizar equidad y justicia en la evaluación, de forma que ni el azar ni la suerte sean factores que intervengan en determinar un mejor o peor desempeño de nuestros estudiantes. Finalmente, los modelos 10 y 11 parecen ser adiciones al banco de exámenes de los últimos años, pues su estructura responde a modelos más estandarizados y, en la práctica, han demostrado su eficiencia para arrojar muestras más objetivas de las capacidades logradas por los alumnos al final del curso. Estos argumentos abonan todavía más a la urgencia de someter nuestros bancos de exámenes de EO a revisiones y actualizaciones constantes.

Los modelos de interacción de este nivel, también son similares a los modelos de producción: parten de presentar un tema polémico, mediante la introducción de un texto que sirve de input, a partir del cual el estudiante define su postura y argumenta a favor o en contra. La diferencia con los modelos de producción estriba en que los de interacción involucran la participación de dos estudiantes, y la confrontación –al menos, eso se espera– de sus posturas, lo que permite presentar argumentos y contraargumentar las ideas de cada uno.

En resumen, los modelos de examen de EO de Español 1, 2, 3 y 4 contienen algunas estructuras sintácticas, formas gramaticales y léxico que consideramos demasiado complejos para los niveles evaluados. Por tanto, es indispensable que todo profesor que se involucre en el diseño de tareas para la evaluación de EO esté consciente de estos aspectos. Hemos encontrado, además, que la estructura de todos los modelos analizados es muy diversa, lo que podría representar un factor que intervenga en un decremento de la validez y confiabilidad de nuestras pruebas; además de constituir condiciones poco equitativas en nuestras evaluaciones. Si la tarea planteada en un modelo es más fácil que la de otro, se debe sobre todo a las falencias en su diseño y a la falta de tablas de especificaciones para la evaluación de EO.

2.5. Cómo se administra la prueba

El examen de EO, a diferencia de todas las pruebas para las otras habilidades (incluida la EE), no se realiza por medio de una computadora. Al final de los cursos, la Coordinación de Español, por medio del Departamento de Apoyo a la Docencia, realiza la organización de los estudiantes de cada grupo en parejas. Cuando el número total de estudiantes del grupo es impar, alguno tendría que realizar su examen de forma individual. No obstante, en la práctica, se llega a solicitar a algún compañero que hizo su examen con otro, que “ayude” al alumno sin pareja para hacer su examen.

Una vez que los estudiantes están dispuestos en pares, se asignan a diferentes salones, que estarán a cargo de dos profesores: uno fungirá como calificador, tomando notas sobre el desempeño de los examinados, y el otro, como entrevistador, interactuando con ellos. Al final de la evaluación de todas las parejas, ambos profesores deliberan en torno a los puntajes que recibirá cada estudiante en dos secciones, producción e interacción, tomando como referencias criteriosales una rúbrica holística y otra analítica.

Cuando ha terminado de calificar a todas las parejas que les han sido asignadas, la dupla de profesores entrega los materiales del examen (aquellos donde están descritas las tareas, las hojas de registro de calificaciones, las rúbricas con las escalas y sus descriptores, así como los contenidos lingüísticos correspondientes a cada nivel) al Departamento de Apoyo a la Docencia, área que se encargará de registrar las calificaciones en el sistema de asuntos escolares del centro.

3. Análisis e interpretación de los instrumentos de EO de los niveles de Español 1, 2, 3 y 4

3.1. Aplicación de las pruebas de EO

Existen en el CEPE algunos documentos que sirven para orientar las acciones de los profesores, durante los exámenes de aprovechamiento de expresión oral de los niveles de Español 0 a Español 7, si bien aquí nos estamos concentrando en los niveles del 1 al 4. Hay un texto en particular que expone de forma detallada el protocolo del procedimiento para la aplicación del examen de EO: el documento del protocolo para la aplicación, proporcionado por la Coordinación de Español.

Ahí se explica que los dos profesores, entrevistador y calificador, pueden cambiar de rol en cada evaluación de parejas. La disposición de profesores y de estudiantes es fija: el calificador debe sentarse detrás del escritorio, apartado de los estudiantes, los cuales toman asiento uno junto al otro, de frente al profesor entrevistador, y de espaldas al calificador. La finalidad de dicha disposición es que los estudiantes no sientan presión por parte de este último, cuyas reacciones y movimientos corporales podrían influir en la calidad de su desempeño.

Veamos la figura en la siguiente página.

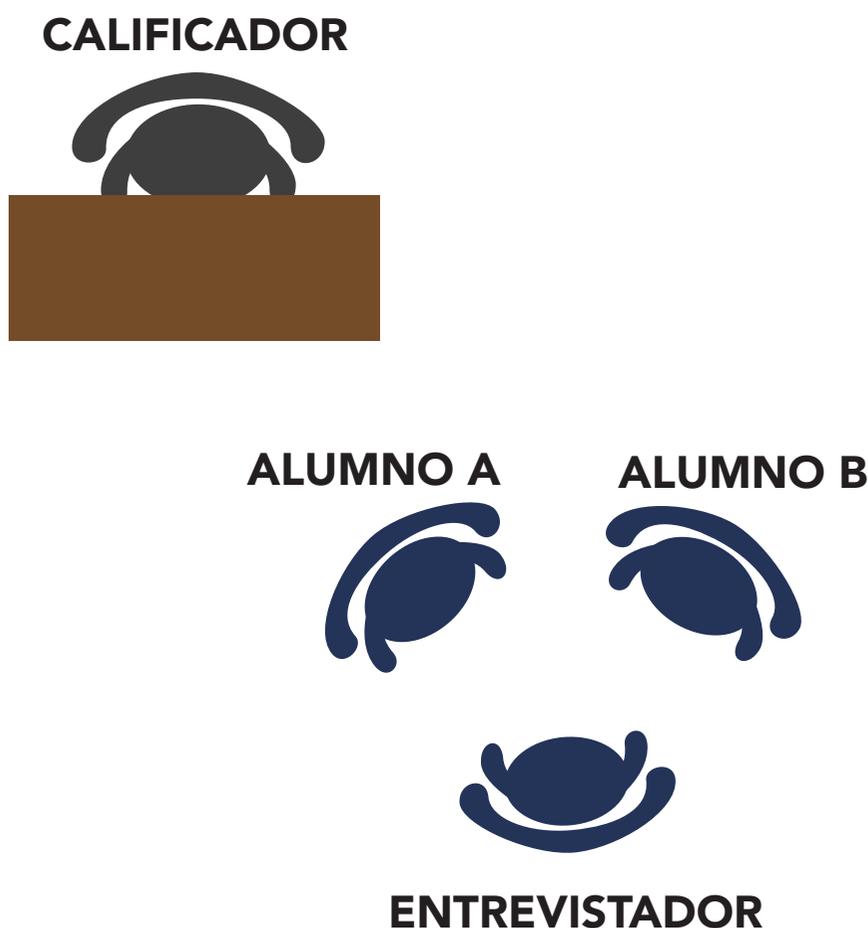


Figura 4. Ubicación de los participantes, profesores y examinados, en el examen de EO (imagen tomada de Campos et al., 2011:148).

El protocolo de aplicación de la prueba de expresión oral consiste en tres etapas: tarea 1. Entrevista entre pares; tarea 2. Producción; y tarea 3. Interacción (juego de roles). Antes de iniciar cualquier prueba, el profesor que funge como entrevistador recibe a la pareja de estudiantes y les señala los lugares que les corresponden. Enseguida, debe explicarles las etapas que componen el examen y cuáles son las tareas que deben realizar en cada una. Finalmente, da las indicaciones para iniciar con la primera de las tareas.

Para la entrevista entre pares, se prepara con anticipación una caja (en la práctica, se utilizan bolsas transparentes de plástico, puesto que son más fáciles de almacenar), en cuyo interior se guardan preguntas escritas en tiras de papel doblado. El entrevistador pide al estudiante A que extraiga un papel, que lo desdoble y le formule la pregunta al estudiante B, quien tendrá un minuto para planificar su respuesta y otro más para expresarla. Una vez terminado el turno del estudiante B, el entrevistador

repetirá el procedimiento para evaluar la respuesta del estudiante A. En todo momento, el entrevistador debe marcar los tiempos, con la finalidad de delimitar perfectamente los turnos de habla. Al final, recoge las tiras de papel, señala que la entrevista entre pares ha terminado y comienza la transición a la segunda tarea.

De acuerdo con el protocolo, esta primera tarea tiene como objetivo establecer un ambiente en el que los estudiantes se sientan relajados y con la confianza suficiente, antes de ser realmente examinados en las tareas 2 y 3. Asimismo, el protocolo indica que durante la tarea 1, el entrevistador debe dejar en claro que la interacción con él será mínima. La razón de que se tome esta medida es que las muestras de lengua arrojadas por la producción de los estudiantes se vea influida lo menos posible por la intervención del entrevistador. Algo muy importante es no olvidar que el calificador permanecerá en silencio, sólo observando, escuchando y evaluando.

Concluido lo anterior, inicia la etapa de producción. El entrevistador le entrega al estudiante A una tarjeta y una imagen (en la práctica, algunas tarjetas incluyen ya la imagen), indicándole que debe cuidar que el estudiante B no las vea. Le explica que cuenta con tres minutos, como máximo, para leer la tarjeta (las instrucciones, el texto de estímulo, si lo hay, y las pautas), observar la imagen y planificar su participación. Posteriormente, solicita al alumno que describa lo que hay en la imagen, siempre evitando mostrársela a su compañero, quien tendrá que poner mucha atención para ser capaz de plantearle una pregunta al final de su producción. La finalidad de la mecánica anterior es que el estudiante que escucha no sea sólo un espectador pasivo, sino que aproveche la producción oral de su compañero para comprobar su propia comprensión auditiva. Sin embargo, hacemos aquí la puntualización de que en la práctica se omite dicha pregunta al final de la participación del compañero.

A continuación, el entrevistador le pide al estudiante A que empiece a hablar, no sin antes señalarle que cuenta con un lapso de entre 1 a 3 minutos para la ejecución de la tarea. Concluida la intervención de su compañero, el estudiante B procede a plantear su pregunta. Una vez respondida, el profesor entrevistador le retira la tarjeta y la imagen al estudiante A, y repite el procedimiento con el estudiante B, utilizando una tarjeta y una imagen distintas. Finalmente, antes de terminar, pregunta a los estudiantes si desean agregar algo; de no ser así, informa que la segunda tarea del examen ha terminado.

En la última de las tareas, el protocolo señala que se evalúen tanto la producción, como la interacción (que consiste en la capacidad para interactuar, tomar el turno de habla, cederlo, formular el tema y hacer uso de vocabulario adecuado a la situación comunicativa) de los estudiantes,

mediante un juego de roles. El entrevistador entrega a cada cual una tarjeta con las pautas para desarrollar dicho juego, por ejemplo, un diálogo entre un médico y su paciente. Asimismo, les indica (empleando un nivel de lengua adecuado al de ellos) que cuentan con un minuto para leer la información de la tarjeta y preparar su actuación, y con un lapso de dos a tres minutos para realizar la tarea. Una vez que los estudiantes se declaran listos, el entrevistador da la instrucción para que inicien el diálogo, durante el cual, el entrevistador dirigirá las acciones al marcar los tiempos.

Terminado el diálogo, se recogen las tarjetas y se da por finalizada la prueba. El profesor evaluador, entonces, les devuelve su identificación y el papel con su nombre, fechas y horarios de exámenes con los que ingresaron al salón. Después de que los estudiantes han abandonado el salón, el profesor evaluador y el entrevistador contarán con tres minutos para acordar los puntajes obtenidos por cada estudiante, en las tareas dos y tres de la prueba.

Igualmente, el protocolo hace las siguientes recomendaciones para los profesores entrevistadores: crear un clima cordial; minimizar la actuación propia, incluso cuando se nota asimetría entre los niveles de lengua de los estudiantes, a quienes a su vez, se les debe brindar la oportunidad de que demuestren el conocimiento y uso de estrategias de compensación, el manejo de la interacción y el componente actitudinal del aprendizaje (cooperación, respeto, tolerancia, respuesta ética y moral en conflictos, etc.); procurar tener un control estricto de los tiempos correspondientes a cada tarea; y, en los casos donde los estudiantes no están satisfechos con la tarea o el tema asignados, explicarles que en las situaciones reales de comunicación, no siempre se está en circunstancias de nuestro gusto o bajo nuestro control, pero que las tareas están diseñadas para que ellos demuestren su nivel de dominio de lengua y su capacidad para salir airoso de cualquier situación comunicativa.

3.2. Quiénes evalúan

En este punto, es importante señalar que los profesores convocados para la aplicación de exámenes orales de aprovechamiento, son todos aquellos que impartieron cursos frente a grupo. Para quienes ya han aplicado exámenes con anterioridad, la labor de realizar estos exámenes no requiere de un gran esfuerzo. En muchos casos, algunos profesores fueron capacitados en un curso-taller, pero otros tantos han aprendido la mecánica de este procedimiento directamente mediante su propia experiencia y el consejo de profesores más experimentados.

En otras ocasiones, cuando el elevado número de estudiantes o factores azarosos provocan que haya falta de maestros, se recurre a prestadores de servicio social y a profesores de español recién ingresados al CEPE.

Lo anterior nos hace cuestionarnos sobre la pertinencia de su participación en las evaluaciones, no en cuanto a sus aptitudes sobre la materia, pero sí en lo tocante a sus conocimientos sobre los contenidos que se evalúan en cada curso de español, pues durante una jornada, suelen evaluarse estudiantes de diferentes niveles.

Por ello, estamos convencidos de que es necesario establecer un curso-taller que se imparta al menos dos veces por año, para entrenar a los nuevos profesores y a los prestadores de servicio social, con el fin de que conozcan a detalle los pormenores de nuestros procedimientos de evaluación. Asimismo, un curso-taller sería igualmente necesario para profesores experimentados, pues siempre es necesario estandarizar los criterios (Lugo, 2020) y llegar a acuerdos sobre la manera correcta de evaluar esta habilidad.

3.3. Las rúbricas

Las rúbricas (o baremos) que se utilizan para la evaluación de aprovechamiento de la EO están organizadas en escalas para los niveles básicos (1, 2, 3 y 4 del programa del CEPE, y A1, A2 y B1 del MCER). Se emplean dos rúbricas: una holística y otra analítica. La primera proporciona “una valoración global”, en la que se consideran tanto la dimensión comunicativa (logro del propósito de la tarea) como la precisión lingüística (pronunciación, fluidez, vocabulario y el uso de la gramática). Se compone de cuatro bandas que califican, en orden descendente, del 3 al 0. La escala analítica considera dos dimensiones: la competencia lingüística y la competencia pragmático-discursiva. En la primera, se consideran los rubros de pronunciación, fluidez y entonación, así como el de rango y control léxico, y además el de precisión gramatical; en tanto que en la segunda, se contemplan los rubros del desarrollo del tema y del de cohesión y coherencia. De manera semejante a la escala holística, se tienen cuatro bandas que califican del 3 al 0. Como punto de referencia para que los profesores calificadores identifiquen los contenidos correspondientes a cada nivel, se cuenta con un documento que recoge los contenidos gramaticales de la serie de libros *Dicho y hecho, Español para extranjeros* (del 1 al 7; la serie fue editada por la UNAM en 2012).

Sin embargo, tras nuestro análisis, consideramos urgente solventar de forma pronta la falta de tablas de especificaciones para la EO, que definan no sólo contenidos gramaticales, sino también contenidos conceptuales, actitudinales y procedimentales. De igual modo, pensamos que no deben emplearse rúbricas genéricas para todos los niveles básicos, sino que es necesario contar con rúbricas específicas para cada nivel, lo que propiciará evaluaciones más precisas y confiables.

Un último punto a tratar en esta sección tiene que ver con el sesgo de los profesores. Desde nuestra perspectiva, a partir de los datos recogidos mediante las videograbaciones de los exámenes, existe una tendencia –con toda probabilidad, inconsciente– generalizada entre la mayoría de profesores calificadores, la cual consiste en enfocarse sólo en los aspectos más evidentes (la pronunciación, fluidez, entonación, el rango y control léxico, y la precisión gramatical) que abarca la competencia lingüística. Esto provoca que los aspectos anteriores sean los más “castigados”, y aquellos de la competencia pragmático-discursiva quedan relegados, por no decir olvidados. Creemos que aquí existe un nicho de investigación muy próspero que debe ser explorado a corto plazo.

4. Conclusiones

4.1. Identificación de mejoras posibles

1. La disposición de los profesores y estudiantes, presentada como se observa en la figura 1, tiene su razón de ser. Como explicábamos, los estudiantes se sientan lado a lado, de espaldas al calificador y de frente al entrevistador; sin embargo, en la práctica las posiciones de los participantes son muy variables, según nuestros propios registros audiovisuales. En la mayoría de los casos, las sillas de los estudiantes son colocadas ligeramente hacia el frente del escritorio, donde se encuentra el calificador, mientras que la posición del entrevistador resulta muy cambiante: al lado del calificador, frente al escritorio y de espaldas al calificador, frente a los estudiantes y al escritorio, etc.

La variabilidad en la disposición de los participantes de los exámenes orales trae como consecuencia que, en los casos donde los estudiantes pueden ver al calificador, éste adquiera un peso en el desempeño de aquellos, como factor de distracción o generador de ansiedad. En estos mismos casos, entrevistador y calificador llegan a intervenir en la prueba, resolviendo dudas de léxico, sobre las instrucciones, o bien tratando de estimular la producción de aquellos alumnos cuyo desempeño no se llega a considerar suficiente.

De igual modo, hallamos que al permitir que el entrevistador y el calificador participen simultáneamente como entrevistadores, se provoca que el tiempo de la entrevista se extienda más allá de lo estipulado. Percibimos que esto ocurre en algunas ocasiones, porque a uno de ellos le parece que podría estimular más la producción de los estudiantes mediante preguntas adicionales, ajenas a los exámenes, o porque su personalidad es extrovertida y de forma espontánea participa.

Las causas de este proceder pueden deberse a diversos factores: por ejemplo, en el momento en que hicimos el registro audiovisual de los exámenes orales, identificamos a un docente que no tenía experiencia alguna en la aplicación de las pruebas orales. Se optó por explicarle de forma rápida cómo es el procedimiento, cuáles son los roles de los profesores participantes, los instrumentos y los exámenes, y cómo se llenan las hojas de calificación.

Además del profesor aludido, advertimos también que para ese fin de curso había otro que aplicaba exámenes por segunda vez, e incluso una docente que lo hacía por tercera vez. Debido al poco tiempo que ambos llevaban impartiendo cursos, no contaban con conocimientos suficientes sobre los contenidos de otros niveles, pero se les asignaron algunas parejas de estudiantes, de niveles que aún desconocían. En estos casos, provocados por falta de personal, es comprensible que se recurra a todos los recursos humanos disponibles. Por otro lado, hay profesores que, aun teniendo experiencia o habiendo recibido una capacitación, son laxos al tomar en cuenta el protocolo para la aplicación del examen de EO.

4.2. Propuestas de mejora

Lo anteriormente descrito es producto de decisiones que se toman súbitamente, determinadas sobre todo por las circunstancias cambiantes de cada fin de cursos, por lo cual son comprensibles. Sin embargo, creemos que a partir del análisis que hemos realizado, es posible adoptar algunas medidas.

Para solucionar el caso de profesores que no cuentan con suficiente experiencia en la aplicación de los exámenes orales, y así evitar que evalúen de forma discrecional, proponemos la realización periódica de un curso de estandarización. En éste, se explicaría a todos los participantes cuál es el protocolo de aplicación de la prueba, cuáles los instrumentos de evaluación y los modelos de exámenes; cómo se componen y de qué forma se usan. Finalmente, recomendamos la realización de ejercicios grupales de calificación de producciones orales de estudiantes, grabadas para tal efecto, con la finalidad de uniformar los criterios, y de familiarizar (y refrescar) a los profesores en los contenidos a evaluar en cada nivel.

Para resolver el tema de la variabilidad en la tipología de los modelos de exámenes, proponemos, en primer lugar, partir por la definición de contenidos y la elaboración de tablas de especificaciones para evaluar cada nivel. Como hemos señalado anteriormente, las pruebas de competencia lingüística, comprensión auditiva y comprensión lectora sí fueron diseñadas a partir de este tipo de tablas (Campos, Espinosa y Jurado, 2011). Por el contrario, las pruebas de expresión oral y escrita, han carecido de ellas desde un principio, y sólo se ha contado con las rúbricas o baremos,

además de los modelos de exámenes correspondientes a cada uno de los niveles del programa de español del CEPE.

Esto no es un detalle menor, pues al no contar con los contenidos establecidos en tablas de especificaciones (cuya falta de cierta forma se subsana con los contenidos establecidos en el programa de estudios de español del CEPE y en los índices de la serie de libros de *Dicho y hecho*), los contenidos de mayor relevancia se dejan al juicio de los profesores diseñadores de las pruebas y, en última instancia, al de los aplicadores.

Es por ello que, con respecto al diseño de las pruebas, podemos afirmar que a partir del análisis realizado de los modelos de exámenes, hemos detectado que no existe uniformidad en la tipología de las pruebas; a saber: algunas presentan textos de entrada y otras no; no se cuenta con un número mínimo ni máximo de palabras que uniforme la extensión de dichos textos; las instrucciones incluyen léxico y estructuras gramaticales fuera de nivel; las pruebas a veces elicitan un solo aspecto gramatical, cuando mediante un diseño adecuado, se pueden elicitarse otros tantos aspectos de relevancia para el nivel que se evalúa.

Finalmente, el análisis que aquí presentamos, si bien no ha sido exhaustivo, arrojó datos de gran relevancia que nos han ayudado a reconocer debilidades y fortalezas de nuestros procedimientos para la evaluación de EO. Sin duda, la experiencia se ha ido acumulando y hoy podemos construir sobre lo que opera bien, a la par que, gracias a lo anterior, es factible reorientar aquellos aspectos que no funcionan o requieren ajustes. Es necesario seguir investigando al respecto; en ese sentido, esperamos que este texto sirva para provocar curiosidad y estimular a que más profesores se involucren en el tema.

Referencias

Campos, E. (2005). Certificado de español como lengua adicional (CELA). Nivel intermedio. *Decires*, 7 (7) 56-66.

Campos, E., Espinosa, A. y Jurado, M. (2011). Exámenes de aprovechamiento computarizados. Un paso más en la evolución académica del departamento de Español. *Decires*, 13, 121-158.

CEPE. (2014) *Programa de Estudios de los cursos de Español y Cultura Mexicana*. México: UNAM.

Gómez, M. (2021). Videoconferencia Equidad de género y no discriminación en ELE. Recuperado 15 de febrero de 2021, de CEPE Videos canal: <https://www.youtube.com/watch?v=YMW9k9BiziU&t=696s>

Lugo, I. [CEPE Videos] (04 de noviembre 2020). “La importancia de la estandarización de los calificadoros en los exámenes de expresión”. Ciclo internacional de conferencias sobre evaluación [Archivo de vídeo]. Recuperado de: https://www.youtube.com/watch?v=OqDHgbKsxpM&t=5209s&ab_channel=CEPEVideos

Referencias recomendadas

Consejo de Europa. (2002). Marco Común Europeo de Referencia para las lenguas: aprendizaje, enseñanza y evaluación. Madrid: Ministerio de Educación, Cultura y Deporte. Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf

Instituto Cervantes. (2007). Plan curricular del Instituto Cervantes. Niveles de referencia para el español. Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/

SECCIÓN III.

LA EXPRESIÓN

ESCRITA

CAPÍTULO 6

LA HABILIDAD DE EXPRESIÓN ESCRITA Y SU EVALUACIÓN

Sandra Luz **Cruz Iturribarri**

Jorge **Salinas Hurtado**

Introducción

Como acto comunicativo, la habilidad de expresión escrita realiza acciones que persiguen objetivos en la vida cotidiana; desde redactar breves mensajes de texto, hasta cartas para protestar ante las autoridades. Según refiere Daniel Cassany (1999: 24), el propósito de esta facultad humana es producir ideas que causen algún efecto en el lector y para lograrlo, es importante que el autor refleje un dominio adecuado de la gramática y sepa escoger cuidadosamente el léxico más apropiado.

La investigación sobre esta habilidad ha cobrado importancia a partir de los años ochenta del siglo pasado; desde entonces han surgido diversas líneas. Una de las que consideramos más importante es el cognitivismo, que por un lado se ha dedicado a investigar las características de los escritores expertos, y por otro, ha descrito las dificultades específicas a las que se enfrentan los escritores de L2 en la planificación y revisión del texto (López de la Torre, 2013).

En el CEPE se ha seguido un enfoque comunicativo para la enseñanza de la habilidad escrita y, a partir del nuevo plan de estudios (que entró en vigor desde 2007), se toman en cuenta las recomendaciones del MCER.

La evaluación de esta habilidad adquiere una importancia crítica, ya que muchas veces se tienen que tomar decisiones que impactan el potencial de un empleado o estudiante, según lo bien que se desempeñe en su comunicación por escrito (Weigle, 2002). Desde los inicios de la enseñanza de segundas lenguas, sobre todo en el método “gramática-traducción”, siempre se había considerado la calificación de pruebas escritas o composiciones; no obstante, la evaluación de la producción escrita continúa presentando grandes retos en la actualidad. Esto podría deberse a la diversidad de propósitos para escribir o a los distintos estilos y géneros, sin embargo, la causa principal sigue siendo la subjetividad de los jueces involucrados en el proceso de evaluar (Weigel, 2002). Este será uno de los puntos que se desarrollará con mayor detalle dentro de este capítulo. Asimismo, se tratarán los rasgos fundamentales y los parámetros que se deben tomar en cuenta al trabajar la evaluación de la escritura en el ámbito del aprovechamiento; es decir, cómo saber si el alumno adquirió los conocimientos y habilidades impartidos durante un curso.

1. Conceptos generales

La escritura es tan indispensable como compleja: la necesitamos, por ejemplo, para enviar un mensaje de texto, para redactar un proyecto escolar, para dirigir una carta a nuestros superiores en el trabajo, para quejarnos de un mal servicio y para una enorme cantidad de situaciones más. Sin embargo, hacer uso de ella no es tarea fácil, pues requerimos de mucha precisión para lograr transmitir lo que verdaderamente pensamos. Con este propósito debemos tener en cuenta ciertos rasgos propios de la escritura, tales como la intención, el contexto y el código compartido.

Además, tenemos que ser cuidadosos con la gramática, la variedad de lengua (en este caso, del español) que vamos a emplear, el registro y, especialmente, con el tipo de texto que queremos escribir.

2. Rasgos fundamentales de la escritura

Como decíamos líneas arriba, para poder escribir necesitamos estar atentos a varios aspectos, principalmente a los que enumeraremos a continuación. Tomar en cuenta estos rasgos y llevarlos a cabo, de acuerdo con los objetivos, hará que una persona que se somete a una prueba sea considerada como competente en la habilidad de la escritura. Sin embargo, en cuanto a la creatividad de los alumnos, no debemos ser exigentes, porque no estamos formando escritores, sino que debemos averiguar lo que el aprendiz es capaz de comunicar por escrito para llevar a cabo determinadas tareas (Bordón, 2006).

2.1 Intención

La expresión escrita es una actividad que consiste en usar las palabras para dotarlas del sentido de aquello que queremos transmitir; es decir, la intención de lo que deseamos compartir debe ser clara, no importa el número de palabras que empleemos, y para ello es necesario tener claro el objetivo a comunicar. De acuerdo con Cassany (1993), podemos escribir para explorar intereses personales, por ejemplo, un diario, una agenda o simplemente una lluvia de ideas. Escribimos también para informar algo (cartas, solicitudes, invitaciones, etc.); para crear, generalmente para deleite propio y de los demás, como en el caso de un poema, un cuento, una canción, entre otros textos de este tipo. Redactamos con la intención de exponer información (noticias, artículos científicos, entrevistas, etc.), y también escribimos para influir o persuadir; en este caso, nos referimos a un artículo de opinión o a un anuncio, por mencionar algunos casos.

La intención suele ser más importante que las propias palabras; por ejemplo, una palabra tan sencilla como 'gracias', normal y obviamente la

utilizamos para expresar agradecimiento, pero si la vemos escrita en el bote de basura de una zona de comida rápida, adquiere la intención de comunicar “tira aquí los restos de comida”; mientras que en el cartel de un pedigüeño, tendrá un matiz de petición e indicará “dame dinero”, o “ayúdame”. Claro que no todo es la intención: esta debe estar inmersa en un contexto (Cassany, 1999).

2.2 Contexto

De acuerdo con Reyes (2008), el contexto es la situación en que el texto se da, el lugar donde se produce e interpreta, y tiene que ver con el espacio, el tiempo y las circunstancias del momento. Cuando escribimos, estamos inmersos en nuestro propio contexto, pero no debemos olvidar el contexto del interlocutor.

Veamos un ejemplo de la importancia del contexto: en un salón de clase, los alumnos están haciendo un examen y la maestra escribe en el pizarrón, “10 minutos”; claramente los alumnos sabrán que en 10 minutos deben entregar su examen; sin embargo, en el mismo salón y durante una clase normal, si la maestra escribe lo mismo, los alumnos tendrán que preguntar a qué se refiere, porque la situación de examen no existe y, por lo tanto, estarán fuera de contexto. El contexto, entonces, es fundamental; nos ayuda a escribir menos, siempre y cuando sepamos elegir bien qué decir y qué dejarle al entorno en que se enuncia el discurso. En este ejemplo, no es necesario que la maestra escriba: “jóvenes, en 10 minutos termina el tiempo asignado para nuestro examen y deberán entregarlo, así que deben darse prisa para poder concluir”.

De acuerdo con Cassany (1999), es necesario relacionar cada mensaje con el contexto extralingüístico, es decir, con el emisor, el destinatario, el lugar, el tiempo y la situación. De este modo, se construye el significado por medio de la asociación entre el propósito, el contexto y la palabra.

2.3 Código compartido

El contexto también incluye el conjunto de creencias, información previa y conocimientos compartidos por los lectores. Para que se dé la comunicación, es necesario que se establezca un código común entre los interlocutores; por ejemplo: una palabra tan sencilla como *jitomate* puede impedir la comunicación, si no hablo con un mexicano, especialmente del centro de México, pues su uso es casi exclusivo de ahí; en otros países hispanohablantes, se utiliza *tomate* (sea verde o rojo), así que en este caso, el lugar juega un papel importante.

Veamos ahora un ejemplo asociado al tiempo; en un buscador de internet encontramos la siguiente frase: “Los pies de foto más bizarros del

periodismo”¹. La gran mayoría entenderá este mensaje como “Los pies de foto más raros”; sin embargo, para algunos cuantos el sentido será incomprendible, porque se apegarán al significado de “valiente” o “generoso”, que son las acepciones registradas en el *Diccionario de la lengua española* (2014). Actualmente, el uso de “bizarro” como algo “raro”, es mucho más común y es el que la mayoría de los jóvenes comprenden y emplean. Así lo podemos constatar en el *Diccionario de americanismos* (2010): “Referido a cosa extraña, rara, insólita”.

Si bien este código compartido es mucho más fácil de interpretar como hablante nativo, no importa de qué país de habla hispana se provenga: los hablantes no nativos lo van adquiriendo, conforme avanzan en su práctica del español, ya sea como estudiante o en la convivencia con el idioma. De tal suerte, poco a poco son capaces de interpretar, además, los diferentes significados de las palabras. Recordemos que existen frases idiomáticas: aquellas que adquieren un significado connotativo diferente, y no se pueden traducir palabra por palabra, como “soltar la sopa” (en el sentido de hacer una confesión), “dar en el blanco” (por adivinar o acertar sobre un tema), o “lavarse las manos” (refiriendo el no aceptar una culpa o responsabilidad). Un escritor competente sabrá en qué momento y en qué tipo de texto será adecuado incluir alguna de estas expresiones.

3. Constructos de la escritura

Para comunicarnos por escrito será necesario, entonces, poder dominar los rasgos de la escritura mencionados. Sin embargo, debemos también discernir entre textos de expresión escrita y de interacción escrita.

3.1 Expresión escrita

Cuando escribimos para informar de algo, contar una historia o simplemente por deleite, estamos recurriendo a la expresión escrita. Ejemplo de estos textos son: cuentos, poemas, editorial periodístico, textos narrativos, ensayos académicos, etc.

3.2 Interacción escrita

Como su nombre lo dice, escribimos para entablar reciprocidad, es decir, queremos provocar una reacción en el interlocutor: que se atienda una queja, se acepte una invitación, se responda a una carta, o bien que se acepten los motivos para obtener una beca o un empleo; estas entre muchas otras actividades que esperamos sean atendidas al leer nuestro texto. De acuerdo con el MCER, al cual parafraseamos, las actividades de interacción

1. Recuperado de: <https://www.cuantarazon.com/919193/los-pies-de-fotos-mas-bizarros-del-periodismo>.

son aquellas donde los participantes tienen varios turnos de producción y recepción, en las cuales se lleva a cabo la negociación de los significados.

Es importante mencionar que la interacción escrita se consideraba un proceso mucho más lento que el de la interacción oral, dado que ésta se daba simultáneamente y la escrita, por lo general, no era así; en la actualidad, podemos hablar de una interacción escrita sincrónica, gracias a todos los recursos que se han desarrollado en los últimos años (Facebook, WhatsApp, Twitter, etc.).

4. Evaluación escrita

El primer paso a considerar es el *para qué* se quiere evaluar. Bachman y Palmer (1996) mencionan dos propósitos:

- a) Inferir acerca de la habilidad de la lengua del sustentante.
- b) Tomar decisiones basadas en esas inferencias.

Para lograr las inferencias deseadas, se consideran dos métodos: directo e indirecto. En el indirecto, se elaboran reactivos para medir el conocimiento de estructuras gramaticales y uso de la lengua. Estos reactivos, mediante pruebas de opción múltiple, evaluaban, por ejemplo, el uso apropiado de un tiempo verbal, o el empleo correcto de marcadores discursivos y elementos de cohesión. Aunque el resultado es objetivo, no se considera que este método refleje verdaderamente el nivel de dominio de esta habilidad por parte del sustentante, y en la actualidad ya no es frecuente que se utilice.

En el método directo, normalmente se pide la redacción de un texto continuo, con un mínimo de cien palabras. En el siguiente capítulo se describirán con detalle los instrumentos de evaluación y las tareas; a continuación, se hablará de la subjetividad implicada en este método directo, y además se tratarán los parámetros considerados para la evaluación de la habilidad escrita.

4.1 Subjetividad en la evaluación escrita

El mayor reto que presenta esta evaluación es su subjetividad, ya que al ser realizada por seres humanos, puede presentar cierto grado de imprecisión, proclive a corregirse. Es necesario, por lo tanto, establecer criterios unívocos que permitan la objetivación de los resultados (Eguiluz, 2009).

Con objeto de valorar adecuadamente la producción escrita, se sugiere considerar cuatro aspectos: adecuación (cumplimiento de todos los requisitos pedidos en la tarea), estructura (organización del texto, empleo

de marcadores discursivos y mecanismos de cohesión), gramática (uso correcto de las reglas, incluye el buen manejo de la ortografía y puntuación) y vocabulario (selección adecuada de las palabras). A los calificadores se les proporciona, por una parte, una descripción de estos aspectos, que sea a la vez exhaustiva y fácil de comprender; por otro lado, se les provee de una escala de valoración numérica, obtenida a través de un análisis en bandas (Eguiluz, 2009).

Para minimizar la subjetividad en la evaluación, Weigle recomienda, además de contar con una escala bien definida, asegurarse de que los calificadores la empleen de manera apropiada, por lo que aconseja continuas sesiones de calibración (Weigle, 2002).

En estas sesiones, por lo regular, se reúne a los expertos para proceder a simular la calificación de diferentes pruebas. Se explica claramente cada parámetro de la escala y los alcances de las bandas. Al final de cada calificación, son revisados los números y comentadas las razones para asignarlos. Con esto, busca evitarse que exista una diferencia notable en el puntaje que asignan los calificadores.

En el CEPE, con el fin de evaluar las producciones escritas al final de cada curso, un calificador, diferente al maestro titular, recibe una rúbrica con puntuaciones para distintos aspectos.

4.2 Coherencia

Este criterio sirve para medir la claridad de las ideas expuestas en el escrito. En otras palabras, la coherencia ayuda a la comprensión del mensaje que busca transmitir el texto; para lograr un mejor resultado, se debe tomar en cuenta la manera en que las ideas se ordenan y distribuyen a lo largo del mismo. El escritor competente debe cuidar que exista la interrelación adecuada entre cada una de las partes del texto (Eguiluz, 2009).

Cuando en un texto, el escritor empieza a hablar de un tema y las ideas siguientes no tienen relación con éste, se habla de una falta de coherencia; otro caso es cuando, para justificar una postura, emplea argumentos que no funcionan para sostenerla.

4.3 Cohesión

La cohesión involucra la interrelación adecuada entre cada una de las partes del texto. Cuando se empieza a escribir en una segunda lengua, el primer paso es redactar ideas aisladas; conforme se avanza en el dominio del idioma, se aprende el uso de los conectores que permitan hilar un discurso más fluido, que logre una mejor conexión entre esas ideas. En la cohesión se cuida que sean respetadas las referencias internas, como la

repetición de elementos en el texto, el uso de pronombres con valor referencial y, además, se espera un uso correcto de los signos de puntuación (Eguiluz, 2009).

Al escribir en una L2, si se está en un nivel inicial, el redactor escribirá frases y oraciones aisladas; sin embargo, conforme vaya progresando, aprenderá el uso de los conectores, empezará con los más simples: “y”, “o”, “pero”, y ya en un nivel avanzado, empleará otros más sofisticados, tales como: “no obstante”, “si bien”, entre otros. A medida que vaya logrando una mayor unidad en su texto, se entenderá que está manejando mejor la cohesión.

4.4 Corrección

Este parámetro analiza el adecuado uso de la gramática, desde el punto de vista morfológico, sintáctico y ortográfico. Como en cualquier disciplina, se apreciarán los avances del escritor, por ejemplo, desde un nivel inicial en el que sólo maneja verbos muy comunes, del tipo “ser” o “estar”, y algunos regulares en presente de indicativo; asimismo, es de esperarse que presentará bastantes problemas de concordancia de género y número, hasta llegar a niveles avanzados, donde se verá el correcto uso de los tiempos compuestos del subjuntivo, por ejemplo.

En esta categoría se señalan los errores que aparecen en la expresión. Eguiluz (2009) menciona que pueden distinguir tres tipos de errores:

- a) *Errores básicos que afectan a contenidos gramaticales que ya se consideran dominados para cierto nivel.*

“Me encanto mucho la ciudad”.

En este ejemplo, si el escritor es de un nivel B1, presenta un error en la conjugación de los verbos de construcción inversa, que se supone ya debió haber superado en los niveles iniciales.

- b) *Errores gramaticales que afectan la comunicación y que pueden entorpecer la comprensión del texto.*

“No, me no tengo comprar”.

En este ejemplo, no se entiende lo que desea comunicar el escritor y, por tanto, se entorpece por completo la comunicación.

- c) *Errores gramaticales que corresponden a conocimientos de niveles superiores y que surgen al intentar elaborar construcciones más complejas.*

“Yo pienso mucho que sea bien para un niño recibir populares o no”.

En este ejemplo se observa que el escritor no domina la construcción de oraciones subordinadas con verbos de opinión (Eguiluz, 2009).

4.5 Alcance

Con este parámetro se evaluará la corrección en el uso del vocabulario, tanto de palabras como de frases. Se toma en cuenta la variedad o riqueza léxicas del texto, pero se deben considerar también las limitaciones que tiene el sustentante, según su nivel de dominio de la segunda lengua; por lo tanto, el grado de exigencia variará respecto a los recursos léxicos que estén contemplados de acuerdo al nivel. Se debe considerar que se recurre a estrategias para evitar el empleo de cuestiones complejas y para expresar la misma idea de una manera más simple.

Eguiluz divide los tipos de errores en dos:

- a) *Incorrecciones de tipo léxico que dificultan la comprensión del texto. Por ejemplo, al enumerar en su escrito los ingredientes para preparar un platillo típico de su país, escribe: "ingredientes para la plato es spagetti, aciete y tamati".*
Como podemos apreciar, hay una falta de concordancia (la plato), palabras mal escritas (espagueti y aceite) y, por último, "tamati", que no existe en español; todo esto provoca que sea difícil entender la redacción.
- b) *Imprecisiones, vocabulario correcto pero inapropiado para el contexto.*
"El chavo que estaba en la calle..."
Aunque la palabra "chavo" es de uso generalizado en México en un registro coloquial, no se considera apropiada para un escrito formal (Eguiluz, 2009).

4.6 Adecuación o cumplimiento

Para algunos, este parámetro es nombrado "cumplimiento" y lo que pretende evaluar son determinadas características propias del texto solicitado; es decir, si se pidió redactar un artículo de opinión, se busca que el escrito responda a los modelos que corresponden a esta tipología textual.

El registro que se use, debe ser adaptado a la situación comunicativa planteada (presentarse, agradecer, solicitar, quejarse, disculparse, entre otras). Con este criterio, se pretenden medir determinadas características, propias del texto solicitado (un artículo de opinión, por ejemplo, responde a una tipología concreta y, por tanto, es necesario que la expresión escrita se ajuste a este modelo). El registro utilizado en el texto tiene que adaptarse a la situación comunicativa planteada, al objetivo y al público destinatario del artículo.

Por otra parte, el contenido del escrito debe cumplir el objetivo comunicativo propuesto. En este sentido, la elaboración del artículo debe respetar las pautas de orientación dadas al inicio de la tarea. Se entiende

que si estas pautas cumplen y se ajustan a la extensión pedida, el texto ha cumplido el objetivo y es adecuado (Eguiluz, 2009).

Conclusión

Como en todas las áreas del conocimiento, se siguen investigando y proponiendo nuevos y mejores métodos, tanto de enseñanza, como de evaluación. Parte importante de estos cambios es que las tareas solicitadas a quienes aprenden una segunda lengua, deben ser de actualidad. Como mencionamos, la escritura ha evolucionado gracias a la tecnología y ahora puede ser sincrónica; por lo tanto, será necesario adecuar las tareas de enseñanza y evaluación a este tipo de escritura.

Por supuesto, la gramática es y será fundamental para comunicarnos; sin embargo, el vocabulario empleado (o código compartido), junto con el contexto y la intención son fundamentales para lograr este propósito. Todo ello debe verse reflejado, tanto en el momento de diseñar una tarea, como en las rúbricas para evaluarla.

Referencias

Bachman, L. y Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bordón, T. (2006). *La evaluación de la lengua en el marco de E/L2: bases y procedimientos*. Madrid: Arco Libros.

Cassany, D. (1993). *La cocina de la escritura*. Barcelona: Empúries.

Cassany, D. (1999). *Construir la escritura*. Barcelona: Paidós.

Consejo de Europa. (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. España: Ministerio de Educación, Cultura y Deporte. Recuperado 14 de abril de 2021, de: <http://cvc.cervantes.es/obref/marco/>

Eguiluz, J. (2009). "Criterios para la evaluación de la producción escrita", En *Monográficos marco ELE*. Recuperado 12 de abril de 2021, de: https://marcoele.com/descargas/expolingua1996_eguiluz-vega.pdf

López de la Torre, J. A. (2013). *El Portfolio Europeo de las Lenguas (PEL): herramienta de evaluación de la composición escrita*. Valencia: Universitat de Valencia.

Reyes, G. (2008). *Cómo escribir bien en español*. Madrid: Arco Libros.

Weigle, S. C. (2002). *Assesing Writing*. Cambridge: Cambridge University Press.

Bibliografía recomendada

ASALE. (2010). *Diccionario de americanismos*. Madrid: Santillana.

ASALE. (2014). *Diccionario de la lengua española*. Madrid: Real Academia Española.

Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

CAPÍTULO 7

LINEAMIENTOS PARA LA CREACIÓN DE INSTRUMENTOS PARA LA EVALUACIÓN DE LA EXPRESIÓN ESCRITA

Sandra Luz **Cruz Ituribarri**

Jorge **Salinas Hurtado**

1. Introducción

En el capítulo anterior, fueron revisados los aspectos más importantes a considerar en la habilidad escrita. Ahora, la atención se centrará en un punto mucho más práctico: la mejor manera de evaluarla, así como brindar un panorama general de los tipos de evaluación.

A veces se asume que el primer paso es sencillamente *poner a las personas a escribir*; sin embargo, inmediatamente surge la necesidad de determinar algunas cuestiones básicas:

- ¿Qué? (lo que queremos evaluar), ¿a quién? (el perfil de las personas que evaluaremos), ¿para qué? (nuestro objetivo) y por último, ¿cómo? (los instrumentos con que lo vamos a hacer).

Por otra parte, es necesario contar con dos componentes fundamentales:

- a) Una o más tareas de escritura, o al menos algunas pautas que indiquen qué escribir.
- b) Un medio para evaluar las muestras que resulten (Weigel, 2002).

Diseñar una prueba involucra mucho más que pensar un tema (describe tu ciudad, narra una anécdota divertida durante un viaje, expresa tus razones a favor o en contra de evitar el consumo de productos animales) y usar nuestro criterio para calificar los escritos. A continuación, se describirán con mayor detalle las consideraciones que se deberán contemplar, en cuanto a los instrumentos y los lineamientos que ayudarán a una evaluación válida y confiable de la habilidad escrita.

2. Características de los instrumentos de evaluación de la escritura con base en su objetivo

Antes de tomar alguna decisión acerca de la tarea o del procedimiento para calificar, es recomendable reflexionar sobre las siguientes preguntas:

- a) ¿Qué se quiere medir? Es decir, cómo se define la habilidad escrita para el propósito de la prueba. ¿Interesa principalmente si los escritores dominan ciertas estructuras gramaticales, por ejemplo, el uso

del imperativo para dar recomendaciones, o lo que importa es qué tan bien manejan la escritura para cumplir con una función comunicativa específica, como describir, narrar, argumentar?

- b) ¿Por qué se quiere medir la habilidad escrita? ¿Qué se va a hacer con la información obtenida de la prueba?
- c) ¿Quiénes van a presentar la prueba? ¿Qué necesitamos saber de ellos, con el propósito de diseñar tareas que les permitan desempeñarse con su máximo potencial?
- d) ¿Quién va a calificar las pruebas y qué criterios o estándares van a usarse? ¿Cómo es posible asegurar que los jueces aplican dichos estándares de manera consistente?
- e) ¿Quién y de qué manera va a usar la información proporcionada por la prueba?
- f) ¿Qué limitaciones existen (tiempo, dinero, materiales y fuerza laboral) para administrar el tipo y cantidad de información que surjan a través de las pruebas aplicadas?
- g) ¿Qué necesitamos saber para que nuestra prueba sea válida y confiable? (Weigel, 2002).

Y, a su vez, todas las respuestas tendrán que relacionarse con el tipo de evaluación que haremos. A continuación, daremos una breve descripción de los cuatro principales tipos.

2.1. Instrumentos de evaluación diagnóstica de la escritura

Como su nombre lo indica, este tipo de evaluación nos da un panorama general del nivel de conocimientos de los sustentantes. Se aplica al principio de un curso y nos proporciona información que nos permitirá adecuar los contenidos a los conocimientos de los alumnos. O bien, hablando concretamente de lenguas extranjeras, nos facilitará colocar a los estudiantes en el nivel adecuado, de acuerdo con los conocimientos demostrados.

Dependiendo del nivel, se diseñan las tareas; por ejemplo, para nivel A1, sería pertinente pedir que los estudiantes hagan una muy breve presentación de ellos mismos (nombre, nacionalidad, estado civil, etc.).

Hay muchas formas de diseñar rúbricas, dependiendo de las especificaciones de los exámenes; una muy práctica es dividirla en dos partes.

- a) Holística: en la que se evalúa qué tan bueno fue el cumplimiento en general de la tarea.
- b) Analítica: en la que se evalúan por separados los distintos aspectos que se requiera evaluar; por ejemplo, la competencia pragmático-discursiva (coherencia y cohesión dentro del texto) y la competencia

lingüística (léxico, gramática, ortografía, puntuación).

Según las necesidades para calificar, se asignan bandas en las que son descritos los niveles de desempeño, cuyos valores por lo general van del 0 al 5; si bien pueden ser menores, según el criterio de la institución. En algunos casos, se emplean tanto la holística como la analítica.

2.2. Instrumentos de la evaluación formativa de la escritura

La evaluación formativa, también conocida como evaluación para el aprendizaje, debe hacerse a lo largo de todo un curso y su finalidad es apoyar tanto a los alumnos como a los profesores a detectar aquello que se está haciendo de manera adecuada; qué problemas hay que solucionar y cuáles son las áreas de oportunidad. El estudiante recibe una adecuada retroalimentación por parte del profesor, que le permite responsabilizarse de su conocimiento y mejorar aquello que no está funcionando; así como continuar con lo que está haciendo de manera adecuada.

Tradicionalmente, en la calificación de los escritos se da una “retroalimentación correctiva”, es decir, se marcan los errores cometidos al escribir; generalmente verbos mal conjugados, falta de concordancia de género y número, uso incorrecto de los artículos, etc. Sin embargo, lo que se sigue cuestionando es qué tan efectiva es esa retroalimentación y si en verdad provoca una mejora en la habilidad escrita. Aún no se ha encontrado una respuesta definitiva (Lee, J., 2003).

Por otra parte, hay investigaciones que muestran que cuando las retroalimentaciones se enfocan en el contenido, sí tienen un mayor impacto y arrojan evidencia de que los estudiantes mejoran considerablemente su habilidad para escribir. Se recomienda que el calificador negocie el significado con el escritor y lo motive a expresarse de la manera más clara posible (Lee, J., 2003).

De acuerdo con Sánchez y Martínez (2020), para brindar una retroalimentación eficaz, debemos seguir los siguientes consejos:

1. Establecer un ambiente de aprendizaje respetuoso.
2. Comunicar los objetivos y metas para la retroalimentación.
3. Retroalimentar a partir de la observación directa.
4. Brindar retroalimentación oportuna, a tiempo y de forma regular.
5. Comenzar la sesión con la autoevaluación del alumno.
6. Reforzar y corregir los desempeños observados.
7. Usar un lenguaje neutral para centrarse en el conocimiento, la conducta o el procedimiento específico.
8. Confirmar la comprensión del alumno y facilitar la aceptación.

9. Concluir con un plan de acción.
10. Reflexionar sobre las habilidades de retroalimentación.
11. Crear oportunidades de desarrollo personal.
12. Hacer de la retroalimentación una cultura institucional.

2.3. Instrumentos de la evaluación del aprovechamiento de la escritura

En este caso, debemos basarnos en el tipo de conocimientos que el alumno ha recibido durante un curso (gramática, léxico, conectores, etc.) y ése será el aprendizaje que debemos solicitarle al finalizar el mismo. Los exámenes de aprovechamiento pueden ser parciales, finales o ambos.

Algunos autores consideran que es una de las tareas más difíciles, porque no implica únicamente el conocimiento de la gramática o el léxico: el escribir conlleva un proceso de análisis más complejo, que puede representar mayores obstáculos para quien escribe. De acuerdo con Antonio Mendoza (1990), algunas de las dificultades más comunes, y que debemos tomar en cuenta al diseñar un examen de aprovechamiento de expresión escrita, son las siguientes:

1. Limitaciones en la aplicación del código.
2. Interferencias con la expresión oral.
3. Falta de dominio del léxico, la sintaxis y las normas de puntuación, lo que puede provocar falta de precisión o errores en la interpretación.
4. Dificultades en la estructuración del texto, organización, jerarquización de ideas y temas.

Estas, entre otras dificultades. Si un alumno no escribe correctamente en su lengua materna, le será doblemente complicado escribir en una segunda.

2.4. Instrumentos para la certificación de la escritura

La certificación tiene como objeto dar por cierto el dominio, en este caso, de la escritura en una segunda lengua. Es independiente de un curso, pues sabemos bien que una segunda lengua no solamente se adquiere en un sistema escolarizado.

Generalmente, un certificado de este tipo se solicita para ingresar a un trabajo, una escuela, para recibir una beca, etc. Entre los instrumentos de certificación del español más reconocidos están: DELE (Diploma de español lengua extranjera, del Instituto Cervantes), CELU (Certificado de español lengua y uso, diversas universidades argentinas), DUCLE (Diploma universitario de competencia en lengua española como lengua extranjera, Universidad Nacional de Rosario, Argentina), CELA (Certificado de español

como lengua adicional, UNAM), EXELEAA (Examen de español como lengua extranjera para el ámbito académico, UNAM) y SIELE (Servicio internacional de evaluación de la lengua española, Instituto Cervantes, Universidad de Salamanca, Universidad de Buenos Aires y Universidad Autónoma de México). Todos ellos con características muy variadas; por ejemplo, el DELE no tiene vigencia, en tanto que el SIELE tiene vigencia de cinco años y el EXELEAA, de dos. La vigencia de un certificado de lengua es de suma importancia, pues todos sabemos que si practicamos constantemente una lengua, podemos incluso mejorar el nivel de dominio, o al menos conservar el mismo; sin embargo, si deja de practicarse, sucede lo contrario: las habilidades se van perdiendo. Por esta razón, recomendamos que siempre se tenga una vigencia mínima de dos años.

3. Lineamientos para elaborar un instrumento de evaluación escrita

El primer paso en la elaboración de un instrumento es, como se comentó al inicio de este capítulo, definir el propósito del mismo, a partir de responder las siguientes preguntas: *¿qué?, ¿cómo?, ¿a quién? y ¿para qué?*

Una vez redactado el propósito, es importante resaltar las palabras clave, aquellas que resulten significativas para su contenido, se deberá escribir su definición con el objeto de contar con un marco teórico que sustente el examen.

La función principal y el propósito de los exámenes es ayudar a la creación de una tabla de especificaciones que, como se explicará más adelante, es la base de cualquier examen.

3.1. Planeación y objetivos

El primer paso en la planeación teórica es tomar en cuenta el programa o plan del curso que queremos evaluar. En múltiples ocasiones se pide en un examen que el alumno escriba una carta y la rúbrica toma en cuenta que desarrolle bien todos los elementos de la misma; pero sucede que quizá algunos de esos elementos no se estudiaron dentro del curso, por eso, para planear un examen, debemos apegarnos muy bien a lo que el alumno asimiló y desarrolló en su curso.

Los objetivos del examen deben mostrar de manera clara qué se espera que el alumno haga, y no debemos confundirlos con los objetivos de enseñanza, que son los que nos marcan la pauta de lo que queremos enseñar, aunque de estos derivan los del examen.

Se recomienda seguir las orientaciones del CER, en cuanto a lo que se debe esperar de un alumno, según su nivel. Por ejemplo:

Un alumno de A1 ya escribe frases y oraciones sencillas y aisladas, así que debería ser capaz de redactar párrafos cortos en los que comparta información básica de su ámbito personal.

En A2, ya empieza a usar conectores simples como 'y', 'o', 'pero', así que los puede emplear para ir uniendo oraciones y lograr descripciones más cohesionadas.

El B1, corresponde a los niveles intermedios en la mayoría de los planes de estudio. En este nivel, aunque los textos siguen siendo sencillos, están mejor cohesionados y abarcan una variedad de temas de la vida cotidiana. El manejo de más tiempos verbales le permite al sustentante escribir narraciones sobre sus propias experiencias.

Los textos de un estudiante de B2 son más claros y más detallados. En ellos se expresa sobre una gran variedad de temas que tengan relación con su especialidad; sintetiza y evalúa información y argumentos que proceden de diferentes formas. Es común que en este nivel se soliciten artículos de opinión y ensayos argumentativos.

El C1 se considera un nivel avanzado y en muchas instituciones corresponde al último curso de su plan de estudios. Los textos cada vez son más claros y presentan una estructura más compleja; tratan temas especializados, los candidatos presentan su postura, resaltan las ideas principales; expresan argumentos que justifican, amplían y defienden sus puntos de vista, ofrecen ideas complementarias, presentan motivos y ejemplos adecuados, y son capaces de finalizar mediante una conclusión apropiada.

Cuando un alumno alcanza el nivel C2, es muy difícil distinguir sus textos escritos de los de un hablante nativo. Son complejos, claros y fluidos; muestran un estilo apropiado y eficaz, así como una estructura lógica que permite al lector encontrar las ideas significativas (Consejo de Europa, 2002).

3.2. Definición de constructos

Una vez que tenemos claros los lineamientos y objetivos, debemos definir cada uno de los términos importantes. Por ejemplo, si hablamos de expresión e interacción, hemos de definir qué entendemos por cada uno de esos términos.

3.3. Elaboración de tablas de especificaciones

Una vez definido el propósito en la tabla de especificaciones se establecerán los siguientes conceptos:

- a) Acción: lo que se desea lograr que realice el sustentante; narrar, exponer, argumentar, describir, etcétera.
- b) Condición: determinar qué es lo que va a propiciar la producción escrita, si ciertas pautas, un “input contextualizado”, preguntas, etc.
- c) Suficiencia: saber qué es lo que confirma que el candidato cumplió con la tarea, por ejemplo, un texto descriptivo de 100 palabras.

Con estos tres conceptos, se redacta el componente que será sometido a una posterior validación, de acuerdo con su pertinencia, claridad, especificidad, inclusión o integralidad.

En la tabla también se acordarán los siguientes parámetros:

- a) Nivel taxonómico: no todas las habilidades cognitivas tienen el mismo grado de complejidad, de tal manera que, autores como Bloom y Marzano, las jerarquizaron en una taxonomía que nos permite clasificar estos conocimientos, por ejemplo:

Aplicar

De acuerdo a la taxonomía de Bloom, revisada por Anderson y Krathwohl (2014), este verbo lo describen como “usar en una nueva situación, resolver problemas mediante la aplicación de conocimiento, hechos o técnicas previamente adquiridas en una manera diferente”. Cuando un sustentante resuelve una tarea relacionada con la habilidad escrita, el alumno aplica además de sus conocimientos de gramática y léxico, sus competencias pragmático-discursiva, sociolingüística e interculturales con objeto de resolver la situación planteada (Centro del profesorado Tenerife Sur, 2017).

- b) Tipo de conocimiento: **procedimientos**
Los procedimientos se refieren a las destrezas involucradas en el acto de escribir; las principales que emplea el sustentante en su producción escrita, con el propósito de cumplir la tarea lo mejor posible, son: 1) planificar (tener claro sobre lo que va a escribir, cómo va a organizar sus ideas), 2) revisar su trabajo con el objeto de detectar posibles errores, así como otras posibilidades para expresarse mejor, y 3) editar para hacer los cambios correspondientes.
- c) Nivel de relevancia: valorado del uno al cuatro, en donde uno significa que es un conocimiento muy poco relevante, y cuatro es completamente relevante. Si un conocimiento es valorado por los expertos con dos o uno, tal vez no sea necesario medirlo.

d) Nivel de dominio: valorado del uno al diez, esto nos indica el número de errores que podemos permitir en un reactivo. Por ejemplo: si quiero que el alumno identifique el pospretérito y elabore diez reactivos al respecto, ¿en cuántos aceptaría que se equivoque? Si el nivel de dominio asignado es ocho, estamos indicando que solamente permitiríamos que se equivoque en dos de diez reactivos. Si asignamos siete o menos, probablemente no será necesario medir ese conocimiento; esto lo decidiremos, finalmente, al cruzar esta columna con el nivel de relevancia.

3.4. Elaboración de pruebas y tareas

Una vez que hemos terminado con la tabla de especificaciones, podemos desarrollar los reactivos o las tareas. En el caso de la expresión e interacción escritas, siempre será más conveniente trabajar con tareas, de acuerdo con el nivel de dificultad alcanzado, veamos un ejemplo:

Tabla de especificaciones de expresión escrita, nivel A2			
Acción	Condición	Suficiencia	Componente
Describir actividades a futuro.	A partir de un correo electrónico de 50 a 60 palabras, redactar una respuesta.	Texto descriptivo de 100 a 120 palabras.	A partir de un correo electrónico de 50 a 60 palabras, redactar un texto descriptivo de 100 a 120 palabras, como respuesta.

Tarea:

Lee el correo electrónico de un amigo y redacta una respuesta. En tu texto debes:

- Saludar
- Responder a sus preguntas
- Decir qué actividades vas a hacer en vacaciones
- Despedirte

¡Hola!, ¿cómo estás?

¿Ya te cambiaste de casa?, habla de eso. ¿Cuándo podemos ir a tomar un café?

Ayer vi a tu hermano y me dijo que vas a viajar por vacaciones, háblame de tus planes, ¿a qué lugar vas?, ¿cuántos días?, ¿qué vas a comer?, yo también tengo vacaciones así que tal vez podemos ir juntos.

Escribe pronto.

Manuel

Figura 8. Ejemplo de tabla de especificaciones de expresión escrita

Como podemos observar, la tarea se apega por completo a la tabla de especificaciones: las pautas propician que se describan actividades en el futuro; el vocabulario del *input* corresponde a A1 y A2, pero no rebasa esos niveles, y el número de palabras se ajusta a lo establecido.

3.5. Validación y experimentación

Ya que hemos elaborado las tareas, debemos entregarlas a un comité de validación para que las revise con detenimiento y haga todos los comentarios y las sugerencias necesarios. Después de una primera revisión, corregiremos las tareas y volverán a pasar por un segundo proceso de validación; tras esto, se hacen los ajustes necesarios y pasa a la etapa de experimentación.

Recordemos que después de la experimentación, se vuelven a hacer correcciones en aquello que no esté funcionando, esto puede incluir modificaciones a las mismas tablas de especificaciones.

4. Conclusiones

A lo largo de este libro se ha enfatizado la importancia de contar con una evaluación objetiva, válida y confiable. El propósito de este capítulo al tratar, de manera breve y práctica, los lineamientos para evaluación de la producción escrita es proporcionar al lector las herramientas necesarias, ya sea para crear los propios en un futuro, o simplemente poder determinar si los que está empleando en ese momento son los más adecuados.

Como se ha visto, la creación de tareas implica mucha reflexión y buen juicio acerca de todos los factores involucrados. Se trata de una labor colegiada que requiere la validación de colegas expertos, ya que debemos recordar siempre que los resultados de estas evaluaciones pueden tener consecuencias significativas en la vida profesional y personal del sustentante.

Referencias

Area Sacristán, E. (2014). Anderson y Krathwohl – Revisando la Taxonomía de Bloom. Eduarea. Recuperado 14 de abril de 2021, de: <https://eduarea.wordpress.com/2014/11/09/anderson-y-krathwohl-revisando-la-taxonomia-de-bloom/>

Centro del profesorado Tenerife Sur (2017). *Kit básico para evaluar y calificar*. Obtenido de La taxonomía de Bloom, una herramienta imprescindible para enseñar y aprender: Recuperado 12 de abril de 2021, de: <http://www3.gobiernodecanarias.org/medusa/edublog/cprofestenerifesur/?s=taxonomia>.

Consejo de Europa. (2002). *Marco común europeo de referencia para las lenguas*. Aprendizaje, enseñanza y evaluación. Recuperado 29 de octubre de 2020, de: https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf

Lee, J. y. VanPatten, B. (2003). *Making Communicative Language Teaching Happen*. Nueva York: McGraw-Hill.

Mendoza, A. (1990). "La evaluación de la expresión escrita: una propuesta de aplicación": Universidad de Cádiz. Recuperado 27 de octubre 27 de 2020, de: <https://rodin.uca.es/xmlui/bitstream/handle/10498/7576/14030652.pdf?sequence=1&isAllowed=y>

Sánchez, M. & Martínez, A. (2020). *EVALUACIÓN del y para EL APRENDIZAJE: instrumentos y estrategias*. México: UNAM.

Weigel, S. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.

Referencia recomendada:

Méndez, M. (2015). "La taxonomía de Bloom, una herramienta imprescindible para enseñar y aprender". Gobierno de Canarias. Recuperado 8 de noviembre de 2020, de: <http://www3.gobiernodecanarias.org/medusa/edublog/cprofestenerifesur/2015/12/03/la-taxonomia-de-bloom-una-herramienta-imprescindible-para-ensenar-y-aprender/>

CAPÍTULO 8

ANÁLISIS DE LOS ACTUALES PROCEDIMIENTOS E INSTRUMENTOS DEL CEPE PARA LA EVALUACIÓN DE LA EXPRESIÓN ESCRITA

Marcela **Nieto Rodríguez**
Edgar René **Pacheco Martínez**

1. Introducción

En este capítulo se presenta un análisis de los instrumentos del CEPE para evaluar el aprovechamiento de expresión escrita como lengua extranjera en inmersión, de los niveles de Español 1, 2, 3 y 4. Se ofrecerá una descripción general de cada uno de dichos instrumentos, así como el análisis del procedimiento de evaluación que realizan los profesores con base en la prueba escrita y la rúbrica. Al final, se expondrán las fortalezas y debilidades de los instrumentos y las acciones que se sugieren con base en los resultados del análisis.

2. Los instrumentos de evaluación de EE de Español 1, 2, 3 y 4

2.1. Qué evalúan

De acuerdo con lo que se ha establecido en el capítulo 2 sobre el diseño del constructo de un instrumento de evaluación, retomamos la importancia de este elemento para enfatizar que garantiza el rigor del proceso evaluativo de la expresión e interacción escritas. De igual modo, subrayamos la importancia de especificar el contenido de las pruebas y de los exámenes en tablas de especificaciones. Sin embargo, carecemos de esas tablas, que son insustituibles cuando se construye un instrumento. Así que, todo el análisis que se expondrá en este capítulo, será presentado con base en seis elementos orientadores:

1. Los contenidos de los libros *Dicho y hecho, español para extranjeros 1, 2, 3 y 4*, que se emplean en las clases del CEPE.
2. Lo que evalúa cada una de las tareas de los instrumentos de evaluación (de acuerdo con nuestras observaciones y el foco en los objetivos).
3. Las respuestas de las pruebas de expresión escrita de los alumnos del Curso intensivo de otoño 2 de 2019.
4. Las diferentes anotaciones y marcas de los profesores calificadoros, que asumimos como muestra de su experiencia en la materia (por ejemplo subrayados, circulados, tachaduras, *palomeos*; inserción de letras o palabras en la redacción del alumno, pequeñas notas que

aclaran la decisión sobre el puntaje del calificador, marcas que el calificador indica como un cambio en la sintaxis del texto del alumno, signos de interrogación que suponemos son para indicar que no se entiende algún detalle, etc.).

5. Los criterios, es decir, la escala con la que los calificadores califican las pruebas.
6. Las hojas en las que se asientan los puntajes de dichos calificadores.

En otras palabras, tomamos en cuenta todo lo que involucra este proceso de evaluación al final de un curso de español para extranjeros en el CEPE, pero como punto central, el examen de aprovechamiento. Aun así, este análisis está en cierta forma incompleto, ya que no contamos con los documentos de planeación y diseño de los exámenes de aprovechamiento de ninguno de los niveles que analizamos. No obstante, pensamos que el resultado de este análisis nos ayudará a reflexionar sobre la importancia de evaluar de forma válida y confiable para satisfacer objetivos y atender necesidades.

2.2. Descripción de los instrumentos

Los instrumentos de la expresión escrita de los niveles 1, 2, 3 y 4 de español que se presentan aquí se utilizaron hasta el último curso intensivo de 2019. En Español 1 se evalúa con base en 2 modelos; en Español 2, con 5 modelos; en Español 3, con 3 modelos; y en Español 4, con 5 modelos de instrumentos. ¿A qué se deberá que exista esta variabilidad de número de modelos por nivel? ¿Por qué hay más modelos para evaluar Español 2 y 4, y menos para Español 1, con tan sólo 2 modelos? No lo sabemos, pero, ¿quizá podemos suponer que ese nivel de lengua se presta más para diseñar pruebas de este tipo? De cualquier modo, veremos a qué se debe este panorama conforme avance el análisis.

Conforme a lo que se explica en el capítulo 7, sobre los lineamientos para elaborar un instrumento de evaluación escrita (punto 3), hemos observado que al no contar con los documentos que detallan el propósito de las evaluaciones escritas (a través de las interrogantes ¿qué?, ¿cómo?, ¿a quién? y ¿para qué?) de las pruebas de estos 4 niveles, nos vemos en la necesidad de analizar cómo están conformadas las tareas, es decir, las partes que integran cada una de ellas: las instrucciones, algunas pautas y, en ciertos casos, el texto de entrada como estímulo para desarrollar la tarea.

A continuación, describimos cada una de las versiones de los instrumentos a las cuales les hemos asignado un número de modelo; si bien, por cuestiones de confidencialidad, no se muestran.

Español 1:

Modelo 1: en éste se da una situación para que el alumno invite a un amigo a una fiesta, bajo un contexto específico, y para explicar cómo es esa fiesta, qué hay y qué pueden hacer; esto con un mínimo de 60 palabras.

Modelo 2: en éste no se proporciona el contexto de la tarea, sólo le es solicitado directamente al alumno lo que tiene que desarrollar con base en cuatro pautas sobre el tema del clima (preguntas comunes del nivel), con un mínimo de 50 palabras. En este modelo se piden menos palabras que en el modelo 1; la diferencia es de 10, que si bien no es tan significativo, podría afectar el desempeño del alumno al restringir el número de palabras en la tarea.

Español 2:

Modelo 1: en éste se da una situación específica, el alumno debe hacer una descripción sobre las bodas en su país. Este debe escribir un correo para desarrollar la tarea, dar información y contestar algunas interrogantes que se presentan de forma indirecta en la misma base del planteamiento. La tarea debe contestarse con un mínimo de 100 palabras.

Modelo 2: éste inicia con una breve situación, las instrucciones y un correo como texto de entrada, en el que se presentan algunas preguntas indirectas de un amigo que trabaja en el extranjero. La tarea debe completarse con un mínimo de 100 palabras.

Modelo 3: a partir de una situación breve y una fotografía de varios coches en una calle inundada y con mucha lluvia, se pide describir lo que pasó en una ciudad, de acuerdo con 4 preguntas directas, en forma de listado. Se pide un mínimo de 100 palabras.

Modelo 4: éste es sumamente corto y plantea una situación breve donde se hace una invitación al alumno, para que éste la rechace y explique por qué. Debe desarrollar la tarea dando algunos ejemplos, con un mínimo de 100 palabras.

Modelo 5: a partir de una situación que recrea la venta de un inmueble, se pide escribir un correo para describir el proceso de compra-venta, contestando a cuatro preguntas indirectas que se formulan en las instrucciones. Se solicitan un mínimo de 100 palabras.

Español 3:

Modelo 1: en éste se solicita escribir sobre la vida de alguien importante para el alumno, ya sea una persona real o ficticia. Para desarrollar la tarea, de forma explícita, le es requerido al sustentante que use dos tiempos verbales del pasado: pretérito y copretérito, con cinco pautas, que en

forma de listado consideran datos importantes para desarrollar la tarea. El alumno debe escribir la tarea con un mínimo de 150 palabras.

Modelo 2: aquí se plantea una situación sobre el posible uso de un presupuesto para que el alumno desarrolle la tarea. Este debe dar tres opciones y además argumentar por qué está a favor de esa opción; también tiene que considerar cuatro imágenes (que sirven de estímulo) y puede usar cinco expresiones (que aparecen en un listado) para dar orden y sentido al texto. Se debe escribir la tarea con un mínimo de 150 palabras.

Modelo 3: en éste se pide redactar una aventura de dos personajes, un perro y un gato. Como estímulo se presenta una secuencia de acciones mediante seis fotografías de los dos animales. No hay pautas que deba seguir el alumno, como en el modelo 1, ni se sugiere usar expresiones, como en el modelo 2. El alumno debe escribir la tarea con un mínimo de 150 palabras.

Español 4:

Modelo 1: aquí el alumno debe escribir su opinión a favor o en contra de la publicidad con base en cuatro pautas listadas. Se solicita desarrollar la tarea con un mínimo de 200 palabras. Debemos decir que hay un error en la tarea, y eso puede confundir al alumno a la hora de hacer su redacción, e incluso también al calificador porque no hay precisión, lo cual repercute en la validez de la evaluación. El error aludido se encuentra en las instrucciones, donde se habla de “expresar su opinión a favor o en contra”, mientras que la tercera pauta dice “las ideas a favor y en contra” (las cursivas son nuestras); con esto, se entiende que, en el primer caso, el alumno debe elegir una de las dos posturas, y en el segundo, que debe desarrollar ambas.

Modelo 2: en éste se plantea una situación breve sobre la posibilidad de ganar un concurso y poder viajar a una playa mexicana. Para obtener este supuesto premio, la dinámica es escribir una historia loca o atrevida, de temática amorosa, con un mínimo de 200 palabras.

Modelo 3: en éste se introduce un texto breve sobre las actividades que hacen los niños en su tiempo libre. El alumno debe escribir su opinión a favor o en contra de dichas actividades, brindando ejemplos. El alumno debe completar la tarea con un mínimo de 200 palabras.

Modelo 4: éste presenta, a manera de noticia, un texto de entrada extenso sobre cinco tradiciones del mundo (que se registran como Patrimonio Cultural de la Humanidad), con el fin de que el alumno desarrolle la tarea con un ejemplo de su país; así, se solicita que mediante un informe, haga una descripción de la tradición, que explique por qué es importante y las medidas necesarias para protegerla. Los requerimientos se dan dentro de

la base de las instrucciones de la tarea. El alumno debe cumplir la tarea con un mínimo de 200 palabras.

Modelo 5: éste presenta como estímulo una frase sobre la idea de la participación de los jóvenes en las redes sociales y su influencia, para que el alumno exprese su postura a favor o en contra. En la base de las instrucciones se pide explícitamente que el sustentante estructure su texto con introducción, desarrollo y conclusiones. Además, que utilice ejemplos y que la tarea cumpla con un mínimo de 200 palabras.

Cabe señalar que los modelos 1 y 5 son semejantes, porque solicitan la opinión del alumno; sin embargo, las cuatro pautas que considera el quinto modelo son diferentes a las que presenta el primero, pues se habla de cómo organizar el texto, con introducción, desarrollo y conclusiones; mientras que en el modelo 1, se solicitan ideas específicas de contenido, la tesis fundamental que defiende, las razones por las que se piensa así, las ideas a favor y en contra, para finalmente llegar a una conclusión.

En la tabla que ofrecemos a continuación, podemos ver la distribución de los modelos por nivel, así como cuántos exámenes fueron aplicados en el último periodo de 2019. Hay que decir que estas pruebas nos las facilitó la Coordinación de Español, porque se aplicaron según la población de alumnos que hubo, y no sabemos si existen otros modelos más.

Antes de presentar dicha relación, vale la pena comentar que algunos modelos aparecen más que otros, es decir que unos se aplican más que otros. En teoría, el sistema escoge el modelo para cada estudiante, por lo que la selección es aleatoria; pero en la práctica, vemos que los números indican que unos modelos aparecen con mucha mayor frecuencia que otros. En Español 1, por ejemplo, aunque hay dos modelos, ¿por qué el modelo 2 aparece 23 veces y el 1, tan sólo 2? Lo lógico sería que hubieran salido, aproximadamente, mitad y mitad. En Español 2, por su parte, los modelos 2 y 5 aparecen con mayor frecuencia que los modelos 1, 3 y 4. En Español 3, el modelo más recurrente es el 3, y muy pocas veces se presentan los modelos 1 y 2; en Español 4, los modelos 1 y 2 son los dominantes (pero tampoco de forma equitativa), mientras que los modelos con menor índice de frecuencia son los 3, 4 y 5. Nos preguntamos a qué se deberá esto: ¿acaso el sistema no está funcionando adecuadamente, o está siendo manipulado de alguna forma? En caso del primer supuesto, sería una tarea primordial revisar cuidadosamente el sistema para hacer los ajustes pertinentes; y en el supuesto de que esté siendo manipulado, nos cuestionaríamos: ¿con qué fin se hace esto?

Veamos la figura en la siguiente página.

Modelo	Español 1	Español 2	Español 3	Español 4
1	2	1	6	12
2	23	29	2	18
3	-	2	32	3
4	-	6	-	1
5	-	15	-	1
Total de exámenes aplicados	25	53	40	35

Figura 9. Tabla comparativa del número de exámenes en que aparece cada modelo

2.3. Cómo y para qué se evalúa

Debemos mencionar que el programa curricular de español del CEPE, que entró en vigor en 2007, fue elaborado con base en el MCER; la división de los grados de dominio de español quedan representados en subniveles, tomando como referencia los 6 niveles que considera el MCER. Es decir, concretamente hablando de los A1, A2 y B1, Español 1 y 2 se corresponden respectivamente con los niveles A1 y A2; por su parte, Español 3 y 4 se compaginan con el nivel B1, pero subdividido en B1.1 y B1.2.

Siguiendo con lo establecido en el capítulo 6, entendemos que el alumno deberá demostrar en su actuación, durante el examen final o de aprovechamiento, las destrezas de que dispone en la expresión escrita (estructurales, gramaticales, léxicas, discursivas, culturales, etc.; es decir, los conocimientos y habilidades adquiridos durante el curso); esto en relación con la tarea que se le propone, bajo un contexto, una intención y un código. El propósito es medir el logro de objetivos definidos en el programa de estudios de cada nivel, así como verificar el perfil de egreso correspondiente.

Al no existir las tablas de especificaciones, situación de la cual hemos hecho mención anteriormente, tuvimos que recurrir a los criterios respecto a la consecución de los objetivos finales de aprendizaje: aquellos relacionados con la evaluación de una actuación concreta del curso de Español 1, 2, 3 y 4 en la expresión escrita. En consecuencia, con base en las instrucciones y en las pautas de cada uno de los instrumentos por nivel, hacemos una inferencia de lo que se pretende medir y, por otro lado, también se

consideró la rúbrica para calificar las pruebas en el punto 3, que describimos al final de este capítulo.

2. 4. Análisis de los instrumentos

En este apartado analizamos los instrumentos de evaluación de los diferentes modelos de Español 1 a 4, tomando como base la descripción del capítulo 7 sobre los procesos metodológicos que garantizan las buenas prácticas en la creación de tareas para la evaluación de expresión escrita; asimismo, lo expuesto en los capítulos 1 y 2, acerca de las características que deben cumplir las tareas para garantizar su validez y confiabilidad.

El primer aspecto que llama la atención sobre los instrumentos que analizamos es que en cada nivel se han seleccionado distintas muestras representativas de las tareas que se espera que el estudiante sea capaz de producir; es decir, se establecen distintas tipologías de tareas en cada nivel como los modelos a seguir para medir lo que el alumno es capaz de hacer en una situación determinada. Esto, desde luego, no es algo negativo; sin embargo, la prueba debe evaluar objetivos de enseñanza previamente establecidos. En palabras simples y llanas: no hay un modelo único que mida lo mismo para cada nivel de lengua, defecto que se observa al no contar con una tabla de especificaciones del examen que registre las funciones lingüísticas que eliciten la competencia comunicativa del alumno, a través del uso de la lengua escrita.

En Español 1, los modelos 1 y 2 presentan distintas actividades de expresión. En el primero es de interacción; en el segundo, de expresión aunque, por este hecho, según la literatura especializada, deberían estar homologadas y no tendría por qué haber diferencias respecto a su contenido. Sin embargo, ambas tareas evalúan diferentes funciones: la primera, invitar a alguien a través de un correo, y la segunda, dar información con la descripción de un lugar; ambas con un requerimiento distinto de palabras mínimas que si bien no es tan significativo, tampoco es igualitario: 60 palabras para el primer modelo y 50 para el segundo. Podemos decir que el modelo 1 se acerca más a una situación comunicativa real y dirigida en la que se da la interacción. Aun así, no tenemos constancia o fe de un documento propio en el contexto del CEPE para conocer cuáles son las características específicas de una tarea que cualquier estudiante de Español 1 podría resolver de forma satisfactoria. Ni tampoco existe asentada en ningún documento la tipología de las tareas de expresión escrita en la que se dé información explícita sobre, por ejemplo, cuántas pautas se deben desarrollar, qué características deben tener dichas pautas, de cuántas palabras debe constar el texto producido por el estudiante, etc.

Por último, podemos decir que los dos modelos evalúan diferentes contenidos y temáticas, por lo que no hay coincidencia de recursos lingüísticos con los que el alumno podrá demostrar sus habilidades: en el modelo 1, se pide escribir una invitación a una fiesta nacional, además de explicar cómo es, qué hay y qué pueden hacer los invitados; a diferencia del modelo 2, que es una tarea más sencilla, en la que el sustentante da información sobre el clima de su país, la ropa que usa en dos estaciones del año, el clima de México, cuál es su preferencia entre ambos y su justificación. En sendos modelos hay descripción, sin embargo, el nivel de complejidad es mayor en el primero, porque el alumno tiene que ser creativo para invitar y dar información relevante de la fiesta de su país, a diferencia del modelo 2, donde sólo escribe sobre el clima.

Para Español 2, en los modelos 1, 2, 3 y 5, de acuerdo con las funciones, se evalúa brindar información mediante la descripción, y en el modelo 5 ocurre algo similar, pero con base en rechazar una invitación. Observamos que el nivel de exigencia lingüística varía en los cinco modelos de dos formas: 1) la extensión de los textos de entrada son diferentes (van desde 62, 75, 28, 46 y 58 palabras, respectivamente, en los modelos 1, 2, 3, 4 y 5), lo cual nos hace pensar que no existen criterios sobre la presentación de la tarea, al ser tan dispares; en este caso propondríamos un ajuste de entre 50 y 60 palabras, que no le representen dificultad al alumno para entender la situación y pueda seguir las instrucciones con unas pautas. Y en relación con el punto anterior, el léxico que se considera para la elaboración de la tarea no se corresponde con el nivel de conocimientos del alumno, ya que es oportuno recordar que la redacción de la tarea ha de cumplir unos requisitos básicos de validez, practicidad e impacto; es decir, que la instrucción para la elaboración del texto de estímulo debe ser clara y adecuada al nivel del alumno.

Por ejemplo, el texto de entrada del modelo 1 cuenta con un mayor número de palabras, por lo que se sale del esquema que tienen los modelos 1, 4 y 5. Además, incluye léxico fuera del nivel A2, como “por fin”, “curiosidad”, “encontraste”, “buscabas”, “cómo te va”. En ese sentido, el modelo 5 también contiene léxico fuera de nivel, como “interesada”, “tendrá”, “podrá”. Respecto al modelo 3, referimos que evalúa dar información sobre un hecho pasado mediante una fotografía y unas pautas, pero el inconveniente está en que el alumno tiene que desarrollar la tarea haciendo una descripción de ese hecho pasado: contar lo que pasó “anoche” con recursos lingüísticos por encima de su nivel A2. En el modelo 4, la redacción de la tarea es dirigida y clara, pero según los criterios básicos de validez, no cumple con este aspecto, porque en la misma instrucción se da información que el propio alumno debería producir: “Tienes una semana con muchísimo trabajo y no puedes ir. Explícale por qué no puedes aceptar su invitación”. Finalmente, las cinco tareas puntualizan que el número mínimo de palabras

debe ser de 100 para que el alumno logre desarrollarlas; no obstante, en ninguna parte se señala el tiempo en que debe ser resuelta la prueba, como tampoco ocurre con el resto de los modelos de Español 1, 3 y 4; por esta omisión del tiempo asignado a las tareas, podemos correr el riesgo de que sean impracticables.

En Español 3, los modelos 1 y 3 evalúan la función de narrar y describir, pero con distintos tipos de tareas. El primero, con pautas claras a seguir, que ofrecen información significativa sobre lo que el estudiante es capaz de hacer y puede corresponder a una experiencia vivida; pero cuenta con una particularidad que los otros dos modelos no consideran, y es el uso de terminología lingüística de forma explícita: "Usa pretérito y copretérito". Por lo tanto, creemos que falta a los principios de validez, practicidad e impacto, ya que en la instrucción misma se le están mencionando los recursos lingüísticos que debe utilizar. Respecto al modelo 3, podemos decir que corresponde al ámbito de la escritura creativa, ya que no hay unas pautas claras a seguir y el alumno tiene que echar mano de su imaginación a partir de una serie de 6 fotografías, es decir, escribir una aventura de un gato y un perro e imaginar el título que le pondrían a la historia; no obstante, evalúa narrar y describir como el modelo 1, que sí presenta objetivos de aprendizaje del nivel adecuado.

El modelo 2 no corresponde con lo que evalúan el 1 y el 3. Es una tarea sin interacción que valora argumentar sobre una probable situación de la vida. Además, se da la posibilidad de elegir unas pautas ("puedes utilizar expresiones como"), por lo que el alumno no está obligado a seguirlas, así que es una tarea que resulta ambigua para su realización. En ese sentido, incluso se proporcionan 4 fotografías que sirven de inspiración para desarrollar la tarea. Por otro lado, la tarea cumple con un objetivo de aprendizaje del nivel (el uso de la probabilidad o conjetura). Sin embargo, no creemos que este objetivo de aprendizaje sea parte de los que deban ser evaluados en español 3. Los tres modelos coinciden en desarrollar la tarea con un mínimo de 150 palabras.

En Español 4 hay cierta coincidencia con los modelos 1, 3, 4 y 5, ya que evalúan la función de expresar una opinión; de forma explícita, se pide argumentar a favor o en contra en los modelos 1, 3 y 5. Sin embargo, el modelo 1 presenta una contradicción, porque (como ya se había subrayado) en la instrucción se le pide al alumno expresar su opinión "a favor o en contra" de la publicidad, pero en la tercera de cuatro pautas se solicitan "las ideas a favor y en contra". Esta redacción de la tarea no cumple con el requisito básico de validez, porque hay contraindicaciones en la tarea misma, no es clara. A pesar de que en cuatro modelos se pide dar una opinión personal sobre un hecho o una situación como texto de entrada, los cuatro modelos presentan pautas o formatos diferentes. A continuación, explicamos en qué sentido:

El modelo 1 registra cuatro pautas específicas: la idea fundamental que defiende o ataca, las razones por las que piensa así, las ideas a favor y en contra y una conclusión; en cambio, el modelo 5 solicita la opinión a favor o en contra con base en una estructura del texto: una introducción, un desarrollo, conclusiones y que se utilicen ejemplos.

El modelo 4 es una tarea que por su formato es diferente a los demás. Las instrucciones junto a un texto de entrada hacen que sea una lectura larga y el formato más extenso de todos: cuenta con 129 palabras y los otros modelos tienen en promedio de 40 a 55. De hecho, bien podría eliminarse el texto de entrada que habla sobre el Patrimonio Cultural de la Humanidad y la tarea se comprendería perfecta y únicamente con las instrucciones restantes. Por otro lado, tenemos nuestras reservas sobre lo que se evalúa, ya que según se leen las instrucciones, detectamos cierta ambigüedad: se habla de presentar un informe sobre una tradición como Patrimonio Cultural de la Humanidad; se pide describirla y explicar por qué se cree que es importante, así como referir las medidas necesarias para protegerla. De este modo, nos preguntamos: ¿se evalúa la presentación de un informe o dar una opinión con argumentos? Observamos que la falta de claridad del objetivo que se solicita a los alumnos, se debe a la poca uniformidad en el diseño de las tareas, aunada a la ausencia de tablas de especificaciones. Finalmente, ahí radican gran parte de los problemas que estamos encontrando.

El modelo 2 no se ajusta a la función que los otros evalúan, pues éste se centra en la función de narrar; contar una historia personal para conseguir algo. Podemos observar que esta función se evalúa en Español 3, por lo que no corresponde a un objetivo de aprendizaje de Español 4. Aunque sí hay coincidencia en el número mínimo de palabras (200) que los alumnos deben usar para realizar la tarea.

Finalmente, como hemos expresado en repetidas ocasiones, es de suma importancia contar con criterios establecidos de forma colegiada, así como con los objetivos que queremos evaluar uniformemente en todos los modelos para lograr una evaluación objetiva, y eso no lo vemos reflejado en varios de los modelos de un mismo nivel.

2.5. Cómo es administrada la prueba

Todos los modelos de examen de la expresión escrita se administran en una plataforma digital en las instalaciones del CEPE, en salas donde hay computadoras para grupos de alumnos en horarios escalonados.

La implementación de esta modalidad computarizada de los exámenes de aprovechamiento se inició a principios del año 2011, únicamente con los exámenes objetivos de CA, CL y Gramática (Campos, 2011). Tiempo

después, se sumó el examen escrito de aprovechamiento por medios computarizados y ésta es la última prueba en aplicarse mediante la computadora después de que el alumno ha resuelto las pruebas objetivas. El hecho de aplicar cuatro pruebas, y en ese orden, nos puede indicar que quizá el alumno está en una situación de fatiga mental, lo cual puede repercutir en el resultado global del examen.

Aunado a este asunto del tiempo, en el formato de todos los modelos de examen de EE de Español 1 a 4, no se indican los minutos con que el alumno cuenta para resolver la tarea, aspecto importante para la actuación de éste con respecto a la tarea y el tiempo adecuado para su realización. Sin embargo, según nuestra experiencia, ésa es información que el profesor comparte con sus estudiantes cuando llega el momento de llevar a cabo los exámenes finales. Así se resuelve y suele indicárseles que la tarea de EE se debe resolver en un lapso de entre 30 y 45 minutos. Sin embargo, el tiempo definido para que el alumno resuelva la prueba debe estar asentado en el diseño mismo del instrumento y éste, además, ha de validarse por segunda ocasión, como se menciona en el capítulo 2.

3. Análisis e interpretación de los instrumentos de EE de Español 1, 2, 3 y 4

3.1. Aplicación de la prueba

Hasta 2019, los exámenes de aprovechamiento se aplicaban en las salas de cómputo, espacios donde los alumnos contestan las tareas de EE en horarios matutinos o vespertinos; siempre son auxiliados por personal académico, el cual supervisa que todo marche adecuadamente.

Al momento de contestar el examen de EE, los alumnos no tienen la posibilidad de elegir un modelo de nivel, sino que el mismo sistema lo asigna de forma aleatoria. Esto no resulta tan relevante si sólo hay dos modelos, como es el caso de Español 1; pero sí en el caso de Español 2 y 4, que tienen 5 modelos cada cual, por lo que en los diferentes horarios de aplicación no habrá mucha coincidencia de un mismo modelo para otro grupo de igual nivel. Eso representa una ventaja, ya que se dan situaciones en las que los alumnos tienden a platicar con detalle acerca del tipo de examen que tomaron, por lo cual pueden divulgar el tema y la parte fundamental de la tarea entre sus compañeros. Así que, mientras más modelos válidos y confiables haya, tendremos una mejor evaluación.

Suponemos que previamente a la presentación de los diferentes modelos que hemos descrito y analizado, existió un primer prototipo de cada nivel de español, que se utilizó para la etapa de experimentación; sin embargo, no tenemos noticia de esto y resulta una etapa importante en el diseño de los instrumentos.

3.2. Rúbrica

La expresión escrita de todas las tareas de Español 1 a 4 se evalúa con base en dos escalas: holística y analítica. La holística considera una valoración global que toma en cuenta la dimensión comunicativo-discursiva (logro del propósito de la tarea + desarrollo coherente y cohesionado de las ideas + convenciones del género textual) + la precisión lingüística (vocabulario + gramática). Esta escala contiene 4 bandas de 0 a 3, que van de menos a más y describen de lo que el alumno es y no capaz. Cada uno de los descriptores se definen por cuatro opciones disyuntivas más una adición, que van de menos a más (tareas insuficientes a satisfactorias). Es decir, de manera amplia, esta escala describe en las bandas el cumplimiento del tema, el logro del discurso, el desarrollo coherente y cohesionado de las ideas, las convenciones del género textual y el uso de vocabulario o de las categorías gramaticales en el mensaje.

Como recién se mencionó, la escala analítica considera la competencia pragmático-discursiva y la competencia lingüística. En el caso de la pragmático-discursiva, hay dos rubros que suponíamos que ya se habían evaluado en la holística: cohesión y coherencia con 4 bandas de 0 a 3; en tanto que la competencia lingüística considera tres rubros: rango y control léxico, precisión gramatical y ortografía; asimismo, con 4 bandas de 0 a 3. Los dos primeros rubros, también suponíamos que ya habían sido evaluados en la escala holística.

Con la descripción de esta rúbrica dividida en dos escalas, creemos que hay criterios que se están evaluando doblemente, además no se tiene en cuenta su actuación, sino sus deficiencias. Un defecto que tiene esta rúbrica es que está redactada a partir de criterios descritos de forma negativa, lo que da como resultado una evaluación desde un enfoque de lo que el estudiante no sabe hacer (Figueras, 2013: 123).

De acuerdo con Figueras, es aconsejable utilizar criterios holísticos para los niveles iniciales y criterios analíticos para los niveles subsecuentes (Figueras, 2013), pero de ninguna manera se recomienda utilizar ambos. Por lo tanto, pensamos que para la evaluación de Español 1 se podría calificar con criterios holísticos, porque en este nivel los alumnos cuentan con recursos lingüísticos básicos para hacer descripciones mientras que para las tareas de Español 2, 3 y 4 se propondría una escala analítica, ya que supone la evaluación de distintos aspectos de la actuación: la adecuación de la situación comunicativa propuesta por la tarea, la coherencia, la cohesión, la corrección gramatical y la ortografía; todo tomando como base las categorías de la misma rúbrica.

Finalmente, es importante decir que debe haber una correlación entre los contenidos de lo que se pretende medir y la rúbrica para validar o

comprobar esos contenidos y que no llegue a ser una mera lista de cotejo ni que se generalice de forma amplia lo que se pretende medir. Así que la rúbrica analítica se podría particularizar, en cierta medida, sobre los contenidos de Español 2 a 4, incluyendo aspectos relevantes de los tres niveles, ya que son pruebas productivas a puntuarse con criterios de calificación.

3.3. Cómo y qué deben evaluar los criterios de evaluación (rúbrica)

Los criterios de evaluación focalizan habitualmente más de tres aspectos, en el caso de la expresión escrita: la adecuación a la tarea propuesta, la riqueza y la corrección lingüísticas. Además, como lo indica Figueras (2013), es muy importante someter toda actividad de evaluación a los procesos de control de calidad con el fin de asegurar su validez y confiabilidad.

Comparando varias escalas, hay algunas que incluyen la categoría de *alcance*, que ayuda a considerar el uso de las frases idiomáticas, los coloquialismos, el uso de ideas complejas y abstractas; en general aporta a la transmisión de diferentes tipos de textos con los recursos que la lengua lo requiere. Como se podrá observar, hemos descrito esta categoría de más a menos.

Es preciso hacer notar que al momento de realizar el análisis no pudimos tener acceso al dato de la procedencia de los alumnos evaluados, ya que el material (la tarea escrita impresa del alumno), sólo contiene el código del expediente del alumno, el nombre del profesor del curso, el nivel, la tarea y la tarea resuelta. Habría sido interesante contar con la nacionalidad del sustentante para tener más datos a analizar y ver el tipo de tarea que se les facilita o dificulta más a los alumnos por población. No obstante, sabemos que la población de los estudiantes extranjeros es diversa, pero predominan los asiáticos (coreanos, chinos y japoneses), europeos y estadounidenses, cuya edad promedio va de 20 a 34 años, según el texto de Campos (2011).

3.4. Quiénes y qué evalúan (en qué se enfocan los calificadores expertos)

Posteriormente a que el alumno resuelve la prueba escrita, se imprime y entrega para ser calificada por el profesor que no tuvo a ese estudiante en su curso. Esta forma de distribuir los exámenes a los profesores calificadores apoya la idea de evaluar objetivamente a los alumnos, aun cuando los exámenes impresos no llevan el nombre del sustentante, pero sí el del profesor que impartió dicho nivel.

Los profesores que califican la prueba de EE de los niveles de Español 1 a 4 son profesores que tienen mucha experiencia en la enseñanza de ELE.

Decíamos que con la finalidad de calificar lo más objetivamente posible, los profesores no deben calificar a sus propios alumnos y, además, dentro de este proceso, deben calificar pruebas del nivel de español que impartieron; es decir, el proceso se realiza de forma cruzada, de tal manera que se asegure que saben cuáles son los contenidos que se imparten en ese nivel, y que se tiene una idea de qué aspectos se evalúan ahí, según los contenidos del curso. Sin embargo, observamos que de 22 profesores, sólo 4 no impartieron el nivel de español que calificaron, lo cual no es tan grave, siempre y cuando el calificador estuviera estandarizado como calificador experto de ese nivel.

El hecho de estandarizar a los calificadores (Lugo, 2020) es de suma importancia, ya que al no tener esa conciencia, cada uno se fija en aspectos diferentes, e incluso podría llegar a omitir el objetivo principal que se está midiendo; es decir, lo que se pretende evaluar con cada examen. Se cuenta con una vasta experiencia en la enseñanza de ELE, pero no con una estandarización en la evaluación de exámenes de aprovechamiento para cada uno de los niveles de español del CEPE.

Si bien ha sido un logro establecer los exámenes por computadora y homogeneizarlos, el aspecto que más se debe trabajar es que los calificadores sigan los criterios de calificación, y eso sólo se logrará gracias a la estandarización de los docentes. Las discrepancias en el modo de calificar se reflejan a través de las anotaciones que hacen los profesores sobre las pruebas impresas. Dichas notas tienen que ver con las categorías que se evalúan en las escalas, por eso retomamos dichos recursos. Como se ha dicho anteriormente, la calificación se hace con dos escalas o rúbricas: una holística y otra analítica. Ambas se utilizan para todos los modelos de exámenes de Español 1 a 4. La primera contiene una categoría denominada con el mismo término de *holística* con cuatro bandas que puntúan de 0 a 3, y la segunda contiene cinco categorías: *cohesión, coherencia, léxico, precisión gramatical y ortografía*; también se califica con 4 bandas, de 0 a 3 puntos. La suma de ambas escalas da un valor numérico total de 18 puntos; esto constituye otro factor importante que describiremos y analizaremos a continuación, junto con las categorías que se evalúan.

Siguiendo nuestras observaciones de todas las pruebas escritas de Español 1 a 4, efectivamente hallamos que los calificadores tienen una conciencia más identificada de qué mide cada una de las categorías. Sin embargo, los diferentes tipos de anotaciones que encontramos, están más enfocadas en la gramática; la gran mayoría, por ejemplo, en la conjugación adecuada de los verbos, la concordancia de género y número, el empleo inadecuado de las preposiciones, etc. Los calificadores también ponen mucha atención en los errores de acentuación, que son de orden ortográfico. Le siguen los aspectos de cohesión del texto y coherencia, con la falta de claridad en la progresión

temática y el escaso uso de conectores; además de la puntuación, con acotaciones sobre el uso excesivo del punto y seguido en oraciones cortas.

El número mínimo de palabras que se solicita para cada prueba no está considerada de forma explícita en la escala. Y es que hay ideas sobreentendidas entre los calificadores; en este caso, la información que se transmite entre los profesores es que el número mínimo de palabras no se premia ni se sanciona. Existe la idea de cuál es la cantidad recomendada para el alumno, pero en situaciones extremas esto provoca que se vea afectada la calificación en la escala holística o en la categoría de léxico, por ejemplo.

Por otro lado, cuando la tarea presenta un alto número de palabras, puede darse el caso de que el calificador se vaya con la falsa impresión de que el alumno cumplió con el propósito de la tarea y otorgue el máximo de puntos en la escala holística, cuando en realidad no es así.

Además, la escala holística presenta un problema, ya que contiene una valoración global de varios aspectos que también se incluyen con la escala analítica; por lo tanto, estamos evaluando por partida doble. Por un lado, en la holística se evalúa la dimensión comunicativo-discursiva (desarrollo coherente y cohesionado de las ideas) y precisión lingüística (vocabulario y gramática), y por otro lado, en la escala analítica se califican cuatro de las cinco categorías ya aludidas: cohesión, coherencia, léxico y precisión gramatical.

Algo que es criticable sobre la configuración de la prueba impresa para su calificación es que la rúbrica no está integrada a la tarea, por lo tanto, no es una labor práctica para el profesor; es decir, tiene que consultar la rúbrica de forma independiente, y también el formato donde el profesor asienta la calificación es un documento aparte. Incluso, hay que decir que con estos tres documentos independientes, el alumno no tiene la posibilidad de ver cuáles son los criterios bajo los que es evaluado al momento de responder la tarea o, en su defecto, para solicitar una revisión de examen, que si bien los casos de este tipo no abundan, claro que los hay. Es una recomendación importante que la literatura de la evaluación señala.

Otro aspecto que se considera importante y al parecer no está establecido rigurosamente, es hasta los cuántos errores del mismo tipo o categoría se empieza a otorgar banda 3, 2, 1 o 0. Eso no está estipulado, aun cuando es un asunto de la estandarización. También se debe tener presente que cuando un error de acentuación se repite varias veces, sólo se debe tomar en cuenta como uno; por ejemplo, si el alumno no acentúa la palabra país cinco veces en su texto, debe valorarse como un único error.

Por último, otro tema que no está presente en el formato de la prueba es

el punto de corte entre aprobado y reprobado. No hay un parámetro que indique el puntaje que debe obtener el alumno para superar el examen con base en el 15 % que se acota en la parte superior de la tarea. En otras palabras, el calificador otorga como máximo 18 puntos cuando califica cualquiera de los exámenes de Español 1 a 4, de acuerdo con la rúbrica, y el puntaje que obtenga el alumno representa el 15 % de su calificación total, para así poder determinar si continúa o no al siguiente nivel.

4. Conclusiones

4.1. Debilidades y fortalezas

El no contar con el propósito de las evaluaciones escritas (al menos no bajo un margen de exactitud adecuado) a través de las interrogantes *¿qué?*, *¿cómo?*, *¿a quién?* y *¿para qué?*, nos lleva a desarrollar instrumentos sin bases bien fundamentadas; como resultado, nos enfrentamos a lo que hemos venido comentando: las pruebas miden diferentes propósitos a través de distintas tareas, y aun cuando intuimos su finalidad, no la tenemos por escrito. Dicho documento, inexistente hasta el momento, debería ser compartido por el equipo de profesores que crea las pruebas, el que las califica y el que las valida o ajusta.

En el capítulo 2, se presenta una propuesta metodológica sobre la creación de exámenes de aprovechamiento, la cual consta de tres etapas: diseño, elaboración y validación, en las que se responden las interrogantes del párrafo anterior. Una de las grandes debilidades de los instrumentos de evaluación de la expresión escrita (que hemos analizado aquí) es que dentro de la etapa del diseño, falta la definición de los contenidos: qué entendemos por expresión escrita y qué esperamos que puedan escribir los alumnos en cada nivel; la longitud, el tema, características textuales, previo al desarrollo de las tareas, ya que no tenemos esos documentos. Los diseñadores deben poner por escrito la definición teórica de estos constructos y, además, elaborar la tabla de especificaciones en la que se defina operacionalmente el constructo, para finalmente validarla. En pocas palabras, han de ofrecer la descripción detallada de las características del dispositivo a desarrollar, lo cual facilitaría la selección de los contenidos.

Respecto a la etapa dos, la de elaboración, tenemos los instrumentos de evaluación escrita desarrollados en su momento: se trata de las tareas de Español 1 a 4 que analizamos en este octavo capítulo. En la configuración de la prueba, se debe comprobar que los procesos relativos a la producción, administración, corrección y puntuación estén estandarizados y preparados. Al parecer, esta parte está cubierta; sin embargo, no tenemos fe de que la corrección de las tareas responda a los criterios

establecidos en las especificaciones, porque (reiteramos) no se cuenta con ellas.

Por otro lado, con respecto a la forma de valorar las pruebas, el calificador apela a su experiencia, pero no metodológicamente hablando, ya que los criterios no están estandarizados. Otro aspecto que tiene que ver con la calificación es que a las tareas de todos los niveles se les asigna un número de puntos exactamente igual, ya que comparten la misma escala o rúbrica; sin embargo, los objetivos que se miden son diferentes, además, no se tienen fijados los puntos de corte para cada uno de los instrumentos por nivel de lengua. Es decir, los parámetros para definir con cuántos puntos el alumno está aprobado o bien, reprueba. Recordemos que los puntos de corte son una dictaminación que consiste en la toma de decisiones sobre la fijación de estándares, de forma colegiada y consensuada. Esta etapa, comúnmente, se lleva a cabo en la experimentación de las pruebas o en su diseño, después de la elaboración de las tareas, y es entonces cuando el equipo técnico de expertos define los puntos de corte.

Por último, hay un banco de tareas poco abundante en los niveles de Español 1 y 3. Desconocemos las fechas de la creación de todas las pruebas escritas y las de los períodos en que han sido administradas, así como cualquier otra característica útil. Tampoco conocemos que exista un informe técnico sobre cada una de las pruebas: un informe documentado en el que se detallan los aspectos más importantes de su proceso de diseño, elaboración y corrección; o algo tan relevante en la confección de los instrumentos como es aplicar los principios básicos de la teoría psicométrica: el análisis psicométrico de los exámenes, es decir, después de que un modelo de examen ha sido aplicado a una población numerosa (de 100 alumnos, por ejemplo), cómo dicho modelo está reaccionando ante los parámetros de validez y confiabilidad. Este trabajo hace falta dentro de este proceso de diseño de exámenes.

4.2. Propuestas

En este apartado queremos ofrecer, de manera puntual, tres propuestas de metodología con el fin de incentivar buenas prácticas de evaluación: la primera está en relación con el diseño de las pruebas; la segunda tiene que ver con los calificadores; y la tercera, con la escala o rúbrica. Las tres, en sí, están estrechamente relacionadas y la primera es la base de las otras dos.

Sobre esta primera propuesta, consideramos que tenemos que empezar desde cero con el diseño de las pruebas de expresión escrita. Es verdad que los instrumentos que hemos analizado pueden ayudar a orientarnos, pero debemos consultar los programas de cada nivel de español y sus

objetivos para responder *qué, a quién, cómo y para qué* queremos medir en la expresión escrita. El diseño de los exámenes ha de incluir la definición de los constructos y las tablas de especificaciones; recordemos que es un proceso complejo que tiene inicio en la etapa de diseño. Una vez que se tengan las tablas de especificaciones validadas, se propone subirlas a la plataforma virtual para que los creadores de tareas tengan a la mano en el mismo sistema esta información que, como hemos comentado en repetidas ocasiones, hemos echado en falta durante este análisis. Así pues, las tareas harían el recorrido que tienen que cumplir (de revisión, corrección, experimentación, aprobación y liberación), hasta la plataforma virtual con sus respectivas tablas.

Como parte de la primera propuesta, recomendamos que respecto a la creación de las pruebas se creen tareas de interacción sobre una situación posible en un entorno cotidiano y significativo para el alumno. Como se cita en el capítulo 6, es necesario relacionar el mensaje con el contexto extralingüístico (el emisor, el destinatario, el lugar, el tiempo y la situación). La literatura señala que siempre será mejor presentar una tarea con una intención comunicativa y con un formato interactivo, que una en la que se le ordene de forma directa al alumno lo que debe hacer. Por otro lado, hay que considerar que éste debe tener cierta familiaridad con la tipología textual, el tema y la exposición de las instrucciones para su realización.

Nuestra segunda propuesta es sobre la necesidad de estandarizar a los calificadores y, con esto, nos referimos absolutamente a todos los profesores para que califiquen lo más homogéneamente posible. Incluso, podría empezarse por estandarizar a los profesores que con regularidad imparten el mismo nivel de español y, en ocasiones consecutivas y posteriormente, de ser posible, estandarizarlos en otros niveles. Es importante que los calificadores estén activos en el cargo de juicio experto, de otra manera se pierde la habilidad; de hecho, se recomienda realizar un análisis periódico de su comportamiento para evaluar el efecto de su entrenamiento y, de ser el caso, mejorar sus prácticas (Lugo, 2020).

La tercera propuesta está relacionada con la estandarización de los calificadores, ya que alguien que cumple bien esta función no sólo debe ser experto en la lengua, también requiere de un entrenamiento específico para conocer y aplicar correctamente las escalas o rúbricas, y eso desde luego forma parte de la estandarización. Por lo anterior, dentro de las escalas proponemos dos aspectos: de estructura y de contenido. En cuanto a la estructura, sugerimos eliminar la escala holística para los niveles 2, 3 y 4, y dejar la analítica en la que se agregaría la categoría de logro del propósito o cumplimiento de la tarea, así que tendríamos seis categorías; para dicha escala, reconsideraríamos el contenido de cada una de éstas, haciendo a su vez ajustes de lo que se evalúa para cada una de las tareas de los niveles. Para el nivel 1, nos quedaríamos con la escala holística; respecto

al contenido, agregaríamos una descripción más detallada de las bandas según lo que se mide.

Se hace esta propuesta ya que, como hemos comentado, en el análisis de la escala o rúbrica, nos parece que hay un error, porque se evalúan doblemente algunas categorías, tanto en la holística como en la analítica.

Con el hecho de proponer los exámenes de forma virtual, se aprovecharía este medio para recomendar que la tarea deje de imprimirse. Además, el calificador podría evaluar directamente en una computadora, a la cual tendría acceso mediante una clave; de esta manera, también se le concedería acceso a las rúbricas y facultad para otorgar las puntuaciones en el mismo sitio de los exámenes. Esto simplificaría el proceso de calificar esta prueba productiva.

Por último, es importantísimo definir el punto de corte que separe dos categorías de cada uno de los exámenes; por ejemplo, las categorías “aprobado” y “no aprobado” dentro de la prueba de aprovechamiento. Creemos que diseñar un instrumento de evaluación es un proceso que requiere seguir una metodología rigurosa, por lo cual no deja de ser una labor compleja. Así que, es necesario subrayar que el desarrollo de un examen recae en la responsabilidad colectiva de los equipos de trabajo, y no es responsabilidad exclusiva de unos cuantos profesores.

Referencias

Campos, E, Espinosa, A. y Jurado, M. (2011). Exámenes de aprovechamiento computarizados. Un paso más en la evolución académica del departamento de Español. *Decires*, 13, 121-158.

Eguiluz, J. (2009). *Criterios para la evaluación de la producción escritos*. Marco ELE, 9, 75-94. Recuperado 31 de enero de 2021, de https://marcoele.com/descargas/expolingua1996_eguiluz-vega.pdf

Figueras, N. y Puig, F. (2013). *Pautas para la evaluación del español como lengua extranjera*. Madrid: Edinumen.

Marco común europeo de referencia para las lenguas. (2002). Madrid: Instituto Cervantes. Recuperado 28 de octubre de 2020, de https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf

Lugo, I. (2020). *La importancia de la estandarización de los calificadores en los exámenes de expresión*. Ciclo internacional de conferencias sobre evaluación [Video]. Recuperado 10 de noviembre de 2020, de https://www.youtube.com/watch?v=OqDHgbKsxpM&t=5209s&ab_channel=CEPEVideos

SECCIÓN IV. LAS TECNOLOGÍAS DE APRENDIZAJE Y CONOCIMIENTO (TAC) EN LA EVALUACIÓN DE EO Y EE

CAPÍTULO 9

BREVE HISTORIA DE LOS DESARROLLOS TECNOLÓGICOS APLICADOS A ELE EN EL CEPE

Maribel **Carmona Herrera**

1. Introducción

El crecimiento del número de estudiantes de español en el mundo, nos enfrenta al reto de avanzar y modernizarnos a la par del aumento de la matrícula. Además, es necesario reforzar los desarrollos tecnológicos, como los que surgieron gracias al apoyo de los académicos del CEPE de la UNAM, de acuerdo con los lineamientos propuestos en el MCER y el esfuerzo conjunto del área de Tecnología Educativa y la Dirección General de Tecnologías de la Información y Comunicación (DGTIC).

El objetivo de este capítulo es describir el proceso de implementación y mejoras que han sufrido los distintos desarrollos tecnológicos aplicados a la evaluación de ELE en el CEPE; específicamente, con respecto a las expresiones oral y escrita en su vinculación con varias tecnologías, las cuales han sido probadas en la evaluación de dichas habilidades. Mediante estos métodos, ha podido prescindirse, además, de los modos convencionales de escritura en papel o los diálogos cara a cara.

En este capítulo también nos referiremos a la interpretación de los resultados obtenidos a partir de la adaptación de los exámenes de aprovechamiento y su aplicación en línea, puesta en marcha desde el 2011 a la fecha. Nos centraremos específicamente en la plataforma de exámenes departamentales (en funcionamiento hasta hoy día), y haremos una reflexión de sus principales debilidades y fortalezas. Esto desde el punto de vista, como trabajo colaborativo, tanto de la plataforma tecnológica como del evaluador, con la finalidad de dotar nuestra nueva propuesta tecnológica (planteada en el capítulo 10 de este libro) de la mayor cantidad de mejoras y correcciones, para obtener herramientas sólidas, con mayor experiencia, actualidad y objetividad en la obtención de resultados finales.

2. Antecedentes

En el CEPE se elaboran y aplican diferentes procedimientos formales de evaluación: 1) exámenes de colocación, 2) de aprovechamiento, 3) de certificación, 4) de admisión para diplomados y especialización en español, y 5) extraordinarios. A través de ellos, se evalúan las cuatro habilidades y el sistema de la lengua (la competencia lingüística).

En el año 2005 se desarrolló la primera plataforma del CEPE, que gestionaba y aplicaba exámenes para colocar de forma automatizada a los estudiantes en sus respectivos niveles de español. Esto se lograba a partir del desempeño de cada alumno en la resolución de varios exámenes progresivos que, de acuerdo con una tabla de equivalencias y número de aciertos, permitían a los examinados seguir avanzando de nivel, quedarse en su primer intento, o incluso bajar de nivel, según sus resultados. Este proceso era muy tardado, tomando en cuenta que los estudiantes presentaban el examen el mismo día que se inscribían a los cursos. Posteriormente, estas pruebas se implementaron de nuevo, pero con una construcción diferente y en una plataforma ligada directamente con el proceso de inscripción, con lo cual se logró una optimización y efectividad en el tiempo y asignación de nivel para el alumno.

En el caso de los exámenes de dominio de la lengua y, posteriormente, de certificación, se tiene un camino más recorrido. Comenzando por el Examen de posesión de la lengua española (EPL), creado en 1997, en el CEPE se desarrolló en 2004 una herramienta para su aplicación. Dicho mecanismo fue uno de los primeros intentos en los que se implementó un módulo, cuyo objetivo consistía en evaluar la competencia oral de un grupo de estudiantes coreanos que, de forma remota, pretendían conseguir un puntaje certero de su dominio de español.

The image shows a screenshot of a web browser displaying the website for the Examen de Posesión de la Lengua Española (EPL). The browser's address bar shows '10 de 11' and 'Zoom automático'. The website header includes the logo of the Universidad Nacional Autónoma de México (UNAM) and the text 'Universidad Nacional Autónoma de México Examen de Posesión de la Lengua Española'. On the right, there is a logo for 'eple CEPE - UNAM'. A sidebar on the left contains a menu with the following items: PRESENTACIÓN, ¿PARA QUÉ EL EPLE?, ¿QUÉ MIDE?, ESTRUCTURA, REGISTRO COREA, and INSTRUCTIVO. The main content area features the 'eple' logo and a section titled '4. Expresión oral. Duración 15 minutos: máximo de puntos 170.' Below this title, the text explains that oral expression is evaluated based on three activities: 1. interview, 2. monologue, and 3. questions about the monologue. It notes that this section is recorded and stored for later evaluation. A recommendation is given to have maximum exposure to Spanish. A list of important aspects to consider is provided, including speaking without interruptions, using a varied vocabulary, and practicing response speed. The section concludes with the heading '1. Entrevista (2 minutos aprox.)'.

Figura 10. Imagen del sitio web del Examen de posesión de la lengua española.

Por otro lado, en el año de 2002 se creó el Certificado de español como lengua adicional (CELA), que se aplicó en papel desde 2003 hasta 2006. En 2007, se hizo una adaptación para aplicar este examen de manera computarizada: el Sistema base de reactivos, que se desarrolló a través de un convenio celebrado entre el CEPE y la DGTIC. Dicho sistema comenzó a aplicarse en esta versión en el año 2013, y a la fecha continúa vigente.

Universidad Nacional Autónoma de México

CIUDAD UNIVERSITARIA POLANCO CHICAGO SAN ANTONIO TEXAS TAXCO LOS ANGELES GATINEAU QUEBEC

CEPE Centro de Enseñanza para Extranjeros

DIRECTORIO QUIENES SOMOS PUBLICATIONES BIBLIOTECA MAPA DEL SITIO PREGUNTAS FRECUENTES ENGLISH

CALENDARIO

CURSOS ESPAÑOL Y CULTURA

ESPECIALIZACIÓN

CERTIFICACIÓN

DIPLOMADOS

ACTIVIDADES EN LÍNEA

ACTIVIDADES CULTURALES

VISITAS GUIADAS

INTRANET

CERTIFICADO DE ESPAÑOL COMO LENGUA ADICIONAL (CELA)

Inicial

El CELA inicial certifica un nivel de competencia comunicativa suficiente para interactuar socialmente en situaciones relacionadas con la vida cotidiana y para comunicar de forma básica, experiencias personales, planes, opiniones, deseos y necesidades.

El CELA inicial certifica el nivel correspondiente al B1 (usuario independiente-nivel umbral del Marco común europeo de referencia para el aprendizaje, enseñanza y evaluación de lenguas.

Intermedio

El CELA intermedio certifica un nivel de competencia comunicativa suficiente para expresarse de modo claro y detallado sobre una amplia gama de temas, en circunstancias normales de comunicación, que no requieran un uso especializado de la lengua.

El CELA intermedio certifica el nivel correspondiente al B2 (Avanzado) del Marco común europeo de referencia para el aprendizaje, enseñanza y evaluación de lenguas.

Superior

El CELA superior, acredita la competencia comunicativa necesaria para desenvolverse sin dificultad en situaciones que requieren un uso preciso y matizado de la lengua y un

Figura 11. Imagen del sitio web del CEPE que informaba sobre el Certificado de español como lengua adicional.



Figura 12. Imagen de la entrega del Sistema base de reactivos, el día 26 de junio de 2008.

En el curso de invierno del 2011, se implementó la modalidad computarizada de siete exámenes de aprovechamiento, correspondientes a cada nivel, que hasta la fecha son aplicados al final de los cursos intensivos de español. Estas pruebas se diseñaron con base en el perfil de egreso, definido en ocho de los nueve niveles que conforman el programa de estudios del CEPE (de iniciación a Español 7), y se han aplicado de forma ininterrumpida desde 2011 hasta el verano de 2020, año en que su aplicación fue afectada debido a la pandemia mundial por COVID-19.

La contingencia sanitaria derivada de lo anterior tomó fuerza en México durante el transcurso de la primavera del 2020, factor que detonó que se le diera mayor relevancia al desarrollo de una nueva plataforma de evaluación de habilidades, en la cual actualmente estamos trabajando. El desglose y especificaciones de esta propuesta serán abordados más adelante, en el capítulo 10.

Finalmente, los exámenes de admisión para diplomados y especialización en español, se gestionan desde los sistemas de seguimiento específicos para cada programa de Formación de Profesores de Español como Lengua Extranjera. Quizá más adelante, valdría la pena considerar la incorporación de todos estos exámenes a la plataforma que estamos desarrollando, con el objetivo de concentrar la totalidad de los resultados para su análisis, seguimiento y administración.

2. Descripción del sistema de gestión de exámenes departamentales o de aprovechamiento

Nos centraremos ahora en el análisis del sistema de gestión de los exámenes de aprovechamiento y colocación con el que cuenta el CEPE

actualmente. Sobre esta base partiremos para desarrollar la plataforma que se propone en este proyecto. En el año 2010, comenzó el desarrollo de este sistema de gestión de exámenes con dos equipos de trabajo bien coordinados y organizados.

Por un lado, en el aspecto académico, para la integración de contenidos y reactivos de acuerdo a los 7 niveles de español, la primera versión de los cuatro exámenes básicos fue el producto del trabajo de varias profesoras de asignatura; para los exámenes intermedios se contó con otro grupo de profesores de tiempo completo. Por otro lado, el equipo técnico de programación estuvo integrado por Maribel Carmona y Roberto Sierra, bajo la supervisión de Rodrigo Galindo, jefe del área de Tecnología Educativa y Cómputo en ese momento (Campos, Espinoza, Jurado, 2011: 130).

Al inicio de la aplicación de estos exámenes (2011), se integraron las versiones con sólo tres secciones para evaluar: la comprensión de lectura, la comprensión auditiva y la competencia lingüística; así, las habilidades de expresión escrita y oral quedaron fuera de la aplicación computarizada, manteniéndolas como objeto de evaluación por medios convencionales.

Posteriormente, en el año de 2015, se integró la versión computarizada de la prueba de expresión escrita, a partir de tareas específicas, de acuerdo con el nivel de español que se requería. Para tenerlas dentro de este sistema de gestión, se crearon unas versiones de exámenes con una cuarta sección, buscando brindarle al profesor encargado de la evaluación, una lectura más rápida, comprensible e integrada de las evaluaciones de las otras habilidades adquiridas durante el curso, mediante hojas de las rúbricas para obtener el puntaje final. Dichas rúbricas, conviene aclarar, se encuentran impresas; en otras palabras, no es posible desglosar o especificar qué criterios toman en cuenta los calificadores para asentar el puntaje final obtenido por el estudiante dentro del sistema mismo.

La evaluación del aprovechamiento de la expresión oral se sigue llevando a cabo por separado de la aplicación computarizada, y se integró un sistema de calificación para estas dos últimas expresiones, con una escala o rúbrica de 18 puntos para cada una, la cual también está fuera del sistema: no proporciona información sobre los rubros y sólo se puede ingresar en ella el puntaje total obtenido.

Poco a poco, se han agregado más versiones de exámenes de cada nivel de español, con una mayor participación de los creadores de reactivos y tareas, quienes continúan integrando versiones adicionales de estas pruebas. En el año 2017, se integraron al sistema los exámenes extraordinarios de español, con la finalidad de aplicarlos a los alumnos que deseaban demostrar su dominio de español en un nivel específico, o que requerían aprobar un curso para avanzar de nivel de forma más rápida.

Finalmente, en el año 2018, se integraron también los exámenes de colocación, los cuales son una integración de reactivos que, de forma progresiva, incrementan su nivel de dificultad; de tal modo, ubican al aspirante en un nivel de español, de acuerdo con unas tablas de valores mínimos de aciertos, con la finalidad de asignarles un grado específico y las asignaturas correspondientes.

3. Catalogación y almacenamiento de datos

El sistema de gestión de exámenes cuenta con una base de datos de reactivos, recursos y exámenes, los cuales están relacionados y categorizados de acuerdo con el nivel de español, tipo y examen que se requiera (departamental, de colocación o extraordinario).

Como se puede apreciar en la siguiente imagen, es posible consultar o probar la sección de reactivos por separado. Este módulo cuenta con las funciones básicas, tanto para administrar el banco de reactivos, como para agregar, editar o eliminar.

En el módulo de recursos, se agrupan los reactivos por tareas específicas, tales como adjuntar un video o un audio para la comprensión auditiva;

ID	NIVEL	PREGUNTA	VER REACTIVO	HABILITADO	EDITAR	ELIMINAR
162	Español 2	Rosa y Hugo hacen un recorrido por el centro porque	Ver	<input checked="" type="checkbox"/>		
163	Español 2	Rosa y Hugo hacen un recorrido en	Ver	<input checked="" type="checkbox"/>		
164	Español 2	El edificio de correos lo diseñó	Ver	<input checked="" type="checkbox"/>		
165	Español 2	En la construcción del edificio se emplearon materiales de	Ver	<input checked="" type="checkbox"/>		
166	Español 2	Actualmente el edificio de correos es	Ver	<input checked="" type="checkbox"/>		
167	Español 2	La celebración del Niñoopa tiene su origen en el siglo	Ver	<input checked="" type="checkbox"/>		
168	Español 2	La celebración es una combinación de características	Ver	<input checked="" type="checkbox"/>		
169	Español 2	El Mayordomo es	Ver	<input checked="" type="checkbox"/>		
170	Español 2	Para cuidar al Niñoopa, hay una lista de espera hasta	Ver	<input checked="" type="checkbox"/>		
171	Español 2	El Niñoopa cambia de casa y Mayordomo en	Ver	<input checked="" type="checkbox"/>		
360	Español 2	El jet pack	Ver	<input checked="" type="checkbox"/>		
359	Español 2	¿Por qué es necesario desarrollar nuevas formas de transporte?	Ver	<input checked="" type="checkbox"/>		
358	Español 2	El tema del texto es	Ver	<input checked="" type="checkbox"/>		
361	Español 2	¿Cuánto debe pesar el conductor del jet pack?	Ver	<input checked="" type="checkbox"/>		
362	Español 2	La información sobre el jet pack no incluye	Ver	<input checked="" type="checkbox"/>		
363	Español 2	El XfV parece un	Ver	<input checked="" type="checkbox"/>		
364	Español 2	Una característica de los dos ejemplos de transportes es que	Ver	<input checked="" type="checkbox"/>		
365	Español 2	¿Cuál de las siguientes afirmaciones es verdadera?	Ver	<input checked="" type="checkbox"/>		

Figura 13. Imagen del sistema de gestión de exámenes finales. Reactivos.

integrar imágenes, una lectura o un periódico en línea, para la comprensión lectora; o bien, a cada recurso se le pueden asociar uno o varios reactivos, que funcionarán como grupos de instrucciones y preguntas, y así ayudarán a la integración de los exámenes.

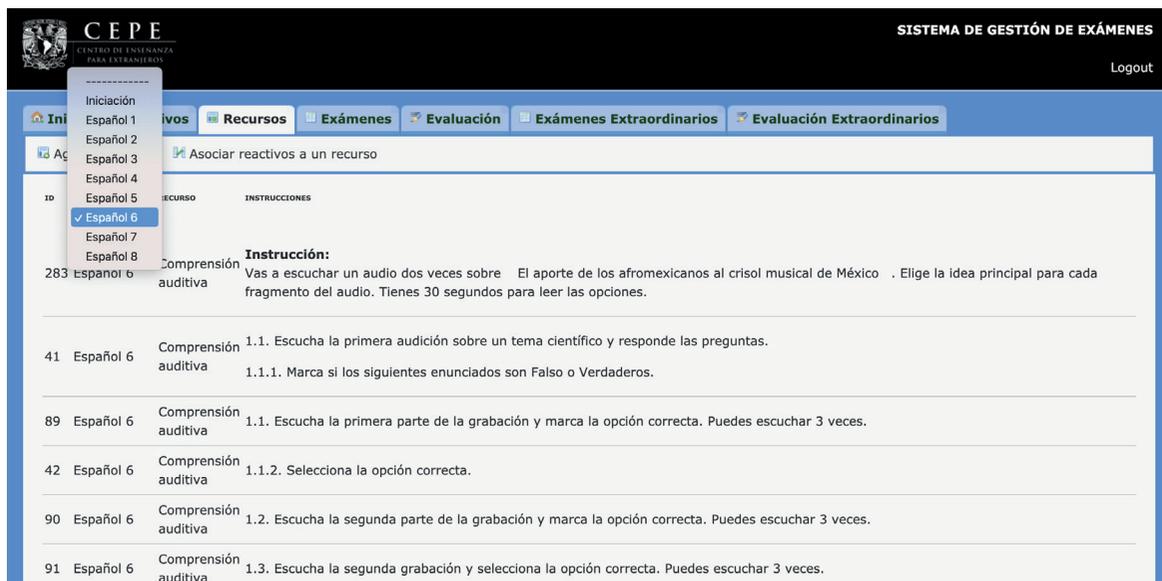


Figura 14. Imagen del sistema de gestión de exámenes finales. Recursos.

El módulo donde se integran todos los datos relacionados y organizados, de forma sistemática y con un orden específico para su resolución, es el del armado final del examen. En esta sección contamos con las versiones de los exámenes aplicados, el número de estudiantes que los contestaron en diversos periodos del año, y aquellos que se encuentran vigentes.



Figura 15. Imagen del sistema de gestión de exámenes finales. Recursos.

Por último, en el módulo de evaluación se integran todos los resultados obtenidos por los estudiantes. Como se aprecia en la siguiente imagen, las habilidades son calificadas de forma separada; no obstante, hace falta integrar la producción oral a este sistema, como parte de un mismo examen, para después brindar la posibilidad de calificarla de forma asíncrona.

Seleccione el año: 2021
 Seleccione la Sesión: Primavera
 Seleccione el período: Primavera 1
 Seleccione el período para desplegar estudiantes: Curso Intensivo de Español Virtual - Intensivo de Español (08/03/2021 - 29/04/2021)
 Seleccione el grupo: 3-EBAI Español 3 Lilia Castellanos

Evaluar alumnos

No. EXPEDIENTE	ALUMNO	EXAMEN PROGRAMADO	FECHA Y HORA SOLUCIÓN	CALIFICACIONES				TOTAL	VER	CALIFICAR	CALIFICAR	CAMBIAR NIVEL	CALIF.
				ESCRITO 45%	REDACCIÓN 15%	ORAL 20%	SCE 20%						
1 2016-4041 17620	Jennifer McDonald	Examen final Versión 12	28-04-2021 09:03 - 10:00	43	15	20	18	96 > 4	Ver	🔗	🔗	Español 4	43
2 2020-0722 17619	Tatál Ferreras	Examen final Versión 12	28-04-2021 09:03 - 10:37	35	14	19	17	85 > 4	Ver	🔗	🔗	Español 4	35
3 2020-0734 17621	William Austin	Examen final Versión 12	28-04-2021 09:03 - 11:03	31	13	17	14	75 > 4	Ver	🔗	🔗	Español 4	31
4 2021-0072 17626	Geoff Swanson	Examen final Versión 12	28-04-2021 09:05 - 10:59	42	15	20	20	97 > 4	Ver	🔗	🔗	Español 4	42

© Centro de Enseñanza para Extranjeros | Universidad Nacional Autónoma de México.

Figura 16. Imagen del sistema de gestión de exámenes finales. Evaluación.

4. Conclusiones

En este capítulo hemos querido mostrar la labor y la experiencia que el CEPE ha acumulado con el tiempo, en materia de elaboración de reactivos, evaluación y certificación; así como en el campo de las tecnologías enfocadas a la elaboración de sistemas, herramientas y aplicaciones de exámenes, tanto de forma presencial como en línea.

Hasta este momento, su funcionalidad ha permitido cubrir los objetivos que se propusieron al inicio de cada proyecto desarrollado. Sin embargo, a lo largo del tiempo, las circunstancias globales, el cambio acelerado de las tecnologías y los medios electrónicos, sin mencionar las necesidades del contexto actual, nos han obligado a reconocer que el sistema de gestión con el que cuenta hoy día el CEPE tiene la capacidad de transformarse en una plataforma más robusta; más necesaria todavía si sopesamos todas las

ventajas, carencias y experiencias de docentes, estudiantes e incluso las propias, como desarrolladores en ambos proyectos.

Estamos seguros de que la propuesta tecnológica derivada del presente proyecto PAPIME (y que se describe con mayor detalle en el próximo capítulo) trascenderá todas las limitantes que hemos experimentado con las anteriores. Confiamos en esto, ya que para llevarla a cabo se cuenta con un grupo interdisciplinario de trabajo, que pertenece a las áreas académicas de español, de evaluación y de tecnología educativa.

Llegamos a la conclusión de que el presente Sistema de Gestión de la Información de Exámenes (MIS, por sus siglas en inglés) debería ser capaz de cubrir determinados objetivos, mediante un uso eficaz de los recursos con los que se cuenta. Sin embargo, este sistema ha quedado limitado a la interacción de varios roles de usuario y a una interfaz diferente, más adecuada para los alumnos que para la administración de centros de aplicación y rúbricas integradas que asistan a los evaluadores de las habilidades oral y escrita. Por todo lo anterior, hemos pensado que es pertinente desarrollar una nueva plataforma virtual, que integre e interactúe con los sistemas con que actualmente cuenta la institución.

Para finalizar, sugerimos que para llevar a buen término este nuevo desarrollo tecnológico, se deben elaborar, además, los siguientes documentos de apoyo:

- a) Preparación de un manual sobre la gestión técnica, con los lineamientos generales del uso y asignación de accesos, para la administración de esta plataforma.
- b) Diagramas de flujo de información y navegación del sitio.
- c) Resultados de las pruebas piloto, así como una bitácora de ajustes y actualizaciones.

Al contar con estos documentos, se sentarán las bases para la creación de esta herramienta que, dadas las actuales circunstancias, no sólo es necesaria, sino también urgente.

Referencias recomendadas

Campos, E. (2005). Certificado de Español como Lengua Adicional (CELA). Nivel intermedio. *Decires*, 7, 55-66. Recuperado de <http://revistadecires.cepe.unam.mx/articulos/art7-3.pdf>

Campos, E., Espinosa, A. y Jurado, M. (2011). Exámenes de aprovechamiento computarizados. Un paso más en la evolución académica del departamento de español. *Decires*, 13, 121-158.

Granda, B. (2008). *Lineamientos para la elaboración de reactivos del examen CELA Básico*. México: UNAM. [Documento electrónico s/r].

Granda, B. (2015). Descripción del examen y avances en el proceso de desarrollo del CELA (Certificado de Español como Lengua Adicional). XII Encuentro Internacional del GERES. (pp. 194-205). Metz: GERES. Recuperado de <https://www.geres-sup.com/revue/les-cahiers-du-geres-n-7/>

Gutierrez, R. (2001). Quién tiene otra lengua vale por dos. Examen de Posesión de la Lengua Española (EPLÉ). *Decires*, 4, 11-21. Recuperado de: <http://revistadecires.cepe.unam.mx/articulos/art4-1.pdf>

Jurado, M. (2007). Propuesta de un banco de ítems para el desarrollo de exámenes que cumplan con los estándares exigidos por el SICELE. IV Congreso Internacional de la Lengua Española. Cartagena: CILE. Recuperado de: <https://congresosdelalengua.es/cartagena/paneles-ponencias/unidad-diversidad/jurado-martha.htm>

Jurado, M. y Pulido, G. (2011). Hacia una propuesta de un modelo de examen para el Certificado Internacional del Español CEPE-UNAM. *Decires*, 8, 49-67. Recuperado de: <http://revistadecires.cepe.unam.mx/articulos/art8-2.pdf>

Instituto Cervantes (2019). El español, una lengua que hablan 580 millones de personas, 483 millones de ellos nativos. Sala de prensa. Recuperado de https://www.cervantes.es/sobre_instituto_cervantes/prensa/2019/noticias/presentacion_anuario_madrid.htm

Desarrollos del CEPE

- Sistema CIE
- Sistema EPLÉ
- Sistema CELA
- Sistema EXELEAA
- Sistema de exámenes de aprovechamiento
- Sistema de exámenes de colocación
- Sistema de exámenes de admisión a diplomados y especialización

CAPÍTULO 10

PLATAFORMA DIGITAL PARA LA APLICACIÓN DE EXÁMENES DE APROVECHAMIENTO Y PARA LA INVESTIGACIÓN EN ELE: UNA PROPUESTA TECNOLÓGICA PARA LA EVALUACIÓN DE LAS HABILIDADES DE EXPRESIÓN ORAL Y ESCRITA

Maribel **Carmona Herrera**

Marcela **Nieto Rodríguez**

Edgar René **Pacheco Martínez**

En este capítulo, presentaremos las características y utilidades de la plataforma, producto final de este proyecto PAPIME; asimismo, profundizaremos en su administración y los beneficios que implicará su puesta en funcionamiento, tanto para la aplicación de exámenes, como para la recolección de datos con fines de investigación sobre ELE.

Con base en lo recién expuesto, el capítulo se desarrollará en tres partes:

1. Características: la importancia de una plataforma digital en la actualidad.
2. Administración: la arquitectura de la plataforma propuesta.
3. Fortalezas, áreas de oportunidad y los procedimientos de evaluación de EO y EE mediante la plataforma.

1. Características: la importancia de una plataforma digital en la actualidad

1.1. Introducción

Como hemos mencionado en capítulos anteriores, desde 2011 en el CEPE, los exámenes de aprovechamiento se han aplicado a través de una plataforma, a la que se accede por computadora. No obstante, después de 10 años, vemos que es imperante la necesidad de aplicar el examen a distancia o virtualmente; así que los cambios en el campo de la evaluación, en su adaptación a las nuevas realidades virtuales y electrónicas, son obligados para que los estudiantes presenten los exámenes desde cualquier parte del mundo.

Consideremos que el aprendizaje en línea es una apuesta cada vez más aceptada y, además, con resultados altamente satisfactorios; de modo que la transformación hacia un entorno digital marcha sin freno y ahora, debido a

la pandemia, se ha acelerado el proceso. Tras más de un año de confinamiento, la digitalización avanzó lo que inicialmente se estimaba para los próximos 5 o 10 años. En definitiva, la situación que estamos viviendo nos conduce a que desarrollemos e implementemos los procesos de evaluación a través de sistemas digitalizados, pero no sólo a causa de la situación actual, sino porque las evaluaciones en plataformas digitales también implican grandes beneficios (Sepúlveda, 2020), tal como sucede con las pruebas SIELE (www.siele.org) o EXELEAA (<https://www.cepe.unam.mx/exeleaa/>); este último, uno de los exámenes de certificación del CEPE.

Las nuevas tecnologías representan un gran empuje en lo referente a softwares para la elaboración y administración de pruebas objetivas de corrección en línea, así como las tareas orales y escritas que sean calificadas a distancia por los expertos. Por otro lado, es importante que la plataforma digital tenga los candados de seguridad necesarios, para evaluar de forma protegida el impacto del proceso de enseñanza-aprendizaje de los estudiantes.

1.2. Objetivos de los instrumentos de evaluación

A nivel educativo, en muchas sociedades la funcionalidad de la evaluación cumple dos objetivos: el social y el pedagógico (Bourdon, 2006). No es una excepción a estos parámetros cuando se evalúa una lengua en L2, en el marco de la instrucción formal. El objetivo social se da por las exigencias que la misma sociedad demanda de la acreditación de programas educativos, mediante la ponderación de una calificación; es decir, calificar a estudiantes de manera individual, en función de su dominio de los contenidos de un determinado curso. En cuanto al área pedagógica, está encausada a la evaluación, la cual se acompaña del proceso de enseñanza-aprendizaje, de tal manera que aporte la retroalimentación según los objetivos de aprendizaje (Sepúlveda, 2020).

1.3. Ventajas de la plataforma digital

Como institución, es recomendable estimar los criterios de calidad: desde el diseño hasta la implementación del instrumento en una plataforma o un software. Por eso, es importante establecer una comunicación eficiente, a favor de la evaluación válida y confiable, entre el equipo académico (expertos en el diseño y planeación del instrumento de evaluación) y el equipo especializado en tecnología educativa (que responde y se ajusta al diseño y necesidades del instrumento). Antes de mencionar las principales ventajas o beneficios, hay que destacar que dentro de la plataforma digital propuesta, se pretenden elaborar las tablas de especificaciones y, con ellas, la elaboración de las tareas de las expresiones; es decir, que

todo el proceso de diseño y aplicación se hará desde la plataforma. Esto, con el objeto de practicar una metodología rigurosa en la evaluación, que va desde contar con todos los documentos involucrados en el proceso, alojados en un espacio virtual y de fácil acceso para todos los participantes del proyecto.

Primera ventaja: se flexibiliza la evaluación, pues a diferencia de la modalidad tradicional, la tecnología da mayor interactividad, más comodidad y, además, potencia los métodos habituales para evaluar. Lo que antes se hacía en papel, ahora se realiza desde la computadora, sin el uso de una plantilla ni lápiz, en un espacio y tiempo determinados; además de que se asignan tiempos establecidos para la realización de cada prueba. Es decir que al implementar la evaluación digital, se eliminan las barreras espacio-temporales, ya que éstas sencillamente dejan de existir, lo cual nos permite integrar y concentrar nuestras evaluaciones en diferentes herramientas y servicios.

Segunda ventaja: ésta resulta relevante por las condiciones actuales del confinamiento por la pandemia: el hecho de digitalizar la evaluación es que sobrepasamos las barreras geográficas, gracias a lo cual, es posible evaluar en cualquier parte del mundo. Por ejemplo, si un alumno tomó clases de español en el CEPE de CU, pero por una emergencia regresó a su país un día antes de la fecha de examen, podrá realizar su prueba de aprovechamiento en alguno de los centros o instituciones autorizados por el CEPE-UNAM.

Otro supuesto: si un alumno atendió sus clases de ELE desde China o Corea, al término del curso también podrá hacer el examen vía remota, desde su casa. Y en el caso de las pruebas de expresión escrita, es importante decir que se incluye un campo de texto y un teclado en español, previendo que en el país donde el alumno se examine, el teclado estándar no cuente con los caracteres necesarios, y que de esa manera, los tenga a su disposición dentro de la plataforma.

Tercera ventaja: se pueden construir bancos de pruebas y de tareas con los cuales nutrir nuestra evaluación y, con ello, mostrar mayor variabilidad. Es decir, al tener un amplio número de modelos de examen, el sistema permite que las diferentes versiones se distribuyan de forma aleatoria para los alumnos que se presentan al examen, y de esa manera no se repita un mismo modelo. Por otro lado, se contará con una nutrida serie de corpórea de datos de producciones escritas y orales, de alumnos que se hayan evaluado en ELE.

Cuarta ventaja: al final del proceso, la información queda almacenada, guardada en los servidores. Esto representa seguridad para la institución, porque toda la información se encripta. Será una forma de proteger la evaluación, para evitar que esté en riesgo de vulnerabilidad.

Quinta ventaja: ésta implica que existe confidencialidad cuando hacemos evaluaciones digitales a cada alumno que se examina. En lugar de publicar su nombre y apellidos, podemos asignar códigos para que, en el proceso de calificación, se tenga total confidencialidad y todo sea anónimo.

Sexta ventaja: ésta tiene relación con la anterior, ya que los calificadores expertos evaluarán de la manera más objetiva posible al no contar con el nombre del alumno, y en el caso de la expresión oral, podrán repetir la grabación cuantas veces sea necesario, lo cual no es posible en la evaluación tradicional, que es “efímera”. En otras palabras, es momentánea y no hay manera de recuperar las tareas orales del alumno, ni tampoco los evaluadores pueden prestar atención al 100% a la producción del alumno, ya que desempeñan diferentes roles durante la aplicación del examen.

Séptima ventaja: en nuestra experiencia con los candidatos que aplican para el examen SIELE, hemos observado que en la prueba de expresión oral, el alumno se siente más cómodo frente a la computadora, pues así graba sus tareas sin la presión del evaluador. Así que, suponemos que ocurrirá algo semejante con nuestros alumnos, en contraste con la forma tradicional.

Finalmente, será más rápido de generar la emisión del resultado del alumno, ya que el sistema calificará las pruebas objetivas –CA, CL y competencia lingüística, que se pretenden integrar posteriormente–, después de ser contestadas; el sistema también repartirá las tareas de expresión escrita y oral a los profesores expertos que se hayan estandarizado para calificar niveles de español en particular. También puede existir la posibilidad de brindar retroalimentación, de forma accesible, a aquellos estudiantes que lo soliciten; pero eso estará determinado por lo que estipule la institución.

2. Administración

2.1. Desarrollo, funcionamiento e implementación

De acuerdo con todo lo mencionado (y con la experiencia de las plataformas y sistemas de gestión, y aplicación de exámenes vistos en el capítulo anterior), ahondaremos de forma más detallada acerca de la estructura y funcionamiento de la plataforma propuesta para este proyecto.

Esta plataforma busca aportar varias ventajas, como la gran flexibilidad en los tiempos, el recurso de nuevos medios digitales, la constante actualización y la reducción de los costos (tanto para la dependencia, como para el alumno) en cada aplicación, y continuidad en su formación académica.

Esta solución, junto con la experiencia profesional de cada involucrado en

el proyecto y las habilidades examinadas en este libro, permite responder de forma eficiente a los desafíos que identificamos como áreas importantes para su atención; todo lo cual proporciona una herramienta y una metodología más adecuadas para garantizar la viabilidad de una aplicación robusta, una calificación objetiva y un seguimiento fácil, intuitivo y seguro durante todo el proceso de evaluación. Los módulos que se tienen contemplados en la arquitectura de la plataforma y su relación con otros sistemas son los siguientes:

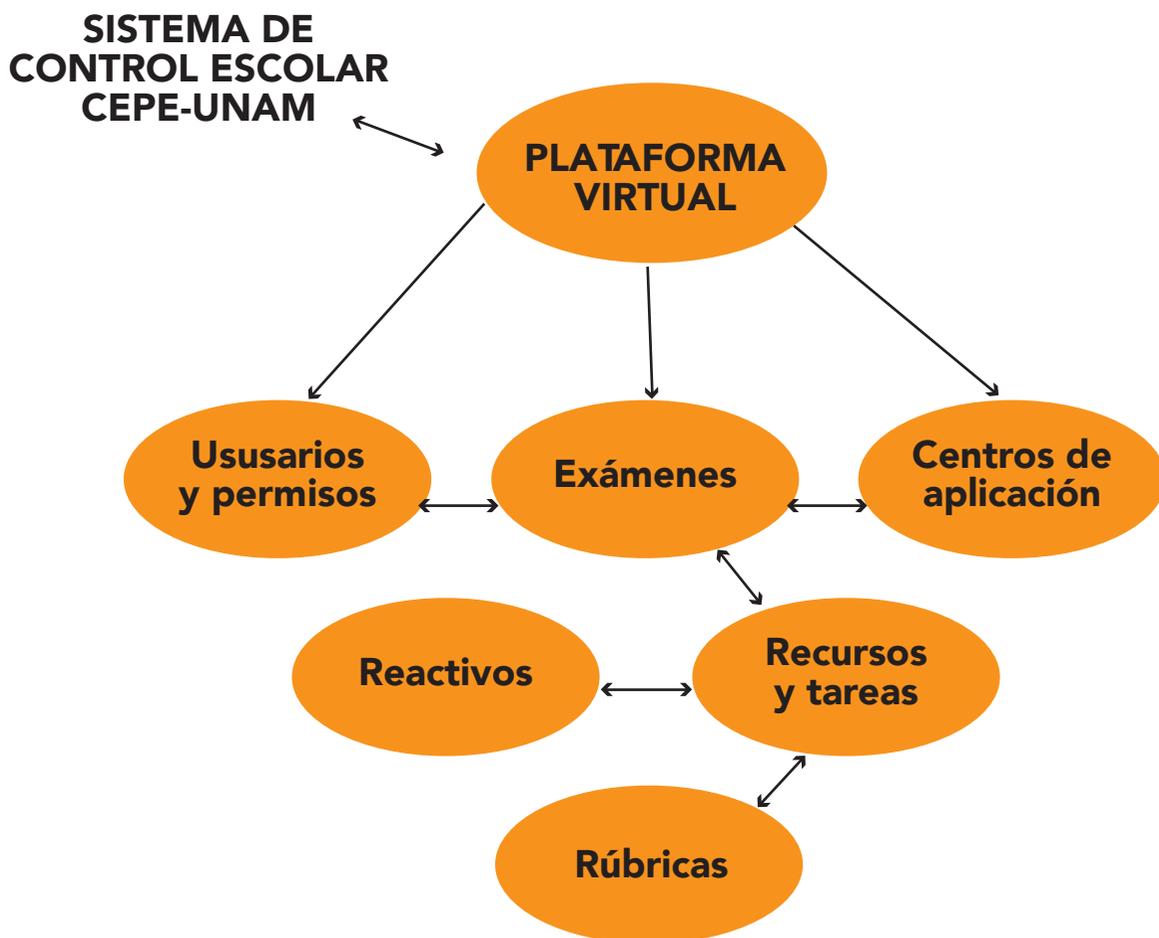


Figura 17. Diagrama del sistema de Control Escolar del CEPE, UNAM.

Configuración de usuarios y permisos

- Administrar los distintos usuarios que podrán tener acceso. En este módulo se agregan, editan y eliminan las personas encargadas de configurar y alimentar la plataforma.
- Controlar el acceso de los diferentes roles de usuarios, todos con un perfil especializado y limitado para el uso y manejo de la plataforma.
- Dar alcance y seguimiento de los distintos módulos que tendrá el sistema, ya que participarán varios expertos elaboradores de reactivos, conformadores de exámenes, evaluadores, administradores generales y, por supuesto, los estudiantes.

Configuración de exámenes

- Crear, listar, editar, eliminar y copiar las distintas versiones que se integren en la plataforma.
- Administrar la publicación y el estatus del examen (integración, pilotaje, liberado).
- Integrar las secciones junto con los recursos y tareas del banco de reactivos.
- Revisar las estadísticas de aplicación.
- Gestionar el tiempo o duración de cada examen, de acuerdo al nivel.
- Gestionar el número de evaluadores por habilidad.
- Configurar el orden aleatorio de preguntas y opciones de respuestas.
- Ajustar la seguridad, vigilancia y supervisión.

Configuración de centros de aplicación

- Administrar los distintos centros en los que se aplicarían los exámenes, ya sean dependencias internas o externas de la UNAM. En este módulo se agregan, editan y eliminan estas sedes.
- Gestionar la asignación de exámenes a los centros aplicadores.
- Administrar las fechas de aplicación de los exámenes liberados.
- Listar las estadísticas de las aplicaciones.

Configuración de reactivos

- Administrar la creación, edición y eliminación de reactivos, para conformar un banco relacionado, indexado y clasificado por nivel de español, habilidad y tipo de reactivo.
- Administrar el estatus de los reactivos y el valor de cada uno, de acuerdo a la complejidad y nivel de español.

- Buscar los reactivos, de acuerdo a los criterios de tipo, nivel, complejidad, etc.

Configuración de recursos y tareas

- Administrar la creación, edición y eliminación de recursos y tareas, para conformar un banco relacionado, indexado y clasificado por nivel de español, habilidad y tipo de recurso o tarea. Por ejemplo: expresión oral, expresión escrita, comprensión de lectura, comprensión auditiva, etc.
- Integrar tablas de especificación, de acuerdo con el nivel de español, para la creación de cada recurso o tarea.
- Buscar los recursos o tareas, de acuerdo a los criterios de tipo, nivel, complejidad, etc.

Configuración de rúbricas

- Dependiendo de la habilidad que se aborde, se podrán configurar las rúbricas para que los evaluadores sean capaces de utilizarlas, y darle así un peso específico a los criterios que se consideren relevantes en una configuración determinada. Por ejemplo: para la EO, se considerarían criterios como pronunciación, vocabulario, coherencia y cohesión (claridad, jerarquía y conexión de ideas) e interacción (adecuación al interlocutor, toma de turnos, etc.), entre otros.
- El sistema hará los cálculos pertinentes, de acuerdo con el puntaje y la habilidad ingresados en este módulo.

Configuración de reportes

- Analizar y obtener informes.
- Analizar el comportamiento de los elementos que integran la plataforma.

Algunas ventajas que se tomaron en cuenta, respecto al sistema de exámenes que tenemos actualmente, son:

Sistema innovador

Genera una mayor satisfacción de las partes implicadas (estudiante, evaluador, administrador de la plataforma).

Flexibilidad

Se adecua a las particularidades de cada centro de formación y aplicación, así como al tipo de habilidad a evaluar, o al tipo de examen a presentar.

Interfaz sencilla

Facilita la configuración y ejecución de los exámenes; elimina la posibilidad de errores que puedan cometer los estudiantes al momento de contestarlo o navegarlo.

Seguridad

Registro con cámara web. Activa la grabación de audio y/o video en la habilidad oral.

Modo de pantalla completa. Activa el modo de pantalla completa para evitar que los participantes abran nuevas ventanas, pestañas, aplicaciones, etc.

Supervisión en bitácoras de acceso, comportamiento y cierre de sesión, tanto de los estudiantes como de los administradores de la plataforma.

1. Tecnología

La tecnología utilizada para la programación de esta plataforma utilizará un software libre, el cual contribuye al desarrollo de nuevas aplicaciones, sin la necesidad de adquirir un software comercial; permite, además, obtener aplicaciones que tienen utilidad y que pueden ser mejoradas. El desarrollo de una plataforma propia facilita los ajustes, correcciones y actualizaciones de forma transparente, adaptable y configurable a las necesidades del usuario.

2. Compatibilidad

Basados en la experiencia anterior en el desarrollo de los sistemas administrativos y académicos del CEPE, la compatibilidad en el diseño, programación e implementación de esta plataforma será completamente transparente; también podrá conectarse con nuestros principales sistemas de gestión de información, así como con el sistema de Control Escolar, para gestionar a los estudiantes, sus calificaciones globales y el seguimiento de sus evaluaciones periódicas.

3. Fortalezas, áreas de mejora y los procedimientos de evaluación de EO y EE por medio de la plataforma

Las características de la comunicación a través de los medios digitales, específicamente en lo que respecta a la educación, con el paso de los años han demostrado que estamos siendo testigos de un cambio de paradigma en esta materia: la interactividad hace posible que la información se transmita de manera bidireccional y diversa, en soportes electrónicos; la multiplicidad de espacio y lugares virtuales de aprendizaje potencia las oportunidades y experiencias de adquirir conocimientos; la

movilidad y la ubicuidad posibilitan que cualquiera con acceso a estos medios entre en contacto con oportunidades de enseñanza y aprendizaje, en cualquier momento y lugar del planeta.

Por lo anterior, es posible afirmar que la tecnología digital aplicada a la educación no sólo constituye una herramienta a su servicio, sino también una verdadera extensión del *espacio* donde esta práctica se lleva a cabo. Los medios digitales son capaces de expandir el territorio donde tiene lugar la enseñanza –antes, dominio exclusivo de la escuela–, y proporcionar nuevos ambientes de aprendizaje, medios de expresión para los profesores y estudiantes, así como nuevas formas de interacción entre ellos. El salón de clases es concebido ahora como un espacio virtual metatópico (del griego μετά, “más allá”, y de τόπος, “lugar”), que ha visto sus fronteras físicas rebasadas. Se habla también de nuevas formas de evaluación, apoyadas por este cambio paradigmático que nos traslada de la escuela a los espacios digitales, como nodos de transmisión de conocimientos: la evaluación auténtica, la evaluación de actuación integrada, la evaluación dinámica, la evaluación participativa, etc. (Bordón, 2007: 236-240).

En todo este contexto, la plataforma digital que estamos desarrollando tiene como objetivo fundamental ofrecer al CEPE (y, por supuesto, a otras instituciones de la UNAM) nuevas alternativas para cubrir un amplio rango de necesidades, tanto del profesorado como de la comunidad estudiantil. De este modo, ya expuesta en la sección anterior la organización arquitectónica de la plataforma digital y sus características técnicas, conviene ahora hacer una reflexión en torno a sus fortalezas y áreas de oportunidad, desde una perspectiva de los profesores y alumnos en su respectivo rol de usuarios.

3.1. Fortalezas y áreas de oportunidad

Si bien algunas predicciones de la ciencia ficción nos han alcanzado –y muchas veces hasta rebasado nuestras expectativas sobre lo que hasta hace poco era el futuro próximo–, aún no hemos llegado al punto en que las máquinas sean capaces de sustituir por completo a los profesores en la tarea de planificar un curso, enseñar y evaluar las capacidades de nuestros alumnos. Es verdad que las tecnologías digitales avanzan y cada año nos sorprenden con capacidades que antes nos parecían imposibles, como el procesamiento del habla y el de la escritura. Como ejemplo de esto, en Bangalore (el Silicon Valley de India), se ha desarrollado ya una inteligencia artificial que, según sus creadores, es capaz de escribir un *best seller* en inglés.

No obstante, hoy día los profesores seguimos siendo, afortunadamente, indispensables. En este preciso momento del siglo XXI, tenemos el reto de

encontrar maneras de encuadrar la tecnología en nuestro trabajo, no para sustituirnos, sino para potenciar nuestras capacidades, mejorar nuestra enseñanza y facilitar el aprendizaje de nuestros alumnos. En efecto, no se trata de un juego en el que uno de los bandos (humanidad y tecnología) deba triunfar sobre el otro, sino de una partida en la que ambos podemos ganar. Desde este punto de vista, hemos partido para escribir este capítulo: la plataforma digital que aquí presentamos tiene como principal propósito modernizar, agilizar y facilitar la administración, aplicación y calificación de los exámenes de aprovechamiento de EO y EE en nuestro centro.

A continuación, desglosamos las fortalezas de esta plataforma digital. Como se ha descrito en la sección anterior, ésta cuenta con varios módulos en su diseño arquitectónico: configuración de reactivos, de exámenes, de sedes de aplicación, de usuarios, de permisos, de rúbricas y de reportes. En seguida se detallan ciertas ventajas que ofrecen algunos de estos módulos para los usuarios, profesores y estudiantes.

3.1.1. En el módulo de configuración de exámenes se tienen diversas utilidades, entre las cuales, la más importante es que los usuarios administradores cuentan con la posibilidad de determinar las fechas de aplicación, de forma automatizada. Asimismo, es factible determinar cuántas y cuáles versiones de exámenes estarán a disposición del sistema para su asignación (aleatoria o no) a los estudiantes.

En segundo lugar, este módulo cuenta con una herramienta para el procesamiento de datos y la elaboración de estadísticas sobre la aplicación de exámenes, así como para el análisis psicométrico de tareas y reactivos. Valdría mucho la pena poder saber, por ejemplo, cuántos estudiantes han participado en la jornada y qué modelos fueron utilizados; cuánto tiempo fue necesario para la realización de las tareas, qué tareas o reactivos representan mayor dificultad y cuáles relativa facilidad; qué grupos nacionales tienen mayores dificultades con ciertas tareas y a cuáles se les facilitan, con el fin de determinar si la L1 es un factor interviniente o no, etc. A fin de alcanzar el objetivo anterior, es imprescindible contar con información clara y detallada, que sería de mucho provecho para la recolección de datos con fines de investigación académica, y para la evaluación de nuestra metodología de enseñanza.

3.1.2. Los usuarios administradores que empleen la plataforma en otras sedes –incluso, en un futuro, la plataforma podría ser utilizada por instituciones externas, nacionales o extranjeras– serán capaces de definir las características de sus propias evaluaciones; esto, claro, dependiendo siempre de sus circunstancias, condiciones y necesidades específicas en el

módulo de configuración de sedes de aplicación. Tal como ocurre en algunos restaurantes, en los cuales el cliente marca en una lista los ingredientes de su platillo favorito, este módulo facilitará a los administradores en sedes externas la configuración de sus exámenes, sus fechas de aplicación, las modalidades de calificación, análisis de resultados, etc.

3.1.3. En relación con el módulo de usuarios profesores, la interfaz permite que el administrador y el coordinador de evaluadores realicen varios tipos de consultas sobre las acciones llevadas a cabo por los maestros, en sus distintos roles: creador de reactivos y tareas, validador, entrevistador en exámenes orales, calificador, calificador en entrenamiento, etc. En este punto, vale la pena detenernos un poco para mencionar que, como se ha señalado en los capítulos 5 y 8, una de las áreas de mejora que encontramos en los actuales procedimientos de los exámenes de aprovechamiento de EO y EE, es que no existe una capacitación constante para estandarizar los criterios de los profesores que participan en estas evaluaciones. La plataforma posibilitará dar seguimiento a los profesores en entrenamiento y a los que ya pasaron por él, tanto para el uso de esta herramienta, como para el de los instrumentos de evaluación y la estandarización de sus criterios como evaluadores. A propósito de esto, se impartiría un curso-taller dos veces al año (como mínimo), para capacitar a nuestros profesores en el empleo de esta plataforma.

En segundo lugar, este módulo también permite que el administrador, el coordinador de evaluadores y los evaluadores mismos realicen seguimientos de estudiantes, de forma individualizada o grupal; esto, según los criterios de búsqueda, que dependerán de las necesidades de análisis para la Coordinación de Español, o para investigación de los académicos interesados.

Por último, mediante este módulo se podrá organizar, de forma automatizada, una serie de catálogos de usuarios profesores (creadores y revisores de tareas y reactivos; calificadores expertos y en entrenamiento; entrevistadores, etc.) y de usuarios estudiantes. Además de lo anterior, dichos catálogos ofrecerán la posibilidad de emitir un historial del comportamiento de cada usuario, para su seguimiento diacrónico y su posterior análisis y evaluación.

3.1.4. El módulo de rúbricas, igualmente, tiene una interfaz amigable con el usuario creador de reactivos y tareas, y con el usuario calificador. Dado que el diseño de los modelos de exámenes y la labor de evaluación requieren que los profesores involucrados estén familiarizados con las rúbricas, es importante que éstas se encuentren a disposición de los usuarios docentes y, en algunos casos, de los estudiantes. Sin embargo, la plataforma ofrece también la alternativa de presentar o de ocultar las rúbricas, dependiendo de la configuración de aplicación de pruebas.

Este módulo cuenta, además, con secciones para la creación de rúbricas, según la habilidad que se desee evaluar, su validación y su pilotaje. En éstas, los profesores evaluadores tienen como apoyo herramientas para dejar comentarios, sugerencias y observaciones, sin que por ello se abandone la tarea de afinar estos instrumentos de evaluación en reuniones colegiadas. Por último, es necesario mencionar la sección de cálculo de puntajes, que arrojará de forma automática los resultados del desempeño de cada estudiante o grupo de estudiantes –por nacionalidad, por grupo de edad, por nivel de escolaridad, etc–; lo cual, de forma adicional, generará concentrados de información que sirvan como materiales para la investigación académica.

3.1.5. El módulo de configuración de reportes no es menos importante que los anteriores, pues por este medio se generarán reportes que recojan datos de relevancia en diferentes niveles: el de los instrumentos de evaluación, como rúbricas y modelos de exámenes; el del rendimiento de los reactivos y tareas; el del comportamiento de los profesores creadores y calificadores, y el de los estudiantes usuarios.

Consideramos que con los actuales procedimientos para evaluar, se pierde mucha información que es de gran valor para investigar la evaluación en nuestro propio centro educativo. Sin embargo, la plataforma digital tiene la capacidad de conservar datos por un largo tiempo, procesarlos y generar información digerible para investigadores, cuya labor en este sentido sólo podría derivar en el mejoramiento de nuestro trabajo.

3.2. Los procedimientos para la evaluación de EO y EE por medio de la plataforma

La administración de las pruebas de EO y EE, tal como se realizan en la actualidad, se describen puntualmente en los capítulos 5 y 8. Dado que la mudanza de las formas presenciales de evaluar a una evaluación en medios digitales –o “e-evaluación”, como se ha denominado en la literatura especializada– implica cambios en varias dimensiones, que serán tratados con mayor holgura en el capítulo 11, describiremos aquí, *grosso modo*, algunos aspectos de los procedimientos de evaluación de EO y EE que proponemos.

Puesto que la plataforma es una herramienta digital con conectividad a internet, permite la aplicación de pruebas de EO y EE en cualquier lugar del planeta y en cualquier huso horario. En el caso de los exámenes orales, posibilita realizarlos en tiempo real, siempre que se necesite, mediante entrevista directa entre profesor y alumno; o bien, es posible también grabar las producciones de los estudiantes para su posterior calificación.

El hecho de que las producciones orales de los examinados puedan ser registradas (tanto en video, como en audio), permite que los calificadores las evalúen en cualquier momento que deseen.

De igual forma, contar con registros audiovisuales y escritos de las producciones de los estudiantes es beneficioso, porque facilita la revisión de exámenes, sobre todo orales, que, de lo contrario, sólo podrían revisarse mediante registros escritos y testimonios de los evaluadores y estudiantes participantes. Sin embargo, lo anterior no es la única ventaja, sino que dichos registros conforman materiales que pueden llegar a constituir corpórea de investigación, mediante los permisos legales necesarios.

Otro aspecto positivo de la plataforma es su gran flexibilidad para configurar diferentes modalidades en la aplicación de los exámenes. Esto, como se ha mencionado líneas atrás, responde a que cada sede podría requerir exámenes con determinadas características, con la finalidad de cubrir necesidades diversas y muy específicas, según sus propios contextos y circunstancias. De igual manera, nos pareció importante que la plataforma tuviera esta flexibilidad con el objetivo de generar datos en condiciones especiales, según los requerimientos de cualquier investigador.

Y, en último lugar, dicha flexibilidad permite que los administradores experimenten con distintos procedimientos, a fin de dar con los más adecuados, según las circunstancias y necesidades de cada cual. La adaptabilidad de la plataforma, estamos convencidos, la convertirá en una herramienta provechosa, no sólo para CEPE-CU, sino también para las sedes de Taxco y Polanco, los CEM y algunas otras dependencias de la UNAM que se interesaran en hacer uso de ella.

Un ejemplo de lo anterior es que el módulo de configuración de exámenes puede determinar si en una prueba oral, por poner un caso, participará un calificador, dos o incluso tres (dependiendo, claro, de la disponibilidad de profesores evaluadores o las necesidades de una investigación en proceso). Asimismo, dado que los exámenes pueden realizarse en tiempo real, o bien registrarse por escrito y en medios audiovisuales para su posterior calificación, los evaluadores revisarán las producciones de forma sincrónica o asincrónica. En cualquiera de los dos casos, los profesores podrán hablar en tiempo real o registrar mensajes de texto en los cuales argumenten el porqué de sus puntajes, a fin de que otros evaluadores conozcan sus razones, y puedan llegar a decisiones consensuadas en torno al desempeño de los estudiantes.

Otro aspecto relevante es que se propone que el profesor que imparte el curso, en los casos en que los administradores de la plataforma lo consideren pertinente, contribuya en las evaluaciones de EO y EE; esto, aportando la información que haya recogido –en forma de bitácora o portafolio de evidencias– a lo largo del curso sobre el aprendizaje de sus

estudiantes. O, incluso, que participe en las evaluaciones de aprovechamiento de EO y EE, contribuyendo con sus apreciaciones y puntajes, que se promediarían con la de los otros evaluadores. La finalidad de experimentar con distintas modalidades de evaluación es dar con las más eficaces, a fin de reducir el factor de subjetividad y lograr evaluaciones lo más objetivas posibles, e incluso, más democráticas. Sobre esto último, la arquitectura de la plataforma permitirá que en un futuro se incluyan utilidades, para que los estudiantes participen en auto y coevaluaciones.

Por último, en lo que respecta a los exámenes orales, sus registros audiovisuales permitirán que los calificadores reproduzcan cuantas veces sean necesarias las actuaciones de los estudiantes. Esto abre un gran abanico de posibilidades para experimentar, también, con el uso de las rúbricas. Por ejemplo, podrá configurarse el examen, a fin de evaluar un rubro por cada reproducción, para que el calificador se enfoque en un aspecto en específico a la vez, o bien, evaluar varios o todos los rubros simultáneamente; esto también será posible en el caso de la expresión escrita.

4. Conclusiones

La plataforma digital para los exámenes de aprovechamiento de EO y EE que aquí presentamos debe tener dos objetivos fundamentales: en primer lugar, modernizar nuestra forma de evaluar las habilidades de producción; y, enseguida, servir como puente de transición, para pasar de modos de evaluación tradicionales a aquellos más innovadores, conforme a las necesidades de los tiempos actuales. Tal aspecto quizá constituya el alcance más importante de este proyecto.

Es verdad que integrar una evaluación en un medio digital conlleva una inversión fuerte de tiempo y diseño. Sin embargo, la principal ganancia en esta ingente tarea será obtener una evaluación objetiva, válida, confiable, estandarizada y, por añadidura, un ahorro mucho mayor a largo plazo en la gestión de este proceso; como por ejemplo, el ahorro de papel y el acortar plazos en la entrega de resultados.

Por otro lado, para prever la buena culminación de este proyecto, es importante considerar aspectos que interactúan entre ellos y que contribuyen a su sostenibilidad. Los tres núcleos importantes son: 1) la relación provechosa entre el profesorado del centro; 2) la colaboración idónea entre estos y la institución, y 3) las condiciones de cooperación que permitan trabajar de forma adecuada, y que se perciban como justas y factibles (Figueras y Puig, 2013). Estas cuestiones organizativas están dadas en el CEPE, ya que el proyecto cuenta con la aceptación del centro y de la comunidad educativa; además, es percibido como innovador y útil. El primer núcleo debe responsabilizarse de cuestiones de contenido (especificaciones de

la prueba, redacción de reactivos y tareas de producción, etc.). El segundo núcleo tiene que ver con las cuestiones organizativas (quién se ocupa de qué, cuándo debe estar lista la prueba para ponerse en práctica, etc.). El tercer núcleo se encarga de las cuestiones administrativas (lo que se cambia en el sistema o en el centro y lo que no, etc.).

Acerca de los aspectos técnicos de la plataforma digital, sus aportaciones al centro son diversos, pero entre los más importantes podemos destacar que, mediante su uso, seremos capaces de realizar pilotajes frecuentes de los bancos de reactivos y de los modelos de exámenes; esto dará solvencia al centro, para el mantenimiento de sus instrumentos de evaluación. Para la primera fase de pruebas de la plataforma, se hará un pilotaje de los instrumentos correspondientes al nivel 3. Este banco de reactivos estará validado, además, mediante una metodología adecuada para su evaluación y seguimiento.

Asimismo, el sistema de la plataforma nos facilitará la administración y operación de las distintas secciones que lo componen, así como una etapa de pruebas, ajustes y retroalimentación. Hemos pensado en la posibilidad de que una vez que el sistema pueda ser completamente utilizado para la aplicación de exámenes departamentales de las habilidades de EO y EE, las autoridades del CEPE puedan considerar factible extenderlo a otros tipos de exámenes, como de colocación, extraordinarios y, en un modo más robusto, los exámenes de certificación.

Referencias

Blázquez, F., Alonso, L. y Yuste, R. (2017). *La evaluación en la era digital. Tecnología educativa*. Madrid: Síntesis.

Bordón, T. y Liskin-Gasparro, J. (2007). Evaluación. En Manel Lacorte (coord.). *Lingüística aplicada del español*. Madrid: Arco Libros, 211-252.

Bordón, T. (2006). *La evaluación de la lengua en el marco de E/L2: bases y procedimientos*. Madrid: Arco Libros.

Figueras, N. y Puig, F. (2013). *Pautas para la evaluación del español como lengua extranjera*. Madrid: Edinumen.

Sepúlveda, J. [CEPE Videos]. (2020, noviembre 4). La gestión de la evaluación mediante una plataforma digital. En Ciclo internacional de conferencias sobre evaluación. [Archivo de Vídeo]. Recuperado de: <https://www.youtube.com/watch?v=OqDHgbKsxpM>

CAPÍTULO 11

EVALUACIÓN DE LAS HABILIDADES DE EXPRESIÓN ORAL Y ESCRITA MEDIANTE LAS TAC

Diana Helena **Martínez Cruz**
Edgar René **Pacheco Martínez**

1. Introducción

En las últimas dos décadas, el interés por la tecnología en su imbricación con la enseñanza de lenguas se ha enfocado principalmente en los usos educativos de los desarrollos tecnológicos. Sin embargo, si consideramos que la enseñanza en general está mudando de lo tradicional a lo que en la literatura especializada se ha dado en llamar el *e-learning* (o aprendizaje electrónico), entendiendo por éste la “formación que se realiza a través de internet”; en otras palabras: “entornos virtuales de aprendizaje, que son espacios que utilizan la tecnología de la información y la comunicación y están diseñados con un propósito eminentemente educativo, donde el alumnado, como eje central del proceso de aprendizaje, utiliza herramientas síncronas y asíncronas para establecer una relación con el objeto de aprendizaje” (Blázquez, Alonso y Yuste, 2017: 37). Con esto en mente, debemos estar conscientes de que nuestras concepciones, prácticas y herramientas sobre la evaluación del proceso de enseñanza-aprendizaje, también tienen que estar sujetas a readaptaciones, conforme a sus propios tiempos y realidades.

Debemos tomar de ejemplo cómo los ambientes virtuales de aprendizaje se adaptan a las necesidades formativas de cada estudiante, y no a la inversa. El *blended learning* (o aprendizaje híbrido), que combina la educación virtual con las clases presenciales; el *mobile learning* (o aprendizaje móvil), que por su accesibilidad permite a un estudiante seguir aprendiendo en cualquier rincón del mundo; los MOOC (o cursos abiertos masivos online), entre tantas otras opciones, son sólo una pequeña muestra de la diversidad de alternativas con que la educación se ajusta a los nuevos tiempos. Esta misma variedad nos obliga a pensar que debemos impulsar un cambio en las formas como evaluamos la enseñanza-aprendizaje del español como LE en nuestro centro y, también, fuera de él.

Douglas (2017: 115-116) imagina, por ejemplo, que en un futuro próximo, el campo de la enseñanza de lenguas cuente con tecnologías que sean capaces de evaluar la producción oral, tal como se hace persona a persona; que bastaría con que el estudiante encienda su dispositivo electrónico para interactuar con una inteligencia artificial que podrá dar puntajes objetivos a las muestras de lengua registradas. En el caso de la expresión oral, todavía no hemos salido del ámbito de la ciencia ficción; no obstante, en el de la expresión escrita ya existe un software capaz de evaluar la

producción de los examinados en nanosegundos, y con mínimos porcentajes de diferencia con respecto a los calificadores humanos. Sobre esto, ahondaremos más adelante.

No cabe duda de que el vínculo entre educación y tecnología nos hace soñar e imaginar todas las posibilidades que surgen y suenan muy promisorias. Sin embargo, se debe tener cuidado de no rendirse ante las maravillas de las nuevas tecnologías, e integrarlas sin cuestionarse qué herramientas son las ideales y qué objetivos son los que buscamos. En otras palabras, no se trata de usar la tecnología por el simple hecho de usarla, pensando que un examen montado en una pantalla de computadora es ya un ejemplo de un uso adecuado de las TIC y las TAC en la educación¹. La integración de la tecnología en el campo educativo tiene que lograrse a través de una planificación cuidadosa, donde los recursos tecnológicos que empleamos sean pertinentes y respondan a nuestras necesidades específicas; también, mediante una reflexión acerca de las formas en que nuestro método de enseñanza, nuestros contenidos, nuestra forma de evaluar (es decir, que todo lo que engloba nuestra filosofía pedagógica) se adapta al nuevo medio en que tiene lugar el proceso de enseñanza-aprendizaje.

A partir de lo anterior, nos parece indispensable y urgente detenernos a pensar cuidadosamente qué implica evaluar lengua mediante las nuevas herramientas digitales. Por tanto, el objetivo de este capítulo es exponer las características que adquiere la evaluación de las habilidades de expresión, oral y escrita, a través de las tecnologías de aprendizaje y conocimiento (TAC), los retos que esto supone, así como los pros y contras de dicho cambio.

2. Evaluación y tecnología

2.1. Criterios de evaluación de la expresión oral y escrita en medios electrónicos

Este apartado retoma los criterios esenciales de las habilidades productivas, punto en el que ahondaremos más adelante. Los conceptos de validez y confiabilidad son inherentes a la evaluación, razón por la cual deben estar presentes durante el diseño de cualquier instrumento. Siguiendo con los contenidos planteados anteriormente, debemos tener en cuenta qué, cómo y para qué evaluar, tanto en la expresión oral como en la

1. Hoy en día, existen muchas denominaciones para las tecnologías aplicadas a la comunicación y a la educación: TIC, TEP y TAC. Román-Mendoza (2018) y Herrera (2019), entre otros, señalan que aún no se llega a un acuerdo unánime en cómo denominarlas. Por TIC entenderemos las tecnologías de la comunicación y la información que tienen como rasgo principal la difusión de información en medios digitales; mientras que cuando hablamos de TEP, nos referimos a las Tecnologías del empoderamiento y la participación, que en sí son las TIC aplicadas al fomento de la participación ciudadana, en temas políticos y sociales. Por otro lado, con TAC hacemos referencia a las Tecnologías del aprendizaje y del conocimiento, aquellas tecnologías aplicadas a la pedagogía y a la formación educativa, con un menor énfasis en la dimensión informática, pero sin soslayarla.

escrita, sin perder de vista que los constructos a evaluar estarán perfectamente definidos de manera operacional. No obstante, al mudarse de un medio convencional a una modalidad electrónica, habrá por supuesto cambios, y se tendrán que hacer todos los ajustes necesarios; así como también considerar a qué población se va a evaluar y, además, sopesar si los medios electrónicos son idóneos para este grupo, y por qué es importante y beneficioso emplear las TAC en determinados casos.

En términos de la definición de constructo, consideramos importante para la expresión escrita, diferenciar la interacción de la expresión. Por otro lado, incluimos para la expresión oral tres constructos: la expresión, la interacción o la mediación y la conversación (Pinilla, 2010; y capítulo 3 de este libro). Eguiluz (2009) considera dentro de los criterios para evaluar esta habilidad: la adecuación, la estructura, la gramática y el vocabulario. En este sentido, es necesario dotar a los calificadores de una descripción detallada sobre estos constructos, además de los instrumentos adecuados, como una escala numérica representada por un análisis de bandas.

Sin embargo, los constructos y los criterios anteriores no deberían cambiar por el hecho de que se encuentren enmarcados en medios electrónicos. En realidad, las TAC han de representar una ventaja para asegurar que este proceso se lleve a cabo de la mejor manera. Con esto debemos resaltar que los criterios a evaluar están sujetos a nuestra definición del constructo y éste, a su vez, podría estar influenciado por el nuevo medio electrónico en que se encuadra.

Sin duda, creemos que éste es un tema que requiere de mayor investigación, puesto que, si bien en la definición de nuestro constructo debemos delimitar lo que vamos a medir, (en este caso, las habilidades oral y escrita), no es posible evitar la idea de que los medios electrónicos determinan significativamente al constructo y, por ende, adquieren un papel muy importante como facilitadores de todo el proceso que implica la aplicación de un examen.

Con esto, no queremos implicar que las TAC sean siempre la mejor opción, sino que el uso de la tecnología en función del constructo a evaluar, es una cuestión que debe someterse a un análisis profundo, pues todavía es difícil calcular el efecto que las TAC producen en el constructo, los criterios e instrumentos de evaluación. Esto, de unos pocos años para acá, se ha constituido en un campo fértil para la investigación, en el que todavía se pueden explorar nuevos caminos donde coincidan enseñanza, aprendizaje y tecnología. Debido a esto, es nuestra intención también poner sobre la mesa la cuestión de en qué medida los criterios de evaluación en las habilidades productivas son independientes del medio, sea convencional o digital, o si estos criterios deberán adaptarse de algún modo al medio que los soporta.

En este sentido, Douglas (2017: 118) plantea algunas preguntas sobre el constructo en su vínculo con la tecnología, entre las que destaca la reflexión de cómo el uso de ésta puede afectar la definición del constructo que deseamos medir. Bajo esta perspectiva, Canale (1986) añade que el uso de las tecnologías da pie a que se puedan medir diferentes constructos de lenguas, con nuevos y mejores medios tecnológicos.

A partir de ambos argumentos, podemos concluir que aún falta mucho por investigar sobre la relación del constructo y el uso de las TAC; así como también, acerca de en qué situaciones sí pueden emplearse dichas herramientas, y en cuáles es mejor optar por medios convencionales. Recordemos que cada vez se acelera más el avance tecnológico y día a día surgen nuevas necesidades que cubrir, lo que también podría acrecentar la *brecha digital* y obligaría a los candidatos a mantenerse actualizados. Es por esta razón que el constructo cobra especial importancia en la reflexión sobre su posibilidad de que se ajuste al medio; es decir, si el uso de las tecnologías afecta la naturaleza del constructo mismo, de los criterios y de nuestros instrumentos.

2.2. La validez y la confiabilidad de la evaluación de la habilidad oral y escrita en medios electrónicos

2.2.1. Validez

Definimos validez como la firmeza o seguridad que puede tener un evaluador sobre su instrumento, para determinar si hay una correlación entre el aprendizaje que se busca evaluar y el contenido de la prueba. De acuerdo con estudios previos (véase capítulo 1 de este volumen), la validez es un concepto que se asocia con la investigación del significado teórico de las puntuaciones que se obtienen a través de los instrumentos, lo cual demuestra cómo están relacionadas las respuestas observables de los examinados, con las interferencias en función de los resultados del instrumento, y el propósito con el que se utilizan (Martínez, 2014: 219).

En este sentido, creemos importante señalar que este atributo, junto con la confiabilidad y la objetividad, deben estar presentes durante todo el proceso de diseño de un instrumento, desde la definición del constructo, hasta la creación de los ítems. Por esta razón, enfatizamos que el uso de la tecnología como el medio de aplicación de nuestro instrumento de evaluación, puede facilitar algunos procesos, siempre y cuando sigamos el método adecuado para construir nuestro examen. Esto implica como primer paso la definición adecuada del constructo que deseamos evaluar y la población a la que estará dirigido nuestro examen. Una vez definido lo

que se pretende medir, será nuestra labor seleccionar cómo podrá intervenir la tecnología como el medio de aplicación de nuestro instrumento, y qué herramientas nos ayudarán a simplificar algunas tareas. Todo lo anterior, sin perder de vista que no forma parte del constructo a evaluar.

De acuerdo con Martínez (2014), hay situaciones que pueden afectar las evidencias de validez; entre ellas, hay algunas que inciden directamente en los ítems del examen, puesto que son afectados por aspectos que no pertenecen a nuestro constructo. Asimismo, instituciones como la ALTE han publicado estándares en los que establecen criterios para la creación de exámenes, aplicación y diseño de pruebas estandarizadas e información a los candidatos. En el caso del *Código de práctica* (1994), se manifiestan las obligaciones en cuatro rubros: confección de exámenes, interpretación de los resultados, esfuerzo por conseguir imparcialidad e información para los candidatos; todo esto, con la finalidad de asegurar un proceso de evaluación justo y de mejorar las prácticas evaluativas.

Si bien ya se ha hablado de métodos de validación (juicio experto, análisis psicométrico, validación de las especificaciones) y de la creación de estos estándares que nos ayudan a crear un buen instrumento, es preciso señalar también ciertos factores ajenos a una evaluación objetiva: errores en la medición por motivos externos, falta de concentración por parte del alumno, problemas técnicos relacionados con la tecnología e, incluso, con la energía eléctrica.

2.2.2. Confiabilidad

El concepto de confiabilidad se entiende como un continuo que comprende desde la consistencia de la medición, hasta la emisión de los resultados de un instrumento; es decir, para que un examen sea confiable, debe cumplir ciertos parámetros que demuestren que una persona obtendrá el mismo puntaje, siempre que se evalúe en las mismas condiciones; en otras palabras, que no intervengan otros factores que puedan alterar los resultados. Si bien en el capítulo 1 de este volumen se enuncian algunas condiciones que pueden alterar la confiabilidad en las pruebas, es preciso señalar que en los medios electrónicos también hay factores externos que impactan en la consistencia de lo que queremos medir.

A continuación, enlistaremos aquellas variables que pueden atañer en el proceso de evaluación y, de manera específica, en la validez o confiabilidad, sobre todo cuando se lleva a cabo en medios electrónicos.

- La brecha digital: cabe aclarar que este concepto se entiende como la desigualdad en el uso de las TIC entre distintos grupos sociales (Rodríguez, 2006: 28). A pesar de que este factor no afecta

directamente la validez de nuestro instrumento, creemos que es una de las variables que se deben controlar cuando elegimos evaluar mediante los medios electrónicos. Para esto, habrá que medir los alcances de nuestro instrumento y saber a qué población está dirigido; de igual manera, tendremos la obligación de informarle al candidato cuáles son los requisitos que se espera que domine para poder presentar el examen. En este caso, el uso de la tecnología y los requerimientos tecnológicos indispensables en cuanto a software y hardware. Para poder solucionar esta situación en la evaluación mediante las TAC, proponemos considerar, como algo importante, ofrecer tutoriales, MOOC o exámenes demo que permitan que el candidato se familiarice con la plataforma de exámenes, de tal forma que éste no sea un impedimento para examinarse, o una variable que pueda afectar su desempeño.

- Los sesgos: este factor repercute de manera directa en nuestro instrumento, por lo que es importante reducirlos al mínimo posible, con la finalidad de evitar inconsistencias en nuestra medición (Martínez, 2014: 241). Para ello, sugerimos, además de los referentes teóricos, definir los lineamientos que eviten la creación de ítems que favorezcan o afecten a una población.
- El entorno de aplicación: este factor tiene que ver con las condiciones ambientales, materiales y de logística que deben existir en el lugar donde se llevan a cabo las evaluaciones a través de las TAC. En los momentos actuales, pensamos en dos escenarios, ya sea un centro educativo que debe contar mínimamente con un aula tranquila, que a su vez tenga un espacio adecuado para la instalación del equipo de cómputo y la conectividad a internet que, de igual modo, deben estar en condiciones óptimas para su uso; o bien, ya sea el lugar en donde se encuentre el examinado, en aquellos casos cuando la evaluación se realice vía remota. Por supuesto que las condiciones que rodean al examinado y al equipo electrónico, a través del cual se realiza el proceso de evaluación, influyen en cierta medida: sobre estos aspectos se hablará con mayor detalle más adelante.

3. La objetividad en la evaluación oral y escrita por medio de las TAC

3.1. Objetividad

Autores como Tristán y Pedraza (2017) definen la objetividad como el resultado de un proceso que se compone de dos partes: la primera incluye el contraste de conocimientos de manera empírica; la segunda toma en cuenta un aspecto que es muy importante (y que trataremos a profundidad más adelante), denominado la intersubjetividad, que se refiere a

que determinado grupo acepta una idea como válida, por medio de un acuerdo convencional sobre determinado objeto. La objetividad en un instrumento, asimismo, debe estar presente en las preguntas (que han de elaborarse con claridad), y en los criterios de calificación.

La intersubjetividad nos obliga a la reflexión, como docentes de ELE, sobre uno de los aspectos más controvertidos: la calificación automatizada de pruebas subjetivas. En los últimos años, se ha optado por emplear sistemas automatizados de calificación, tal como lo menciona Xi (2012: 439), quien considera que los examinadores entienden los desafíos de capacitar, monitorear y contar con un gran número de evaluadores en las aplicaciones a gran escala y, por esto, aprecian los beneficios que la calificación automatizada podría ofrecer, en términos de producir puntuaciones más rápidas y fiables. Llegando a este punto, podemos afirmar que el avance de los medios electrónicos aporta grandes beneficios en las aplicaciones masivas de exámenes.

A pesar de estas ventajas, no ha sido posible conciliar si la calificación en las pruebas subjetivas (expresión oral y escrita) podría ser del todo, o en un alto grado, objetiva. Lo que sí podemos asegurar es que las TAC facilitan las tareas requeridas para tener la certeza de una calificación precisa. Esto ya ha sido regulado por organismos especializados que han propuesto una serie de estándares (SICELE y ALTE) que buscan unificar las buenas prácticas de evaluación. Es en este aspecto donde toma una gran importancia el uso de las TIC y las TAC, ya que los estándares proponen una serie de tareas, tales como hacer una doble revisión, tener las pruebas orales y escritas a disposición de los calificadores, etc.

3.2. Mantener la objetividad: un reto

Las TAC ofrecen, como se ha señalado, una variedad muy amplia de opciones y herramientas para la enseñanza, lo cual amplía el rango de modalidades en que se pueden diseñar procedimientos de evaluación. Esta diversidad implica no sólo nuevas formas de aprendizaje, sino también nuevos instrumentos, estrategias y técnicas evaluadoras, pero al mismo tiempo, dicha diversidad suele ser empleada como contraargumento para alertar de que una evaluación a través de las TAC, deja de ser objetiva. La anterior no es una objeción que se deba minimizar, sino por el contrario: es una preocupación seria y auténtica que surge del cambio de medio en el que se lleva a cabo el procedimiento de evaluación, asistido por las TAC.

Una de las principales razones de preocupación se encuentra en la posibilidad de aplicar exámenes y pruebas de forma remota, lo que crea un vacío de control en las acciones que un examinado pudiera realizar, con el fin de alterar a su favor el proceso de evaluación, al copiar o al ser sustituido por

otra persona con un mayor dominio de la lengua. Sin embargo, algunos exámenes de lengua asistidos por computadora, como el ICB TOEFL y Phone Pass, cuentan ya con candados de seguridad suficientes para evitar o disuadir a los estudiantes de hacer trampa en este aspecto.

Una segunda objeción relacionada con la reducción de objetividad es la falta de autenticidad de las pruebas. No son pocos los profesores de ELE que han señalado que por mencionar un caso, un examen oral realizado cara a cara establece una situación de comunicación, aunque no al cien por ciento, más cercana a las situaciones comunicativas auténticas, que tienen lugar en la vida cotidiana fuera del contexto escolar. No obstante lo anterior, siguiendo a Fulcher y Davidson (2012: 432), es posible argumentar que las situaciones que se plantean en cualquier examen nunca serán auténticas; por lo tanto, la discusión debe centrarse en el grado de autenticidad que podemos ofrecer adaptándonos siempre al medio, ajustando el constructo de lengua que tenemos como objetivo a evaluar, especificando claramente los contenidos y diseñando exámenes y pruebas que se adecuen a los soportes digitales.

Un ejemplo de cómo podemos lograr salvar los obstáculos de una comunicación a distancia nos lo ofrece Sawaki (2012: 432):

Muchas tareas de comprensión auditiva de la vida real incluyen información visual. Por ejemplo, en una conversación cara a cara, las expresiones faciales y los gestos de un interlocutor podrían ayudar a quien escucha a comprender mejor el mensaje de quien habla. De igual modo, cuando un estudiante escucha una clase en un salón, un profesor podría proporcionar ayudas visuales mientras presenta su tema de forma oral. El contenido de audio se puede fácilmente integrar en los exámenes de lengua asistidos por computadora (CALT, por sus siglas en inglés) a fin de incrementar su autenticidad. Como ya se ha mencionado, sin embargo, la integración de multimedia en CALT está progresando, aunque lentamente. Esto se debe a que la disponibilidad de información adicional mientras se escucha, podría cambiar la naturaleza del constructo para el que el examen ha sido diseñado.

De lo anterior, se destaca la capacidad de las TAC para el empleo de multimedia que, usada de manera adecuada, aumenta el número de recursos que pueden contribuir a incrementar la autenticidad de una prueba, acercándola a situaciones de comunicación reales. Reiteramos: ningún examen de lengua en un contexto de educación formal implica una situación auténtica de comunicación, pero existen factores que pueden ayudarnos a recrear las circunstancias o condiciones. Lo mismo ocurre en los exámenes y pruebas aplicados a través de las TAC.

Un tercer factor que juega en contra de las TAC y los medios computarizados se encuentra en cuestiones tan simples como que el monitor de una

computadora o la pantalla de un dispositivo –tableta o teléfono celular– pueden funcionar como distractores. De acuerdo con Sawaki (2012: 432), se ha realizado una investigación al respecto, cuyos resultados son intrigantes, pues arrojan datos que nos hacen pensar que la atención de los examinados puede no ser retenida por largo tiempo; que su campo de visión se reduce para evitar otros factores de distracción, o que la falta de retroalimentación audiovisual sobre la situación comunicativa afecta realmente el modo en que se comprende la información de entrada, en los exámenes realizados a través de las TAC. Es interesante, sin duda, pero no hay nada definitivo mientras no se realice otra investigación más profunda sobre este tema en específico.

4. Resultados y retroalimentación

Las herramientas digitales tienen, como una de sus características esenciales, la rapidez tanto para el procesamiento de información y datos, como para la presentación de reportes y resultados. El *software* y las inteligencias artificiales diseñadas como *e-raters* (o calificadores electrónicos) para la evaluación de expresión oral y escrita, como *Phone Pass* y *Criterion*, presentan resultados de forma inmediata, o incluso en un periodo no mayor a veinticuatro horas (Douglas, 2010: 131-132).

Exámenes como SIELE, EXELEAA y CELA, aunque ya se aplican a través de plataformas digitales, aún demoran entre uno y varios días en la entrega de resultados para las pruebas de expresión oral y escrita. Sin embargo, para las pruebas objetivas, las calificaciones se realizan de forma automatizada, lo que representa una enorme ventaja en cuanto al ahorro de tiempo y de recursos humanos y materiales.

Otro aspecto muy importante es, por supuesto, el de la retroalimentación. Los medios electrónicos permiten muchas formas de registrar retroalimentación desde diferentes fuentes –calificadores, docentes y estudiantes– y a través de distintos medios –video, audio y por escrito–. Las TAC, como hemos afirmado a lo largo de este capítulo, poseen la virtud de ofrecer varios recursos para la transferencia de información y, en lo concerniente a la enseñanza de lenguas, para proporcionar retroalimentación que ayude, tanto a maestros como a estudiantes, a reevaluar el proceso de enseñanza y aprendizaje.

Sin embargo, en el presente los seres humanos no somos la única fuente para proveer retroalimentación, sino también los *e-raters*, como en el caso de *Versant* y del *TOEFL iBT* (Douglas, 2010: 131). *Criterion*, a su vez, también proporciona retroalimentación detallada tanto a estudiantes como a profesores, y ofrece igualmente información de carácter formativo “potencialmente útil para las actividades de revisión y aprendizaje”

(Douglas, 2010: 132).

Así pues, es muy importante abordar también la pertinencia de emitir retroalimentación para los alumnos, no sólo como evaluación formativa, sino también al final de los exámenes de aprovechamiento. Nos parece muy importante que la evaluación sumativa no se perciba como algo de carácter punitivo, sino como una oportunidad para identificar fortalezas y áreas de mejora; lo cual es factible que ocurra si se le proporciona retroalimentación al estudiante, justo al final de los cursos. Las TAC pueden facilitarnos el retroalimentar a nuestros estudiantes por varios medios de comunicación. El papel del *feedback* en el mejoramiento del aprendizaje es muy relevante, y su imbricación con las nuevas tecnologías tiene un futuro prometedor que, igualmente, necesita de investigación por parte de nosotros, los docentes.

5. Conclusiones

Hemos mencionado aquí la existencia de diversos tipos de tecnologías, tales como las TIC, las TEP y las TAC, que se distinguen entre sí por el énfasis que hace cada una en aspectos como la información y la comunicación, la participación y el empoderamiento de los ciudadanos en las sociedades globalizadas, y el aprendizaje y la comunicación del conocimiento. Esta diversidad vuelve complicada una elección adecuada de los medios tecnológicos con fines educativos, pues siempre está presente caer en el error de dejarse seducir por la novedad de estos medios, aplicarlos sin una planificación cuidada ni objetivos claros, y presentar viejos métodos de enseñanza bajo deslumbrantes ropajes tecnológicos.

Así pues, cuando se habla de integrar las nuevas tecnologías al campo de la enseñanza, es indispensable tener muy en cuenta que las TAC nos proporcionan un amplio abanico de posibilidades que bien podemos aprovechar. En el caso de la enseñanza de ELE, las TAC nos permiten explorar diversos tipos de evaluación, como la coevaluación, la autoevaluación, la evaluación auténtica, la e-evaluación, etc. (Blázquez *et al.*, 2017: 27-29). En lo anterior, no se descarta incluso que el alumno comience a participar cada vez de forma más activa en su propia evaluación, lo que únicamente es posible gracias a los desarrollos en las nuevas herramientas tecnológicas.

Asimismo, las TAC potencian lo que desde décadas atrás se ha denominado como la evaluación para el aprendizaje, que ha adquirido una relevancia mayor que la evaluación del aprendizaje. Esto nos hace pensar en la disciplina de la evaluación, ya no como un control ni como un "castigo", o "etiquetar con números" el desempeño de nuestros estudiantes, sino como una oportunidad para replantear nuestros modos de evaluar el aprendizaje hacia una nueva perspectiva, la evaluación es más bien un

diálogo y negociación para el conocimiento, en lo cual adquiere mayor importancia cómo aprenden nuestros estudiantes, que lo que aprenden (Blázquez, *et. al* 2017: 29).

Por otro lado, si bien las inteligencias artificiales para la evaluación de EO y EE avanzan a pasos agigantados, y las diferencias entre los calificadores humanos y no humanos es mínima en cuanto a porcentajes de confiabilidad, tal como destaca Douglas (2010: 129-130), aún no es posible inclinar la balanza a favor de las TAC, en cuanto a aspectos como la validez y a la relación de ésta con los constructos que, sin duda, tienen que adaptarse a los nuevos medios en que se lleva a cabo la evaluación.

Las mayores dificultades que se presentan al trasladar procedimientos de evaluación de lengua, como en el caso de las habilidades de producción oral y escrita, radican en la gran variabilidad de las muestras de lengua. Y aunque es posible programar herramientas que salven estos escollos, todavía no se puede prescindir totalmente de los seres humanos y, desde nuestra perspectiva personal, no creemos que sea lo más deseable. Por el contrario, estamos convencidos de que la llegada de las TAC no implica nuestra futura extinción, sino más bien la aparición de una serie de herramientas que pueden ayudarnos a expandir nuestras capacidades y a mejorar el proceso de enseñanza-aprendizaje.

Desde este particular punto de vista, las TAC suman, no restan. Y, de igual modo, nos plantean el reto de asumir que la presencia de las nuevas tecnologías no puede ni debe soslayarse, sino que tenemos la gran oportunidad de integrarlas a la totalidad de nuestras prácticas docentes: en específico, a la evaluación del y para el aprendizaje de nuestros estudiantes siempre con miras a nuestra mejora continua.

Referencias

ALTE. *Recursos - manuales y documentos de referencia gratuitos.* (2021). Recuperado 5 de febrero de 2021, de <https://es.alte.org/Materials>

Blázquez, F., Alonso, L. y Yuste, R. (2017). *La evaluación en la era digital. Tecnología educativa.* Madrid: Síntesis.

Douglas, D. (2017). *Understanding Language Testing.* Nueva York: Routledge.

Eguiluz, J. (2009). *Criterios para la evaluación de la producción escritos.* *Marco ELE*, 9, 75-94. Recuperado 31 de enero de 2021, de https://marcoele.com/descargas/expolingua1996_eguiluz-vega.pdf

ETS Global. (2019). *Criterion. The tool to help students plan, write and revise their essays.* ETS Global. Recuperado 28 de enero de 2021, de <https://www.etsglobal.org/fr/en/test-type-family/criterion>

EST. (2021). *Explicación de sus calificaciones TOEFL iBT. ¿Qué significa su calificación? TOEFL EST.* Recuperado 28 de enero de 2021, de <https://www.ets.org/es/toefl/test-takers/ibt/scores/understanding#mybest>

Herrera, Francisco (2019). *La dimensión digital del aula de español.* En Jiménez, F. y Rufat, A. (eds.). *Manual de formación para profesores de ELE* (pp. 233-252). Madrid: SGEL.

Martínez, M. R., Hernández, M. J. y Hernández, M. V. (2014). *Psicometría.* Madrid: Alianza Editorial.

McNamara, D., Allen, L. y Crossley, S. (2019). *WAT: Writing Assessment Tool.* Companion Proceedings 9th International Conference on Learning Analytics & Knowledge (LAK19). Recuperado 28 de enero de 2021, de <http://wa.utscic.edu.au/wp-content/uploads/2019/03/McNamara-Allen-Crossley-2019-WAT-Writing-Assessment-Tool.pdf>

ORDINATE Corporation Technical Report (1999). *PhonePass Testing: Structure and Construct.* Recuperado 28 de enero de 2021, de <http://www.ordinate.com/pdf/StructureAndConstruct990826.pdf>

Pérez, S., Fluin, S., & Query, D. (2020). *Extempore* (Nº de versión 1, aplicación móvil). Deeloh Technologies. Recuperado 5 de febrero de 2021, de <https://extemporeapp.com>

Pinilla, R. (2010). *Tipos de interacción oral y entornos situacionales: una tipología de las actividades de comunicación oral.* *Marco ELE*, 10, 97-104.

Román-Mendoza, E. (2018). *Aprender a aprender en la era digital. Tecnopedagogía crítica para la enseñanza del español LE/L2.* Londres: Routledge.

Rodríguez, A. (2006). *La brecha digital y sus determinantes*. México: UNAM.

Sawaki, Yasuyo. (2012). *Technology in language testing*. Fulcher, G. y Davidson, F. (eds.). *The Routledge Handbook of Language testing* (pp. 426-437). Londres: Routledge.

Sistema Internacional de Certificación del Español como Lengua Extranjera. (2021). Recuperado 29 de enero de 2021, de Asociación SICELE website: <https://asociacionsicele.org/es/node/36>

Tristán, A. y Pedraza, N. (2017). *La Objetividad en las Pruebas Estandarizadas*. *Revista Iberoamericana De Evaluación Educativa*, 10 (1).

Versant Spanish Test. *Test Description and Validation Summary*. (2018). Pearson Education. Recuperado 28 de enero de 2021, de <https://www.pearson.com/content/dam/one-dot-com/one-dot-com/english/SupportingDocs/Versant/ValidationSummary/Versant-Spanish-Test-Description-Validation-Summary.pdf>

Xi, Xiaoming. (2012). *Technology in language testing*. Fulcher, G. y Davidson, F. (eds.). *The Routledge Handbook of Language testing* (pp. 438-452). Londres: Routledge.

Bibliografía recomendada

Alderson, J. Charles. (2006). *Diagnosing foreign language proficiency: The interface between assessment and learning*. Londres: Continuum.

Chapelle, C. y Douglas, D. (2012). *Assessing language through computer technology*. Cambridge: Cambridge University Press.

Chapelle, C. y Sauro, S. (eds.). (2017). *The Handbook of Teaching and Second Language Teaching and Learning*. Oxford: Wiley Blackwell.

Fulcher, G. y Davidson, F. (2017). *The Routledge Handbook of Language Testing*. Londres: Routledge.

Navarro, F., Ávila Reyes, N. y Gómez Vera, G. (2019). *Validez y Justicia: hacia una evaluación significativa en pruebas estandarizadas de escritura*. *Revista Meta: Avaliação*, 11 (31), 1-35.

Oliden, P. E. (2003). *Sobre la validez de los tests*. *Psicothema*, 15 (2), 315-321.

Rodríguez, G. e Ibarra, M. (2018). *E-evaluación orientada al e-aprendizaje estratégico en educación superior*. Madrid: Narcea.

A

acento. Véase *pronunciación*.

acreditación. Documento o registro otorgado por una institución que avala el resultado obtenido por un estudiante al finalizar un curso o nivel.

adecuación.

I. Cumplimiento o resolución de todos los requisitos o pautas que se piden para la ejecución de una tarea.

II. Según el contexto de la evaluación de la expresión oral o escrita, adaptación discursiva del registro empleado por un hablante.

a distancia. Modalidad educativa basada en el uso de las tecnologías digitales, en la cual el estudiante aprende de forma independiente o semi independiente, sin la necesidad de asistir a un aula física.

administración. Puesta en marcha de un instrumento hecho por una institución.

administrador. Usuario encargado del diseño, la programación y regulación de una plataforma digital y sus contenidos.

alcance. Grado de conocimiento lingüístico que un estudiante posee sobre un nivel, el cual le permite desenvolverse de forma oral o escrita en diversas situaciones.

Ambiente Virtual de Aprendizaje (AVA). Plataforma digital educativa en donde el profesor y los alumnos interactúan de forma sincrónica o asincrónica.

análisis psicométrico. Método compuesto por un conjunto de herramientas estadísticas que permiten aportar evidencia de la validez y confiabilidad de un instrumento de evaluación.

aplicación.

I. Puesta en marcha de un instrumento por parte del examinador hacia un examinado.

II. Programa de software diseñado para su utilización en una plataforma digital, en donde se puede encontrar una tarea o un instrumento de evaluación.

aprendizaje en línea. Proceso educativo que se realiza en un entorno digital, con o sin un profesor.

aprendizaje mixto/híbrido (*blended learning*). Modalidad de estudio que combina el aprendizaje en línea y las clases presenciales.

aprendizaje móvil (*mobile learning*). Forma de aprendizaje en la que los contenidos se pueden consultar mediante dispositivos móviles con acceso a internet, en cualquier lugar y a cualquier hora.

aprovechamiento. Conocimiento y habilidad que se adquirió por las y los estudiantes durante el proceso de enseñanza-aprendizaje de un curso o nivel determinado.

B

banco de pruebas o tareas. Conjunto de reactivos que se almacenan para su posterior utilización en diferentes test de evaluación.

banda. Eje vertical de una escala que designa un valor numérico para calificar la actuación de un alumno.

Blended Learning. Véase *aprendizaje mixto/híbrido*.

C

calibración. Disminución del error de medida mediante el ajuste de los reactivos de un instrumento de evaluación.

calificación. Expresión numérica que el evaluador otorga para saber si el estudiante puede ser promovido o no de nivel o curso.

calificador. Persona que puntúa una tarea o prueba.

calificador electrónico. Software con inteligencia artificial capaz de procesar, evaluar y retroalimentar muestras escritas y orales de lenguas humanas.

calificador experto. Profesor de ELE que ha sido estandarizado y acreditado para calificar tareas de expresión dentro de una institución.

CALT (*Computer Assisted Language Test*). Véase *examen de lengua asistido por computadora*.

candado de seguridad. Certificado digital de una página web que confirma su origen y autenticidad.

CBT (*Computer Based Test*). Véase *examen basado en computadora*.

certificación. Respaldo institucional que avala el dominio de las competencias lingüísticas que posee un hablante.

código. Combinación reglamentada de un conjunto de signos que permiten la transmisión de un mensaje.

cognitivismo. Teoría psicológica que estudia los procesos mentales superiores, relacionados con la asimilación y organización de los conocimientos.

coherencia. Claridad en la exposición de las ideas para la comprensión de un mensaje.

cohesión. Concordancia entre cada una de las ideas semánticas, sintácticas y pragmáticas del mensaje.

confiabilidad. Cualidad de un instrumento de evaluación que determina el grado de estabilidad y consistencia de los resultados obtenidos, aun cuando este instrumento haya sido aplicado al mismo individuo en distintas ocasiones.

confidencialidad. Protección de los datos personales de un individuo para que no sean divulgados sin su consentimiento.

constructo. Concepto o característica que se pretende medir con un instrumento diseñado.

contenido mínimo. Conocimiento indispensable que el alumno de un determinado nivel debe poseer para desempeñarse apropiadamente en el siguiente.

contexto. Serie de circunstancias que modifican el intercambio comunicativo de los hablantes.

conversación. Actividad espontánea de interacción comunicativa libre que se caracteriza por la alternancia de turnos entre dos o más interlocutores en relación simétrica.

corrección. Categoría asociada a la competencia gramatical y a las competencias lingüísticas en la que se evalúa el uso de la lengua con apego a la norma de la lengua meta; por ejemplo: la ausencia de errores en las oraciones.

criterios de evaluación. Descripción de la actuación que se espera del estudiante frente a una tarea determinada.

cumplimiento. Véase adecuación.

Curso Online Masivo Abierto (COMA). Véase MOOC.

D

descriptores. Explicación detallada de lo que hace el alumno y el nivel de concreción con respecto a la tarea que se evalúa.

dificultad (de criterio). Puntaje promedio obtenido por todos los examinados en relación al puntaje máximo posible.

discriminación. Factor que compite con la respuesta correcta.

doble corrección. Resultado que se obtiene al promediar dos calificaciones hechas por un par de expertos sobre una misma prueba.

E

e-evaluación. Véase *evaluación digital*.

entonación. Variación de la pronunciación, el tono y el volumen de la voz que otorga una carga semántica diferente a una palabra en un enunciado.

e-rater. Véase *calificador electrónico*.

e-rúbrica. Herramienta digital que auxilia al docente en la elaboración e implementación de una rúbrica para la evaluación dentro de un ambiente virtual.

error de medida. Diferencia de medición que resulta de la comparación entre el valor verdadero y el valor obtenido en un instrumento de evaluación.

escala. Conjunto de elementos que permiten valorar los conocimientos de un estudiante en un nivel determinado.

escala analítica. Instrumento de evaluación que desglosa categorías específicas sobre lo que hace el estudiante con respecto a una tarea de expresión.

escala holística. Instrumento de evaluación que ofrece una descripción rápida y razonablemente sencilla de lo que hace un estudiante con respecto a una tarea de expresión.

estándar. Modelo fundamentado en principios psicométricos que sirve para diseñar, elaborar, aplicar o valorar un instrumento de evaluación y los efectos de su uso.

estandarización.

I. Proceso mediante el cual se conviene y adecua el proceso de creación y aplicación de un modelo de instrumento a una norma establecida.

II. Formación que los expertos en evaluación deben seguir para calificar

de forma justa e imparcial los contenidos de una prueba.

estrategias cualitativas. Técnica de investigación centrada en la obtención de datos (no numéricos) derivados de la observación y descripción del objeto de estudio.

estrategias cuantitativas. Técnica de investigación que recopila sistemáticamente datos y los analiza para la comprobación de hipótesis.

estructura. Organización del texto, empleo de marcadores discursivos y mecanismos de cohesión.

evaluación. Proceso mediante el cual se obtiene información, se emiten juicios de valor y se toman decisiones al respecto.

evaluación del aprendizaje. Valoración vinculada al aprovechamiento de un curso, así como a la evaluación certificativa.

evaluación diagnóstica. Valoración que se centra en el tipo y nivel de conocimientos que tienen los alumnos antes de iniciar un curso o asignatura. A partir de esta evaluación, el docente puede tomar decisiones objetivas respecto a su actuación pedagógica.

evaluación digital. Integración de los recursos informáticos para crear y aplicar un instrumento de evaluación dentro de un ambiente virtual.

evaluación formativa. Recopilación de información a lo largo de todo el proceso de enseñanza y aprendizaje que permite la regulación y el mejoramiento de las condiciones pedagógicas.

evaluación sumativa. Valoración que el docente realiza al término de un curso o nivel, con el propósito de determinar si los estudiantes alcanzaron o no los objetivos propuestos.

evaluador. Véase calificador.

evidencia. Verificación existente entre la relación de la prueba con el resultado obtenido por el estudiante.

examen. Instrumento objetivo, válido y confiable que mide las características y dominios educativos de un individuo.

examen a distancia o virtualmente. Forma de evaluación en la cual el alumno realiza el examen de forma remota a través de una plataforma.

examen basado en computadora. Prueba diseñada y aplicada en un entorno computarizado.

examen computarizado. Prueba que se realiza en una computadora dentro de las instalaciones de una institución educativa.

examen de aprovechamiento. Test realizado al final del curso que mide el cumplimiento de los objetivos.

examen de lengua asistido por computadora. Prueba evaluada mediante un sistema computarizado que analiza las respuestas de forma automática, siguiendo tres principios: la eficiencia de aplicación de la prueba, la estandarización de resultados y la innovación tecnológica.

examinador. Véase *calificador*.

experimentación. Véase *pilotaje*.

expresión escrita. Producción textual realizada por un usuario de la lengua, la cual es hecha para diferentes fines.

expresión oral. Producción del discurso en la que una persona emite un mensaje para sí misma, para otra o para varios individuos, dotada de elementos paralingüísticos, socioculturales y pragmáticos.

F

feedback. Véase *retroalimentación*.

fiabilidad. Véase *confiabilidad*.

fluidez (oralidad). Calidad que indica la velocidad del habla del locutor, cuya soltura es similar a la de un hablante nativo.

función pragmática. Uso de la lengua que establece una relación entre el enunciado, el contexto y los interlocutores.

G

gramática. Conjunto de normas lingüísticas que determinan la manera en que las palabras se relacionan dentro de una oración.

H

herramientas digitales asíncronas. Sistema de comunicación virtual empleado para interactuar sin necesidad de que el o los usuarios estén conectados en el mismo espacio y tiempo.

herramientas digitales síncronas. Sistema de comunicación virtual que permite la interacción en tiempo real con el o los participantes.

I

input. Muestra oral o escrita de la lengua meta a la que el aprendiente está expuesto.

instrucciones. Serie de pasos que un estudiante sigue para la resolución de una tarea o una prueba, la cual debe ser precisa sin otorgar poca o demasiada información al estudiante.

instrumento de certificación. Prueba mediante la cual una institución de prestigio acredita el dominio que un estudiante posee de una lengua.

instrumento de evaluación. Herramienta diseñada a partir del currículo de un curso que permite obtener la información de los resultados sobre un proceso de enseñanza o aprendizaje.

intención. Propósito comunicativo que un participante transmite por medio del discurso oral o escrito y que puede alterar aquello que se diga en un contexto determinado.

interacción. Actividad comunicativa en la que existe un intercambio oral o escrito entre dos o más participantes que resuelven una tarea en común.

interacción oral. Intercambio de información o conocimientos entre dos o más hablantes enmarcados en una situación comunicativa particular con expectativas y patrones de comportamiento, en relación asimétrica.

interlengua. Sistema lingüístico que el estudiante de una segunda lengua va construyendo durante su proceso de aprendizaje, independiente de la versión final de una L1 o una L2.

interlocutor. Participante que puede intervenir en una conversación.

ítem. Ejercicio, problema o pregunta a resolver para medir las habilidades y conocimientos del examinado.

J

juicio de experto. Dictamen emitido por un calificador estandarizado sobre el instrumento y los resultados obtenidos de una prueba.

justicia. Principio de imparcialidad con el que un calificador experto busca garantizar una puntuación clara, equitativa y válida para todos los estudiantes.

L

léxico. Unidad semántica compositiva de una lengua.

lineamiento. Determinación de los propósitos que se pretenden obtener mediante la elaboración de un instrumento de evaluación.

M

marcador del discurso. Expresión lingüística que codifica instrucciones de procesamiento y vincula segmentos discursivos para guiar la interpretación de enunciados.

medio computarizado. Instrumento de evaluación montado en una computadora con el propósito de almacenarlo para acceder a él posteriormente.

medio electrónico. Herramienta de comunicación utilizada durante el proceso de enseñanza y de aprendizaje, cuyo soporte está en la red. Se puede acceder a ella por medio de diversos canales como computadoras, teléfonos inteligentes, tabletas, etc.

método. Secuencia de procedimientos basados en un enfoque que servirá para determinar la forma en la cual se enseñará una lengua.

MIS (Management Information System). Véase *sistema de información de gestión*.

Mobile Learning. Véase *aprendizaje móvil*.

modelo de examen. Muestra preliminar de un test correspondiente a un nivel específico.

monitorización. Revisión periódica a los servidores donde se alojan los sistemas informáticos para supervisar el uso del CPU, el consumo de memoria, las entradas y salidas, la red, el espacio en disco, el procesador, el número de peticiones o saturación de conexiones, etc.

MOOC (Massive Open Online Course). Curso en línea (generalmente gratuito) de corta duración disponible para cualquier persona en cualquier parte del planeta.

muestra.

I.(Oral/escrita) Recopilación específica de datos producida por el estudiante que se utiliza para su posterior análisis y evaluación.

II. (En psicometría) Subconjunto de personas representativas de un sector poblacional, tomado para efectos de una investigación.

N

nociones

I. Conceptos elementales que una persona posee sobre una disciplina o asignatura específica

II. Conceptos elementales basados en el análisis de la lengua y que identifican una serie de categorías lingüísticas semántico-gramaticales.

O

objetividad. Ausencia de las percepciones individuales de un examinador frente a un objeto de estudio.

objetivo (del examen). Descripción de aquello que se pretende que el alumno realice durante la prueba. Su creación parte de los objetivos de aprendizaje.

oralidad. Emisión de ideas o conceptos mediante la voz.

P

parámetro. Conjunto de características que delimitan el constructo de una evaluación.

pautas. Guía estructurada en puntos a resolver de las tareas de expresión.

pilotaje. Puesta en marcha del instrumento de evaluación para estimar los tiempos de ejecución y los resultados.

plataforma digital (o virtual) . Sistema en el que ejecutan aplicaciones, servicios o herramientas bajo un mismo entorno, dando a los usuarios y administradores, la posibilidad de acceder a ella a través de Internet desde cualquier lugar y en todo momento. Generalmente, se emplean para la educación a distancia e intentan simular las mismas condiciones de aprendizaje que se registran en un aula.

ponderación. Importancia que asigna una institución a un criterio de evaluación con respecto a una calificación.

pregunta. Véase *ítem*.

pronunciación. Articulación de los sonidos que propicia el entendimiento de las palabras.

prueba. Véase *examen*.

prueba objetiva. Instrumento de evaluación constituido por ítems con una respuesta única.

prueba subjetiva. Instrumento de evaluación en el cual un examinador puntúa la actuación de un usuario.

psicometría. Disciplina encargada de la medición y cuantificación del comportamiento humano, incluyendo la teoría de los test y el análisis de los ítems o reactivos que componen las pruebas.

puntaje. Sumatoria total de puntos obtenidos en una prueba.

punto de corte. Cantidad mínima de puntos que el estudiante necesita superar en un instrumento de acuerdo con su desempeño para determinar su nivel de logro.

R

reactivo. Véase *ítem*.

registro. Variación lingüística de una lengua que se modifica dependiendo de la situación comunicativa en la que el hablante se encuentre.

remoto. Véase *A distancia*.

resolución de la tarea. Acto mediante el cual el alumno responde pautas con las que debe solucionar una serie de reactivos.

retroalimentación. Juicio emitido por un docente con el que se pretende alentar el progreso o señalar cómo puede mejorar un estudiante su proceso de aprendizaje.

rubro. Aspecto parcial que se considera en las rúbricas, por ejemplo, en la expresión oral: gramática, léxico, entonación, pronunciación, fluidez, cohesión, coherencia y resolución de la tarea; en el caso de la expresión escrita: cohesión, coherencia, léxico, gramática, ortografía y resolución de la tarea.

rúbrica. Véase *escala*.

S

sintaxis. Ordenamiento lógico de las palabras regido por sus funciones gramaticales.

sistema de aplicación (de exámenes). Sistema informático que permite a sus usuarios realizar tareas determinadas. Algunos ejemplos de estos sistemas son: los procesadores de texto, las hojas de cálculo, los reproductores de audio y vídeo, etc.

sistema de información de gestión. Sistema informático que recopila y administra datos de múltiples niveles mediante una interacción colaborativa, con la finalidad de generar diagnósticos y retroalimentaciones que permitan evaluar de forma constante el mismo sistema u otros para la toma de decisiones.

subjetividad en la evaluación (oral/escrita). Juicio basado en la percepción del evaluador respecto a la producción oral o escrita del estudiante.

T

tabla de especificaciones. Herramienta de organización lógica y secuenciada de los componentes que integrarán el instrumento de evaluación.

TAC (Tecnología del Aprendizaje y el Conocimiento). Véase *tecnología educativa*.

tarea. Actividad que el estudiante debe realizar para demostrar los conocimientos y habilidades definidos previamente en los objetivos de aprendizaje o de evaluación.

tecnología educativa. Recursos técnicos utilizados en el proceso de enseñanza, como plataformas digitales con materiales didácticos, lúdicos, blogs, vlogs, foros, etc.

test. Véase *examen*.

texto. Producción de ideas gramaticales y sintácticamente estructuradas en un documento de acuerdo con un tema determinado.

texto de entrada. Introducción que se pone antes de resolver una tarea.

texto de input. Véase *texto de entrada*.

TIC (Tecnologías de la Información y la Comunicación). Soporte digital en el que se basan los recursos tecnológicos para difundir información.

V

validación (de las pruebas). Proceso que analiza cómo se han elaborado la prueba y sus contenidos.

validez. Grado de adecuación que un instrumento posee para cuantificar y evaluar el constructo que se pretende medir.

validez de constructo. Correlación existente entre el instrumento diseñado, los fundamentos teóricos y las prácticas de clase de una institución educativa.

validez de contenido. Juicio derivado de la relación entre el constructo a evaluar y la forma del instrumento de evaluación.

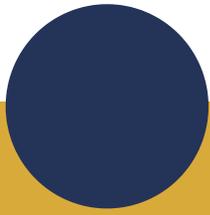
validez empírica. Comprobación práctica del desempeño de una prueba.

versión de examen. Véase *modelo de examen*.

vocabulario. Véase *léxico*.

**La evaluación de EO
y EE en ELE: retos y
propuestas** se editó en la
Ciudad de México durante el
primer semestre de 2021, a manera
de publicación digital, con formato de
salida PDF. Se utilizó la fuente tipográfica
Avenir Roman para el texto, y Bebas Neue
Pro para los títulos. Se hizo la correc-
ción de estilo, orgullosamente, en
Linux. Para la maquetación y
formación se usó Adobe
InDesign 2021.





Este libro es uno de los productos del proyecto PAPIME Nuevos procedimientos para la evaluación de las habilidades de expresión escrita y expresión oral en estudiantes de español como lengua extranjera en el Centro de Enseñanza para Extranjeros y tiene como propósito divulgar la investigación que se realizó en el Departamento de Evaluación del CEPE como antecedente teórico para el desarrollo de la Plataforma integral de exámenes y evaluación (Pliexeval) que aloja propuestas de exámenes de aprovechamiento de expresión oral y escrita, instrumentos de evaluación de dichas habilidades, y herramientas para la creación, gestión y aplicación de exámenes.

El hilo conductor de los capítulos que integran el volumen es la evaluación de las habilidades de expresión, tanto oral como escrita, en los exámenes de aprovechamiento de ELE. Nuestros objetivos específicos son introducir a los profesores de lengua en los conceptos fundamentales de la psicometría, ciencia de gran importancia para la disciplina de la evaluación; proporcionar herramientas teórico-prácticas para elaborar y reflexionar sobre los instrumentos de evaluación de expresión oral y escrita para sus clases; y presentar cómo las herramientas tecnológicas se han ido integrando en la evaluación de ELE.

