

REVISTA DE EVALUACIÓN
E INVESTIGACIÓN EDUCATIVA

Número 3 | 2021

R3



CENEVAL®



@SomosCeneval

INNOVACIONES EN LA **NUEVA GENERACIÓN** **DE EXÁMENES** DEL CENEVAL II

CENEVAL[®]

Revista de evaluación e investigación educativa

Ceneval Investiga

Presentación

El Centro Nacional de Evaluación para la Educación Superior (Ceneval) presenta a la comunidad de académicos e investigadores en educación, evaluación y al público en general el tercer número de su nueva revista digital especializada en temas de evaluación e investigación educativa. *Ceneval Investiga*, cuya periodicidad es cuatrimestral, busca constituirse en un foro para mostrar investigaciones y ensayos realizados por personal del Centro y de otras instituciones y organismos interesados en dichos temas.

En esta ocasión se continúa con la exposición de algunas de las innovaciones más relevantes de nuestra nueva generación de pruebas, que se ha puesto en operación este año. En julio se estrenaron los nuevos Exámenes Nacionales de Ingreso a la Educación Media Superior (EXANI-I), a la Educación Superior (EXANI-II) y al Posgrado (EXANI-III); a partir de noviembre se aplicará el examen diagnóstico al concluir el componente de formación básica establecido en el Marco Curricular Común de la Educación Media Superior (DOMINA-BACH Básico) y, en diciembre, los nuevos Exámenes Generales para el Egreso de la Licenciatura (EGEL Plus).

En primer lugar, el rector del Sistema de Universidades Estatales de Oaxaca (SUNEO), doctor Modesto Seara Vázquez, describe brevemente el modelo de estas universidades, así como su experiencia en la evaluación de los aprendizajes de estudiantes. Se explica asimismo cómo se diseñaron el nuevo DOMINA-BACH Básico y el EGEL Plus Disciplinar del Proyecto Egresos de Licenciatura, se describe el Examen de Expresión Escrita en Español (EXPREESE) y se describe la justificación de dos de las principales innovaciones de esta generación de exámenes: los detalles técnicos para el establecimiento del Modelo Único para la Evaluación de las Habilidades Socioemocionales y la incorporación de reactivos de innovación en los EGEL Plus.

Los académicos e investigadores interesados en los temas de esta revista pueden enviar sus colaboraciones a revista.investigacion@ceneval.edu.mx para que sean evaluadas e incluidas en números subsecuentes. El propósito es que *Ceneval Investiga* se constituya en un instrumento más que contribuya con nuestro mandato social: mejorar la calidad educativa de nuestro país.

Contenido

Entrevista
al rector Modesto Seara Vázquez

06

El nuevo DOMINA-BACH Básico

18

El EGEL Plus
Disciplinar del Proyecto Egresado de Licenciatura

26

EXPRESE Examen de Expresión Escrita en Español
Evaluación de las habilidades para comunicarse de forma escrita

34

Hacia un modelo único para la medición
de habilidades socioemocionales en el Ceneval

42

Los reactivos de innovación

52



Entrevista al rector Modesto Seara Vázquez

El doctor Modesto Seara Vázquez, rector del Sistema de Universidades Estatales de Oaxaca (Suneo), es un catedrático reconocido en el ámbito internacional. Es investigador nacional emérito del Sistema Nacional de Investigadores, profesor titular C de tiempo completo de Derecho Internacional y Relaciones Internacionales (UNAM) y presidente de honor de la Asociación Mexicana de Estudios Internacionales.

De acuerdo con lo publicado en su biografía (<http://www.modestoseara.com/index.html#>), 1988 fue un año decisivo en su vida: un antiguo alumno suyo de la Facultad de Derecho de la UNAM, Heladio Ramírez López, entonces gobernador de Oaxaca, le pidió un proyecto para una universidad en la ciudad de Huajuapán de León, en la región mixteca oaxaqueña, ubicada al norte del estado, en una zona alta, deteriorada, sin vegetación y de difícil acceso. El proyecto

consistía en construir una universidad que respondiera a las necesidades locales.

Con este proyecto se dio inicio a un nuevo modelo, concebido como una universidad para el desarrollo. La definición que le dio a la universidad buscaba trascender su función educativa y entenderla como un instrumento cultural para transformar la sociedad, mediante la enseñanza, investigación, difusión de la cultura y promoción del desarrollo. La Universidad Tecnológica de la Mixteca inició sus labores en 1990, con 48 alumnos, cinco profesores y dos aulas, en medio de un terreno bautizado por los humoristas locales como “el desierto de Seara”. Fue el inicio de una gran aventura que llevó a un proyecto académico universitario sin precedentes. En 2016, la UTM tenía alrededor de 220 profesores-investigadores de tiempo completo, 1 500 estudiantes,

diez licenciaturas e ingenierías, 15 maestrías, cinco doctorados y ocho institutos de investigación, en 208 hectáreas con más de 100 edificios. Pero lo más importante es la calidad: sus alumnos rutinariamente consiguen los primeros lugares nacionales en las pruebas que hace el Ceneval de ingeniería en computación, electrónica, industrial o ciencias empresariales. En 2008 y 2011, sus estudiantes consiguieron también los primeros lugares en pruebas internacionales.

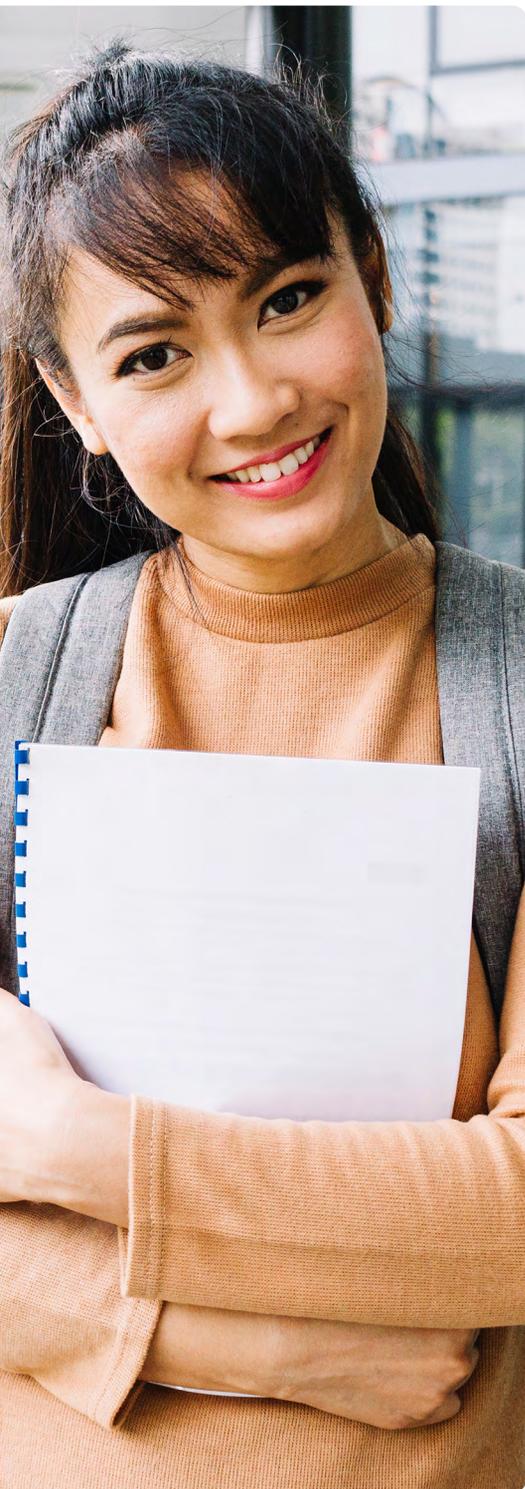
El Suneo cuenta ya con diez universidades (Universidad Tecnológica de la Mixteca, Universidad del Mar, Universidad del Istmo, Universidad del Papaloapan, Universidad de la Cañada, Universidad de la Sierra Juárez, Universidad de la Sierra Sur, Universidad de la Costa, Universidad de Chalcatongo y NovaUniversitas), con 17 campus que cubren la totalidad del territorio del estado de Oaxaca.

En una visita reciente (17 y 18 de junio) a tres campus del Sistema de Universidades Estatales de Oaxaca (SUNEO), el Ceneval tuvo la oportunidad de platicar ampliamente con el Dr. Modesto Seara Vázquez, rector general de este sistema educativo. Aquí se presenta la entrevista realizada en el marco de dicha visita.

¿Qué es para usted una institución de educación superior de buena calidad?

En una institución que se limita a la educación superior, en sus tres niveles básicos, el factor fundamental que condiciona o determina su calidad es la calidad del profesorado, que es el factor más importante de todos. Pero también hay factores complementarios, como la calidad de las instalaciones y los servicios, los equipos de todo tipo, las bibliotecas y redes informáticas, el ambiente (tipo campus o tipo urbano, de las que yo llamo universidades-cemento) y el modelo académico.





Para mí, una verdadera universidad no es una simple escuela, por buena que sea en esa función. Yo defino a la universidad como un instrumento cultural para transformar y desarrollar la sociedad. Por ello, debe asumir muchas funciones, entre las cuales están la enseñanza, la investigación, la difusión de la cultura y la promoción del desarrollo, además de ofrecer una serie de servicios a la sociedad. También debe asumir su responsabilidad como transmisora de valores sociales positivos para formar ciudadanos responsables.

La universidad no puede ser simplemente capacitadora de profesionales, por buenos que éstos sean en su campo de especialidad.

¿Podría describir brevemente el modelo de las universidades del SUNEQ?

Partiendo de la definición que ofrecí de universidad como instrumento cultural de transformación y desarrollo de la sociedad, elaboramos el modelo que toma en cuenta la realidad oaxaqueña. No consideramos conveniente duplicar la oferta educativa de la capital y nos orientamos al interior del estado, en donde se encuentran los jóvenes cuya situación socioeconómica es de las más deterioradas del país. Para superar esas condiciones diseñamos un modelo orientado a conseguir la más alta calidad, y ello requiere la creación de un sistema intensivo, de tiempo completo y dedicación exclusiva que, además, trate de corregir las deficiencias formativas que traen consigo los alumnos.

Entre los rasgos que se pueden destacar están los cursos propedéuticos para seleccionar a los alumnos que reúnen el mínimo de condiciones y habituarlos a la disciplina de este modelo universitario, así como el reforzamiento del conocimiento de las materias básicas en los diversos campos del conocimiento; el estudio de materias formativas como Historia del

Pensamiento Filosófico, para explicar cómo la humanidad se enfrentó al problema del conocimiento a lo largo de la historia, y Teoría General de Sistemas, para desarrollar la capacidad de abstracción y de ordenación de la realidad. También se hizo obligatoria la lectura de una novela al mes y la presentación de un comentario escrito de cuatro páginas para mejorar la capacidad de comunicación e inculcar el hábito de la lectura.

En cada plan de estudios se incluye al menos una materia de administración de empresas para que los alumnos sepan cómo manejar una empresa y no se limiten únicamente a la búsqueda de un empleo; estudio del idioma inglés, en el que se establece como norma general de calidad en su enseñanza, que los profesores sean nativos del idioma que enseñan, lo cual también funciona para los idiomas optativos (chino mandarín, francés y alemán); estancias profesionales en instituciones o empresas afines a su carrera, con una duración de 2 meses cada verano, al concluir el 6° y el 8° semestres; titulación optativa de tesis profesional o Examen General de Conocimientos con el Ceneval.

La mayoría de nuestros campus se ubican en bosques cercanos a las ciudades; dentro de ellos se aplican normas estrictas de mantenimiento y limpieza. Cuentan con los servicios de cafetería, enfermería, auditorio, biblioteca, redes de cómputo, laboratorios y talleres, planta de tratamiento de aguas residuales, vivienda para un número limitado de profesores, despachos individuales para profesores, espacios deportivos, etcétera.

Asimismo, se está extendiendo la producción de energía solar; la Universidad Tecnológica de la Mixteca es prácticamente autosuficiente y la Universidad del Istmo ha avanzado mucho en ese camino. Además, se inició la puesta en servicio de depuradoras de agua para que todos los universitarios puedan disponer del líquido de forma gratuita, lo que evita el consumo de refrescos o aguas de sabores con edulcorantes. La primera se concluyó en la Universidad de la Sierra Sur y el mismo proyecto se va a replicar en el resto de los campus.





¿Cómo evalúa el SUNEО el logro de resultados de aprendizaje de sus alumnos en los programas educativos?

En un sistema de trabajo intenso como el que tenemos debe haber procedimientos de evaluación frecuentes para evitar que los alumnos se limiten a asistir a las clases y dejar el estudio para el periodo de exámenes finales, como sucede en otros modelos de universidad. Más o menos cada tres semanas se realizan exámenes parciales, lo que los obliga a estudiar desde el principio. Complementariamente, los profesores evalúan a los alumnos tomando en cuenta los trabajos de clase, las participaciones en clase y las prácticas. Los exámenes finales tienen un mayor peso. La asistencia a clases es obligatoria y se requiere un mínimo de 85% de asistencia.

El idioma inglés es un requisito, pero no forma parte del plan de estudios, debido a que los alumnos llegan con niveles de conocimiento diferente (se les clasifica en el momento de su ingreso) y también avanzan a un ritmo diferente. Para titularse requieren aprobar el examen TOEFL.

El control de lecturas se realiza cada mes y el alumno debe cumplir con dicho requisito para poder presentarse a exámenes.

Los EGEL del Ceneval son el principal método que tenemos para medir nuestro nivel de competitividad al final de la carrera, dado que miden y comparan resultados. Particularmente, nos permiten comparar la calidad entre universidades en su conjunto y no se limitan a comparar el nivel de conocimiento entre un número restringido de alumnos.

De modo indirecto evaluamos nuestra calidad en función de la empleabilidad de los estudiantes, en los planos nacional e internacional. No nos referimos al nivel administrativo de

los empleos, puesto que el origen socioeconómico de nuestros alumnos los coloca en desventaja al no formar parte de las élites sociales y económicas. Los contratan por su capacidad profesional y por los valores positivos que han asimilado en los años universitarios.

¿Cuáles han sido los beneficios de aplicar los exámenes EGEL en sus universidades?

La calidad de la enseñanza debe medirse esencialmente por los resultados y esa es la función de los EGEL. La comparación de los niveles de calidad entre individuos sólo permite medir y comparar la calidad de la enseñanza, pero los EGEL nos permiten comparar generaciones completas, y eso es lo que refleja el valor académico relativo de las universidades. Las calificaciones de excelencia para individuos no tienen el mismo interés para nosotros porque, aunque miden el nivel académico de éstos, que puede ser muy bueno, no reflejan el del conjunto completo.

Para nosotros, los EGEL y los numerosos premios recibidos en los ámbitos nacional e internacional han sido fundamentales para probar nuestra calidad académica, dado que no estamos en condiciones de gastar en difusión lo que invierten otras universidades.

Es así como se ha dado a conocer que las Universidades Estatales de Oaxaca están entre las mejores en Computación, Mecatrónica, Empresariales, Enfermería, Administración Turística, Relaciones Internacionales, etcétera. Como es natural, los resultados varían cada año, pero se sostiene un nivel de calidad.





¿Cómo es la inserción laboral de los egresados de las instituciones del SUNEEO?

Al principio fue difícil, pues los empleadores preferían a jóvenes “con mejor presentación” (traje y corbata) y procedentes de universidades “con prestigio” (léase privadas). Posteriormente, cuando las universidades empezaron a ser más conocidas, esos empleadores se quejaron de que nuestros alumnos eran “muy presumidos” y preferían irse a otros lados. Al preguntarles cuánto les ofrecían como salario me respondían que lo que era normal (según ellos). Les dije que con esos salarios no podían esperar otra cosa.

En términos generales, nuestros alumnos no tienen problema para encontrar empleo. Unos prefieren quedarse en su tierra, sobre todo muchos de la carrera de Enfermería, quienes dicen que primero tienen que cumplir con sus comunidades. Otros son muy demandados fuera del país, particularmente los de Enfermería y los de las carreras relacionadas con la computación.

También hay un número creciente de egresados en la administración federal del estado de Oaxaca.

¿Cómo promueven las universidades del SUNEEO la movilidad social?

Ésta es una función esencial de las Universidades Estatales de Oaxaca y su eficacia está determinada por la calidad profesional. Nos aseguramos de que nuestros alumnos reciban una capacitación profesional de alta calidad y una sólida formación, con valores positivos, que los vuelven muy competitivos y les aseguran su inserción en la sociedad. Esto, además de resolver su problema personal, los convierte (lo que es más importante para nosotros) en factores eficaces de transformación social.

En la medida en que nuestras universidades actúan como factores de movilidad individual y de transformación social, han asegurado la estabilidad de la sociedad, que no puede ser estática, sino dinámica.

Estos papeles no se aseguran sólo con la función de enseñanza, sino que se refuerzan con otras funciones como la investigación científica, la difusión cultural, la promoción del desarrollo y la prestación de servicios a las comunidades.

¿Qué opina de los *rankings* internacionales?

Para decirlo claramente, es una farsa gigantesca, una estrategia de mercadotecnia para algunas universidades y un buen negocio para las empresas que los elaboran, incluyendo periódicos nacionales y extranjeros.

Las empresas dedicadas a elaborar *rankings* han proliferado por todo el mundo y cada una utiliza diversos criterios, que son limitados y diferentes, y llevan a resultados muy distintos, aunque a veces coincidan en algunas universidades. Basta leer la lista de supuestas universidades de excelencia, tanto nacionales como internacionales, para darse cuenta de la falta de seriedad de tales clasificaciones.

En una parte de esas clasificaciones, para ser incluidos, se requiere hacer una inversión que no todas las universidades consideran conveniente. Esto es particularmente cierto para algunos diarios mexicanos, de los que hemos recibido las ofertas de espacios publicitarios y la información de los costos correspondientes.

Rara vez se evalúa la calidad académica terminal y cuando se habla de resultados en realidad se hace referencia a los niveles altos que los egresados de algunas instituciones alcanzan en las empresas, que generalmente son jóvenes procedentes de





sectores sociales privilegiados que nacen “con la mesa puesta”. Si de los puestos administrativos se pasa a los de carácter técnico, donde lo que vale es la calidad académica, el asunto es diferente.

En los medios internacionales se han denunciado las prácticas corruptas de universidades que falsifican datos, contratan profesores de prestigio temporalmente, para los periodos de evaluación, y presentan como profesores de planta a simples conferencistas con una tenue relación con las universidades en cuestión.

Frecuentemente se olvida que las universidades son instituciones vivas, que no pueden mantener una calidad uniforme porque ésta depende de ciertas variables, como la calidad de los profesores, que cambia con el tiempo (para bien o para mal), la estabilidad de la propia institución, que puede atravesar periodos de agitación que afectan a su funcionamiento (rectores incompetentes o corruptos, organizaciones sindicales agresivas, movimientos estudiantiles espurios, politización que convierte la cátedra en tribuna, interferencia de gobiernos no democráticos, incrustación de grupos parasitarios que se reparten el presupuesto, y *last, but not least*, lo que he definido como “mediocratización por sedimentación”, en la cual las sucesivas direcciones van dejando capas de burócratas o académicos mediocres que se perpetúan en sus puestos).

La terminología utilizada respecto a las universidades también induce a confusión. Así, a veces se habla de grandes universidades para referirse en realidad a universidades grandes, que no es lo mismo; universidades importantes, que pueden no ser las de mejor calidad, pero cuya presencia en un país las convierte en fundamentales, mientras que universidades de más calidad, pero pequeñas, no tienen tanta trascendencia. También podría hablar de las universidades de “vamos a...”, que se la pasan anunciando grandes proyectos científicos, pero no llegan nunca a universidades de “hemos...”, que presenten

resultados. En fin, podríamos enriquecer el vocabulario “calificatorio” de las universidades con el término de “universidades zombis”, para referirnos a instituciones que hace tiempo perdieron su relevancia y pretenden vivir de glorias pasadas.

¿Cómo debería ser un sistema de evaluación para la educación superior en México?

Para nosotros (SUNEO), el Ceneval es muy importante para la dimensión educativa, en particular con los EGEL, porque además de servirnos para demostrar la calidad de nuestras universidades, también nos sirve para analizar las fallas que podemos tener en las diversas áreas del conocimiento y buscar las correcciones necesarias.

Finalmente, se debe recordar que los procedimientos de medición de la calidad universitaria vigentes abarcan sólo la dimensión educativa y una verdadera universidad es mucho más que eso.

¿Cuáles son los retos actuales para las universidades del SUNEO?

Son múltiples los retos que enfrentamos, pues se trata de un sistema que, tres décadas después de haberse iniciado, sigue siendo un proyecto en construcción. Hay que continuar con la ampliación de la infraestructura y del equipamiento para responder a la creciente demanda. En algunos casos (Universidad de la Costa en Pinotepa Nacional, Universidad de Chalcatongo, Universidad del Istmo en Juchitán y NovaUniversitas en Ocotlán), se debe ampliar la oferta educativa para rentabilizar la infraestructura ya existente y ofrecer un abanico más amplio de opciones educativas.





Hemos tenido un gran apoyo por parte de los gobiernos federal y estatal, en lo que se refiere a infraestructura y equipamiento, pero contradictoriamente se ha congelado la creación de plazas docentes, lo que se ha traducido en una carga de trabajo intolerable para los profesores y en la posibilidad de ampliar la oferta educativa. Debe tomarse en cuenta que nuestras universidades responden a un modelo intensivo de trabajo, que además no se limitan a la dimensión educativa, sino que tienen otras dimensiones socialmente fundamentales, como la investigación científica y las de tipo cultural y promoción del desarrollo, que están produciendo un impacto transformador muy grande en las comunidades.

¿Qué opina acerca de la gratuidad en la educación superior pública?

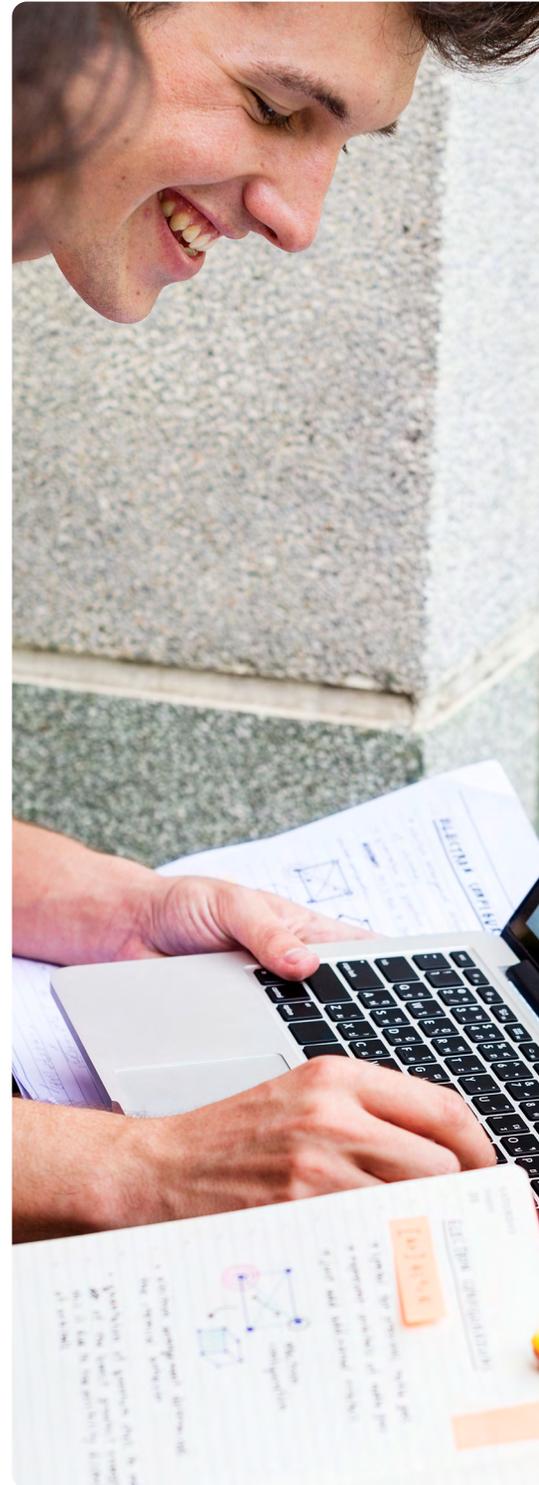
Mi opinión al respecto puede parecer contradictoria a primera vista, pero creo que es totalmente lógica: las universidades no son solamente el instrumento principal para asegurar el desarrollo económico y social de un pueblo, sino que son el único instrumento que garantiza el progreso y la posibilidad de recuperación tras las crisis que aparecen cíclicamente. El desarrollo se puede fomentar a base de inversiones, pero cuando aparece una crisis que paraliza la economía, sea cual sea su causa, la recuperación sólo es posible si hay una población de alto nivel cultural. Un ejemplo claro de ello son Alemania y Japón, que fueron derrotados y destruidos tras la Segunda Guerra Mundial, y que en poco tiempo se situaron a la cabeza de los países desarrollados.

Para asegurar el desarrollo (económico y social) de los países, el único instrumento eficaz es la educación, que tiene que ser de calidad. La inversión en educación y en el terreno más amplio de la ciencia y la tecnología es lo que asegura la independencia y soberanía de los pueblos. Para ser congruentes

con las anteriores afirmaciones, hay que llegar a la conclusión de que la educación no es un gasto, sino una inversión, la inversión más rentable y, por ello, debe ser gratuita.

Eso es totalmente verdad en una sociedad igualitaria, en la que las cargas y los beneficios se reparten equitativamente. Pero cuando nos encontramos en una sociedad desigual, con unos pocos ricos y muchos pobres, la gratuidad de la enseñanza llevaría a la paradoja de asegurar educación gratis, con el dinero del pueblo, a los hijos de las clases privilegiadas y eso no nos parece muy justo, así que lo correcto sería tener cuotas diferenciadas, desde las que cubren el costo total, hasta la gratuidad completa y el apoyo con becas de mantenimiento. Las cuotas se deben determinar en función de la capacidad de pago y no sólo del mérito académico, sin perjuicio de mantener un control sobre la calidad. La función del Estado no es otorgar diplomas que sirvan de prueba de conocimientos que no se tienen. Si lo hiciéramos así, estaríamos creando inválidos sociales permanentes.

Junio de 2021





El nuevo DOMINA-BACH Básico

**María del Rosario Cruz Rodríguez
y Marcela Guadalupe Cano Camacho**
Dirección de los Exámenes Nacionales
de Ingreso, Ceneval

En atención a su compromiso por impulsar instrumentos de evaluación de vanguardia y ofrecer servicios acordes con las necesidades de sus usuarios para contribuir a la calidad de la educación en México, el Centro Nacional de Evaluación para la Educación Superior (Ceneval) se dio a la tarea de diseñar una nueva generación de instrumentos que respondan a los requerimientos educativos actuales.

Así nació el proyecto denominado DOMINA-BACH, cuyo objetivo es aportar información acerca de los conocimientos disciplinares y las habilidades de los estudiantes en dos momentos clave de su formación en el bachillerato. Esta evaluación tiene como antecedente el

DOMINA Competencias Disciplinarias (DOMINA-CD), diseñado por el Ceneval en 2014 con el propósito de brindar información a las instituciones formadoras de bachilleres; tan solo en 2019 se empleó para evaluar a 48 mil sustentantes de 56 instituciones en educación media superior públicas y privadas.

No obstante, dado el contexto educativo actual, resultaba necesaria una actualización. El modelo de la Nueva Escuela Mexicana busca favorecer el desarrollo integral de niñas, niños, adolescentes y jóvenes, y orienta la construcción de una trayectoria educativa que asegure hasta los 23 años de edad una línea progresiva de desarrollo de determinadas competencias. Al visualizar el modelo y su carácter progresivo en la formación de competencias, la evaluación, al ser un proceso continuo, es una herramienta latente, pues para conocer el avance es necesario identificar los contenidos, conocimientos y

habilidades que ya han sido logrados por los estudiantes y que pueden potenciar su desempeño académico, o bien aquellos que están en proceso y deben ser focalizados como áreas de mejora.

Es así que el desarrollo del DOMINA-BACH se alinea a la organización curricular actual de los planes y programas de estudio de las diversas modalidades y subsistemas establecida en el Marco Curricular Común (MCC) de la educación media superior (EMS), fundamentado en la Reforma Integral de la Educación Media Superior (RIEMS).

Para dar cumplimiento a su objetivo, el proyecto está integrado por dos instrumentos de carácter diagnóstico: el DOMINA-BACH Básico y el DOMINA-BACH Avanzado. El primero se encamina a la evaluación del componente de formación básica (común y obligatoria) del bachillerato (segundo año o su equivalente), mientras que el segundo evaluará el componente de formación propedéutica de este nivel educativo (tercer año o su equivalente).

El DOMINA-BACH Básico entrará en operación en noviembre próximo, mientras que el Avanzado estará a disposición de las instituciones usuarias en el otoño de 2022.

DOMINA-BACH Básico

El DOMINA-BACH Básico es un examen diagnóstico que evalúa los conocimientos disciplinares comunes y las habilidades de comprensión lectora y redacción indirecta adquiridos por los estudiantes de bachillerato al concluir el componente de formación básica establecido en el MCC de la EMS, antes de que los estudiantes cursen las áreas de formación propedéutica.

Su propósito es ofrecer información a las instituciones formadoras de bachilleres acerca del nivel de dominio de los conocimientos disciplinares comunes y de las habilidades de comprensión lectora y redacción indirecta, en español e inglés,





adquiridos por los estudiantes de bachillerato al concluir el componente de formación básica establecido en el MCC, para orientar la toma de decisiones remediales. De esta forma, el examen está dirigido a estudiantes que están por concluir el segundo año del bachillerato o su equivalente.

El carácter diagnóstico del examen reside en que explora los conocimientos y las habilidades que poseen los alumnos, con la intención de detectar posibles dificultades o áreas de oportunidad en su proceso de enseñanza aprendizaje y así ofrecer información de juicio que permita la implementación de acciones de mejora, como la intervención docente, los cursos remediales, etcétera. El instrumento no está ligado a decisiones de egreso del nivel medio superior o de aprobación de asignaturas, sino a un proceso de retroalimentación institucional.

El siguiente apartado describe el procedimiento para la determinación de los contenidos del examen, particularmente lo relacionado con los campos disciplinares, ya que lo relativo a las habilidades de comunicación en español e inglés, al tratarse de áreas transversales que se evalúan en el DOMINA-BACH Básico y en otros instrumentos del Ceneval, se aborda en otros números de *Ceneval Investiga*.

Delimitación del objeto de medida de los campos disciplinares

Debido a que el examen evalúa los conocimientos disciplinares de una serie de asignaturas que se relacionan con el proceso de formación académica en los dos primeros años del bachillerato, la selección de los contenidos por evaluar responde a la estructura curricular del MCC de la EMS. En virtud de lo anterior, la delimitación de las áreas tiene fundamento en lo siguiente:

- a) Documentos oficiales para la EMS en México: MCC; planes de estudio de referencia del componente básico del MCC de la EMS y diversos acuerdos. La Secretaría de Educación Pública (SEP) y la Subsecretaría de Educación Media Superior (SEMS) han consolidado desde 2008 la atención de los desafíos de cobertura, calidad y equidad en la EMS a través de la RIEMS. En ella se establecen, entre otros puntos importantes, las estructuras curriculares para la formación académica de los estudiantes de este nivel educativo permitiendo que coincida toda la diversidad de subsistemas en propósitos comunes. Por lo tanto, la evaluación de las áreas disciplinares de este instrumento se ciñe a lo descrito en los documentos oficiales emitidos por las instancias correspondientes.
- b) Programas de estudio de la Dirección General de Bachillerato (DGB): bachillerato general y bachillerato tecnológico; del Colegio de Bachilleres (CB); del Colegio Nacional de Educación Profesional Técnica (Conalep), de la Dirección General de Educación Tecnológica Agropecuaria y Ciencias del Mar (DGETAyCM), y de la Dirección General de Educación Tecnológica Industrial y de Servicios (DGETI) por parte de la SEP y la SEMS. Programas de estudio de la Universidad del Valle de México (UVM). Y, por último, programas de estudio de la Escuela Nacional Preparatoria (ENP) y del Colegio de Ciencias y Humanidades (CCH) de la Universidad Nacional Autónoma de México (UNAM). Los de ésta última se consideraron por ser una entidad emisora de certificados de bachillerato y con un gran impacto en la educación media superior nacional.





El análisis de esta información permitió generar una propuesta de evaluación de cuatro campos disciplinares y dos habilidades de comunicación. Para el primer caso, se realizó un mapeo curricular de los contenidos de las asignaturas impartidas en los diversos subsistemas en cada campo. Así, se identificaron aquellos que son parte del componente de formación básica y, por ende, comunes a todos los subsistemas dentro de los dos primeros años de la EMS. Dicha propuesta de asignaturas a evaluar por campo disciplinar fue elaborada de manera colegiada y sancionada en el Consejo de Planeación Académica del Ceneval.

Una vez aprobada la propuesta, se convocó a dos diferentes Comités Académicos de Diseño para cada asignatura para revisar los contenidos identificados como comunes, determinar aquellos que son clave y preponderantes en la formación de nivel medio superior previo al inicio de la formación propedéutica, definir de manera detallada las áreas y subáreas consideradas en el instrumento de evaluación y diseñar las especificaciones que guían la elaboración de los reactivos de opción múltiple.

Finalmente, en atención a lo referido en la metodología del Ceneval para el desarrollo de instrumentos de evaluación, los productos de estos cuerpos colegiados fueron sometidos a consideración de un Comité Académico de Validación del Objeto de Medida para garantizar la congruencia entre los elementos que lo integran: perfil referencial, marco de sustento, estructura y especificaciones de reactivos.

Estructura del DOMINA-BACH Básico

De las actividades anteriores se obtuvo una estructura integrada por seis áreas:

- › Cuatro que corresponden a conocimientos disciplinares mínimos que, de acuerdo con la organización curricular, deben alcanzarse al término del componente de formación básica del bachillerato establecido en el MCC de las EMS. Estos campos disciplinares se encuentran a su vez conformados por distintas asignaturas que son de carácter común y obligatorio en todas las modalidades o subsistemas del bachillerato, las cuales se reflejan como subáreas en el instrumento de evaluación.
- › Una que atiende a las habilidades de comunicación de comprensión lectora y redacción indirecta en español, que se han identificado como esenciales para la vida escolar de los estudiantes y que se relacionan estrechamente con el logro académico. Se evalúan no sólo a partir de los referentes curriculares nacionales propios de la EMS, sino también con base en importantes referentes internacionales como la prueba PISA de la OCDE o el Laboratorio Latinoamericano de la Unesco.
- › Una que atiende a las habilidades de comunicación de comprensión lectora y redacción indirecta en inglés como lengua extranjera, que se evalúan en un nivel B1 del Marco Común Europeo de Referencia para las Lenguas.

Cada asignatura o habilidad se evalúa a través de 20 reactivos de opción múltiple con tres opciones de respuesta. La longitud total del examen es de 280 reactivos que los estudiantes resuelven en dos sesiones, cada una de 4 horas con un receso de 2 horas entre ellas. El instrumento de evaluación se aplica en cualquiera de las tres modalidades disponibles en el Ceneval (impresa, en línea o *Examen desde casa*).

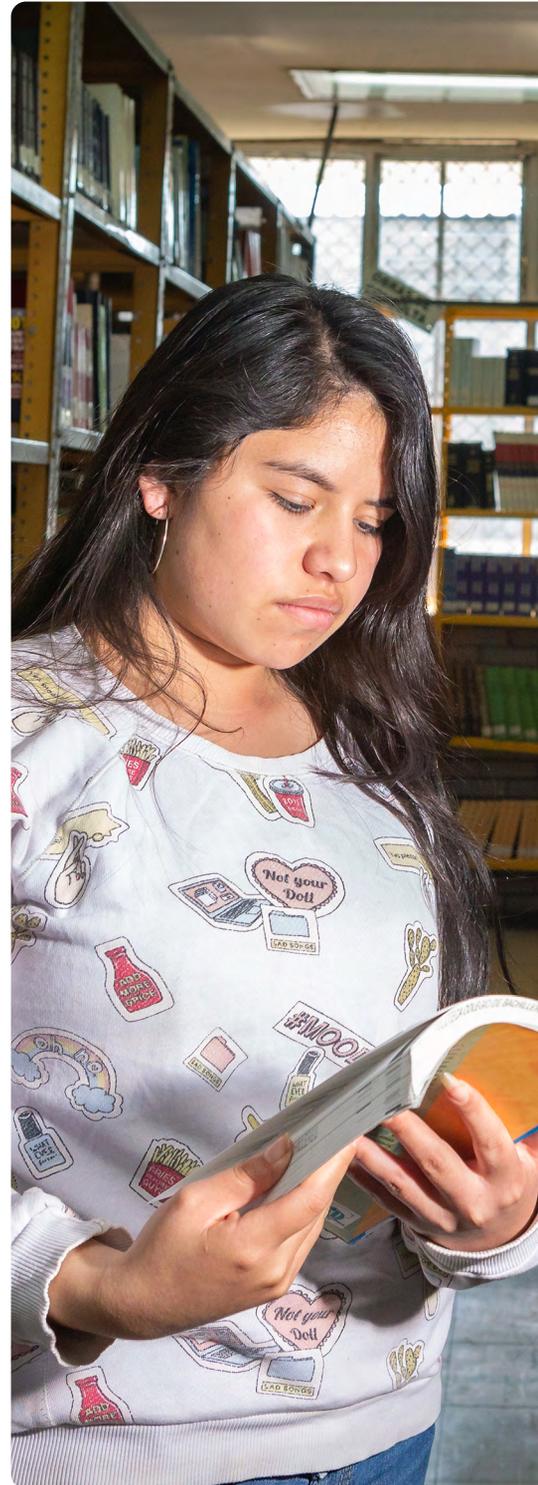


Tabla 1. Estructura del DOMINA-BACH Básico

| Área (Campo disciplinar/Habilidad) | Subárea (Asignatura/Habilidad) | Reactivos por subárea | Reactivos por área |
|---|---|--------------------------|-----------------------|
| Matemáticas | Álgebra | 20 | 60 |
| | Geometría y Trigonometría | 20 | |
| | Precálculo | 20 | |
| Ciencias sociales | Introducción a las Ciencias Sociales | 20 | 40 |
| | Historia de México | 20 | |
| Habilidades de comunicación en español | Comprensión lectora | 20 | 40 |
| | Redacción indirecta | 20 | |
| Ciencias experimentales | Química | 20 | 60 |
| | Física | 20 | |
| | Biología | 20 | |
| Humanidades | Ética | 20 | 40 |
| | Literatura | 20 | |
| Habilidades de comunicación en inglés | Comprensión lectora | 20 | 40 |
| | Redacción indirecta | 20 | |
| Total de reactivos | | 280 | 280 |

El DOMINA-BACH Básico, que el Ceneval pone a disposición de las instituciones de educación media superior, es un instrumento de evaluación que se sustenta en estándares internacionales y que proveerá de información externa que permita reconocer los procesos de aprendizaje consolidados y aquellos en los que es preciso poner énfasis para el logro de sus objetivos.

Lecturas recomendadas

- Allalouf, A. & Alderoqui-Pinus, D. (2012a). Guest Editorial: Dissemination of Measurement Concepts and Knowledge to the Public. *Educational Measurement: Issues and Practice*, 31(2), 1.
- Allalouf, A. & Alderoqui-Pinus, D. (2012b). Museum Exhibition on Testing and Measurement: Scientific Principles, Social Impact of Testing, and Dialogue With the Public. *Educational Measurement: Issues and Practice*, 31(2), 14-20.
- Subsecretaría de Educación Media Superior-Secretaría de Educación Pública (2019). *La Nueva Escuela Mexicana: principios y orientaciones pedagógicas*, México, SEP, disponible en: <http://dfa.edomex.gob.mx/sites/dfa.edomex.gob.mx/files/files/NEM%20principios%20y%20orientacio%C3%ADn%20pedago%C3%ADgica.pdf>





El EGEL Plus Disciplinar del Proyecto Egresada de Licenciatura

**Alejandra Zúñiga Bohigas
y Laura Tayde Prieto López**
Dirección de los Exámenes Generales
para el Egreso de la Licenciatura, Ceneval

En la educación superior, específicamente en el nivel de licenciatura, no existe una estructura uniforme de los currículos de institución a institución, de modo que los planes de estudio son cambiantes y ponen énfasis en distintos aspectos.

Si bien los estudiantes pueden beneficiarse de la multiplicidad de ofertas de las instituciones de educación superior (IES), autores como Medina y Guzmán sostienen que lo importante es que en el área de concentración (lo específico de la carrera) se encuentren los aspectos indispensables de la profesión y que en las áreas

transversales los estudiantes obtengan las experiencias de aprendizaje idóneas para la vida profesional y social.

Lo anterior alude también a las expectativas de empleabilidad del egresado universitario. En el estudio *Educación Superior en México, Resultados y relevancia para el mercado laboral*, la Organización para la Cooperación y el Desarrollo Económicos (OCDE) señala que “pese a que no hay datos representativos para evaluar las competencias de los egresados en México, algunos signos apuntan a unos niveles insuficientes de competencias, tanto transversales como de disciplinas específicas”.

Por su parte, Hazelkorn, Coates y McCormick indican que a la hora de debatir y evaluar la calidad en la educación superior los factores que se plantean son, entre otros, la experiencia del

estudiante y los resultados de aprendizaje, así como la adquisición de competencias disciplinares específicas y de competencias transversales.

Por lo anterior, la renovación de los Exámenes Generales para el Egreso de la Licenciatura (EGEL) no se limita a una actualización de los contenidos de las diferentes profesiones, sino que plantea el desarrollo de una familia de diferentes instrumentos que, en conjunto, constituyen una evaluación integral: el *Proyecto Egresada de Licenciatura* destinado a evaluar el aprendizaje de los egresados en las dos áreas de interés a las que se ha hecho referencia: la disciplinar específica de la carrera y la transversal de lenguaje y comunicación.

En específico, el EGEL Plus —instrumento integrado por dos secciones: el *EGEL Plus Disciplinar* (EGEL+D) y el *EGEL Plus de Lenguaje y Comunicación* (EGEL+LyC)— constituye la base obligatoria y el núcleo del Proyecto Egresada de Licenciatura, ya que, entre otros atributos, fue construido con una metodología sólida con probada validez para la construcción de pruebas criterioles, por lo que sus resultados servirán como indicador para evaluar el aprendizaje de los recién egresados y la eficacia de los programas educativos de las IES que lo utilicen; serán asimismo uno de los referentes más representativos para la mejora de la calidad de la educación superior en México.

EGEL Plus Disciplinar: ¿por qué una evaluación criterial desde el currículo?

Sin entrar en discusiones en torno a la calidad y pertinencia de cada currículo, si un currículo determinado es lo que se enseña o sólo lo que se quisiera o se debería enseñar (y lo que se debería aprender), si es socialmente pertinente o si pudiera mejorarse, lo cierto es que el currículo es la base de





las evaluaciones educativas. Con éstas se evalúa lo que el diseñador de los planes, la escuela o la autoridad educativa espera del estudiante o del egresado como resultado de un proceso de educación formal.

El EGEL Plus Disciplinar es un examen netamente educativo y el referente es lo que se estudia, lo que la universidad aporta o pretende aportar a la formación del egresado, y no una imagen de lo que debería ser idealmente el contenido de la disciplina, su pertinencia social ni lo que hará el egresado cuando tenga un empleo. Estas consideraciones, así como el contenido de la prueba, están determinados por el currículo, en el que ya algunos expertos han definido los componentes de lo que se ha de enseñar, lo que se espera que el egresado aprenda y, por lo tanto, lo que se debe evaluar.

En términos técnicos, según Anthony Nitko, la evaluación con referencia a un criterio alude a interpretar la puntuación de un estudiante comparándola con lo que se espera que aprenda como resultado de la enseñanza de un plan de estudios determinado. El dominio respecto a los objetivos del currículo o resultados de aprendizaje es lo que se denomina *criterio*. El autor sostiene que la referencia a criterio sólo tiene validez cuando se ha evaluado a un estudiante en todo el dominio de los resultados de aprendizaje curriculares, o bien cuando se han seguido procedimientos especiales para obtener una muestra representativa de éstos, por lo que la determinación de dichos resultados de aprendizaje importantes es el núcleo del proceso de construcción de un examen.

Así pues, el diseño del EGEL Plus Disciplinar buscó alcanzar la coherencia entre el currículo planeado, el aprendido y el evaluado, a partir de medidas técnicas y metodológicas, para conseguir una amplia representatividad, especialmente con la participación de diferentes comités académicos y cientos de especialistas, así como de los cinco referentes de alineamiento curricular que a continuación se describen, adaptados del modelo de Nitko (1994) para el desarrollo de pruebas criterioales alineadas a

currículo; de la propuesta *Encuestas de currículo implementado* (*Surveys of Enacted Curriculum*), de Porter y Smithson (2001), y del modelo Webb, desarrollado por Norman Webb (1997):

- i) *resultados de aprendizaje indispensables*:¹ conocimientos y habilidades pertinentes, relevantes y universales que reflejan el perfil que un egresado tipo debe alcanzar al concluir su formación de la licenciatura;
- ii) *conurrencia categórica de contenido*: indica si los resultados de aprendizaje, conocimientos o habilidades analizados y que forman parte de la prueba incorporan los contenidos temáticos del perfil de egreso o programa de estudios de la disciplina;
- iii) *profundidad del conocimiento*: demanda cognitiva o proceso cognitivo (comprensión, aplicación, análisis y evaluación²) que el sustentante ha de evidenciar en los resultados de aprendizaje, conocimientos o habilidades analizados y que forman parte de la prueba;
- iv) *amplitud del conocimiento*: cobertura de los resultados de aprendizaje indispensables del perfil de egreso o programa de estudios, a partir de los conocimientos y las habilidades que los representan y que han sido incorporados en la prueba;
- v) *balance de representación*; considera una distribución adecuada de los conocimientos y las habilidades analizados y que forman parte de la prueba entre los resultados de aprendizaje indispensables del perfil de egreso o programa de estudios de la disciplina.

1 Nitko (1994) hace alusión a resultados de aprendizaje importantes; sin embargo, para propósitos del EGEL Plus Disciplinar se ha hecho la precisión de adjetivarlos como *indispensables*.

2 El EGEL Plus Disciplinar hace uso de la Taxonomía de Bloom, particularmente la versión revisada por sus discípulos y publicada en 2001, la cual establece seis niveles cognitivos. No obstante, el nivel de orden superior definido como *crear* no es usado en el EGEL+D dadas las limitaciones para medirlo en un examen de respuesta cerrada o seleccionada. Por su parte, la categoría del nivel más bajo, *recordar*, también fue excluida, en virtud de que involucra procesos cognitivos que aluden únicamente a "reconocer o recordar conocimientos de memoria".





Estos planteamientos proveen de líneas base para configurar cierta lógica y condiciones que nutren la estrategia metodológica para la delimitación y estructuración del universo de medida del EGEL Plus Disciplinar, partiendo de la consideración de que, ante la ausencia de planes de estudio nacionales y únicos o de un marco curricular común, es preciso retomar y realizar el contraste con experiencias internacionales y nacionales para adecuarlas a las propias necesidades de la evaluación.

EGEL Plus Disciplinar (EGEL+D): principales características de la evaluación

Principios

Evaluación estandarizada de alta calidad, externa al sistema educativo nacional (SEN) y a cualquier otra iniciativa gubernamental, que proporciona información objetiva, válida y confiable sobre resultados de egreso de procesos educativos formales de licenciatura, asegurando la transparencia de sus procesos mediante de la participación colegiada de expertos, así como la equidad, imparcialidad y responsabilidad ética en su desarrollo y resultados.

Objetivo y propósitos

Evaluar el aprendizaje de los estudiantes que han concluido o están por concluir un plan de estudios de licenciatura de educación superior, con el propósito de determinar:

- a) el nivel de desempeño de los egresados;
- b) su grado de dominio en los conocimientos y las habilidades curriculares indispensables al término de la formación académica vinculada a su profesión.

Con ello se pretende proporcionar a las instituciones y a los egresados información integral, relevante, válida y confiable sobre los resultados de desempeño asociados a los programas educativos.

Población objetivo

Estudiantes que estén inscritos en alguna IES y que hayan concluido o estén por concluir estudios formales de licenciatura en alguno de los 39 programas educativos que conforman este proyecto.

Tipo de evaluación

- › *Por su funcionalidad:* evaluación sumativa de egreso
- › *Por el referente para su diseño y construcción:* alineada a currículo
- › *Por el tipo de instrumento que emplea:* prueba objetiva con una longitud de 140 a 160 reactivos de respuesta cerrada o seleccionada, con diseño de opción múltiple. Este tipo de reactivos están constituidos por un enunciado incompleto o una pregunta (base del reactivo) en la que se plantea la demanda de la tarea a partir del aprendizaje o proceso cognitivo involucrado, y se presentan tres opciones de respuesta, de las cuales sólo una es la correcta y el resto son distractores plausibles.





- › *Por el referente para la interpretación de sus resultados:* criterial, con un sistema de interpretación en tres niveles de desempeño: Aún no satisfactorio, Satisfactorio y Sobresaliente.
- › *Por las consecuencias derivadas de sus resultados:* de alto impacto académico, ya que serán referente para la incorporación al Padrón-EGEL: Programas de Alto Rendimiento Académico y el otorgamiento del Premio-Ceneval al Desempeño de Excelencia-EGEL.

Aplicación

- › Tres modalidades de aplicación: examen impreso, examen en línea (presencial) y *Examen desde casa*.
- › Dos sesiones de aplicación con duración total de 6 horas.

Nueve EGEL Plus se integran con reactivos de innovación que requieren del uso de una computadora para ser respondidos, por lo que solo se aplican en las modalidades en línea (presencialmente) y *Examen desde casa*. Este tipo de reactivos favorece la interactividad y la precisión de las mediciones, ya que permite un rango más amplio de posibles respuestas.

Lecturas recomendadas

- Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., Raths, J., y Wittrock, M. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives* (Complete edition). New York: Longman.
- Hazelkorn, E., H. Coates and A. McCormick (2018). *Quality, performance and accountability: Emergent challenges in the global era*. En *Research Handbook on Quality, Performance and Accountability in Higher Education*. Cheltenham: Edward Elgar Publishing.
- Medina Cueva, L., & Guzmán Hernández (2011). *Innovación curricular en Instituciones de Educación Superior*. Ciudad de México: ANUIES.
- Nitko, A. (1994, julio). *A model for developing curriculum-driven criterion-referenced and norm-referenced national examinations for certification and selection of students*, ponencia presentada en la Conferencia Internacional sobre Evaluación y Medición Educativas, de la Asociación para el Estudio de la Evaluación Educativa en Sudáfrica (ASSESA). Recuperado el 14 de noviembre de 2020, de: <https://eric.ed.gov/?id=ED377200>
- OCDE (2019a). *Educación Superior en México. Resultados y relevancia para el mercado laboral*. París: OECD Publishing. Obtenido de: https://www.oecd.org/centrodemexico/medios/educacion_superior_en_mexico.pdf
- Porter, A. & Smithson, J. (2001). *Defining, Developing, and Using Curriculum Indicators*. USA: University of Wisconsin-Madison, pp. 9-15. Recuperado el 25 de noviembre de 2020, de: https://www.researchgate.net/publication/252368536_Defining_Developing_and_Using_Curriculum_Indicators
- Tobón, S., Rial Sánchez, A., Carretero, M. Á., & García, J. A. (2006). *Competencias y Calidad en Educación Superior*. Bogotá: Alma Mater.
- Webb, N. (1997). *Criteria for alignment of expectations and assessments in Mathematics and Science education*. Madison, WI: University of Wisconsin-Madison; National Institute of Science Education, NISE Research Monograph, No. 6. Recuperado el 30 de noviembre de 2020 de: https://www.researchgate.net/publication/234731918_Criteria_for_Alignment_of_Expectations_and_Assessments_in_Mathematics_and_Science_Education_Research_Monograph_No_6





**Tania L. Jaramillo Reyes
y Salvador Saulés Estrada**
Subdirección de Exámenes
Transversales, Ceneval

La escritura es una de las tecnologías más avanzadas que ha desarrollado el ser humano. Los elementos que se involucran al redactar un texto abarcan habilidades socioculturales y cognitivas que, al fortalecerse adecuadamente, propician una comunicación exitosa con los lectores.

En el Ceneval se han generado pruebas que miden el nivel de competencia y las habilidades para comunicarse por escrito. Las evidencias se recaban, por ejemplo, de forma indirecta a partir de respuestas a reactivos de opción múltiple. Tal es el caso del área de redacción de los EXANI, los DOMINA-BACH, o el EGEL Plus. En estas pruebas, los sustentantes seleccionan fragmentos textuales a partir de ciertos criterios determi-

EXPRESE

Examen de Expresión Escrita en Español

Evaluación de las habilidades
para comunicarse de forma escrita

nados. No obstante, también se puede recabar evidencia de forma directa acerca de la habilidad que posee un sustentante para redactar. Un ejemplo de ello es el Examen de Expresión Escrita en Español, conocido como EXPRESE.

El Ceneval ha elaborado dos materiales que sustentan este examen: el *Marco teórico* y la *Guía para el Sinodal*. Buena parte de la información incluida en este artículo procede de estas dos importantes fuentes.

Expresión escrita: tipologías y discursos

Aunque los textos son fenómenos lingüísticos que pueden adquirir múltiples formas, como aquellos que se han adaptado a los tiempos de las tecnologías de la información para comunicarse de forma inmediata (correo electrónico, mensaje de texto, tuit o publicación de Facebook), en los que la redacción suele ser sencilla, con rasgos espontáneos cercanos a la oralidad o producto de decisiones inmediatas, también existen otros textos con mayor formalidad que

pueden ser agrupados por sus características de acuerdo con sus objetivos o funcionalidades. Tanto los textos inmediatos como los más formales pueden ser organizados a partir de su tipología textual; algunas teorías proponen cinco tipos: narrativo, descriptivo, argumentativo, explicativo y conversacional.³

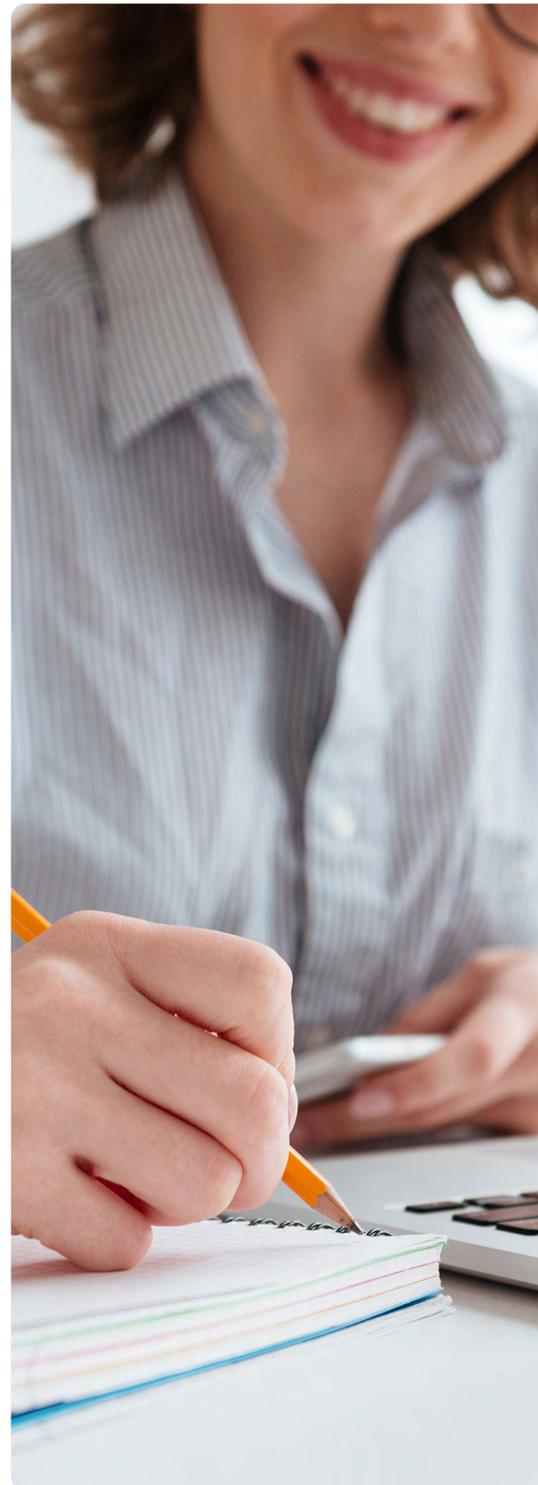
La escritura se complejiza desde otra perspectiva cuando se elige un tipo de discurso. Desde el enfoque sociolingüístico, un escrito como acto social tiene el propósito de comunicar, por lo que se adapta al contexto y al público al que va dirigido el mensaje. No se trata sólo de los conocimientos sintácticos y semánticos que permiten organizar ideas y dar una estructura, sino de la elección y dominio del discurso que deseamos emplear, sea para dar sustento a una tesis, crear una descripción detallada en el interior de una historia o dar a conocer una noticia.

Al elaborar un escrito, la persona debe plantearse una “etnografía de la escritura”. En 1996, Grabe y Kaplan propusieron un modelo desde una perspectiva que comprende a la escritura como una manifestación con características específicas y no como la derivación de la lengua hablada. Este modelo en términos generales se basa en las preguntas siguientes,⁴ que sirven como guía al acto comunicativo escrito: ¿quién escribe qué?, ¿a quién?, ¿para qué propósito?, ¿por qué?, ¿cuándo?, ¿dónde? y ¿cómo?

En este sentido, Salvador Mata considera que “la expresión escrita representa el más alto nivel del aprendizaje lingüístico, por cuanto en ella se integran experiencias y aprendizajes relacionados con todas las habilidades lingüísticas (escuchar, hablar y leer) y se ponen en funcionamiento todas las dimensiones del sistema lingüístico (fonológica, morfo-sintáctica, léxico-semántica y pragmática)”.

³ Bassols y Torrent (1996)

⁴ Ceneval (2015)





¿Cómo se logra la habilidad para escribir?

Una de las mayores aportaciones para comprender el desarrollo de las habilidades para expresarse de forma escrita fue el modelo de Hayes y Flower. En este modelo se tratan los componentes, no sólo internos sino externos, que influyen en quien escribe: “la memoria a largo plazo que incluye tres tipos de áreas de conocimiento, tema general del texto, el acto comunicativo y el conocimiento lingüístico vinculado a la producción textual, y finalmente, el proceso general de composición escrita, compuesto por los tres procesos básicos de planificación, traducción y revisión, dirigidos por un cuarto proceso de control, la monitorización, cuyo objetivo es regular la secuencia recursiva de la composición escrita”.⁵

A partir de este modelo surgieron otros, como el de los ya citados Grabe y Kaplan, quienes enriquecieron la idea de que la escritura es un acto comunicativo compuesto por dos principales bloques: los procesos cognitivos de la o el escritor y el contexto de la persona.

Para estos autores, en los aspectos externos se encuentra el conjunto que conforma el contexto: una situación que contiene un escenario, una tarea, un texto y un tema, a partir de lo cual se generará un texto, según el desempeño. El componente interno o “memoria verbal de trabajo”, por su parte, está “integrado por el establecimiento de metas, el procesamiento verbal y la producción del procesamiento [...]. El establecimiento de metas activa el procesamiento verbal, el cual se compone de la competencia del lenguaje, el conocimiento de mundo y el ensamble del procesamiento en línea. Los dos primeros elementos son parte de la memoria a largo plazo, requerida para el procesamiento de la tarea”.⁶

5 Rodríguez *et. al.*, 2010

6 Ceneval (2015)

Existen asimismo enfoques relacionados directamente con la forma en que se adquiere la habilidad de la escritura en los sistemas escolares, como el Modelo General de Discurso escrito de Anneli Vähäpassi (1988), el cual postula que existen al menos tres habilidades principales involucradas en la expresión escrita vinculadas a un tipo de discurso:

1. Discurso documental: cuando sólo se *reproduce*, pues “el escritor registra material con poca o nula modificación (copiar, tomar nota, dar respuestas cortas a preguntas en libros de texto)”. Esta movilización de información se encuentra en un primer nivel de complejidad.
2. Discurso reporteril: se refleja al *organizar* cierta información, como “fenómenos, conceptos o estados mentales ya conocidos”. Incluye el discurso narrativo, el descriptivo y el explicativo, los cuales tienen un nivel dos de complejidad, en cuanto a habilidades cognitivas involucradas.
3. Discurso exploratorio: se *inventa o genera* nueva información, como “conceptos, estados mentales o fenómenos nuevos o alternativos”. Este nivel, el más complejo, se observa en un texto argumentativo o literario.

Estos niveles no funcionan de manera aislada; pueden estar los tres en un mismo texto, por ejemplo, en un ensayo argumentativo; puede haber reproducción textual de información al hacer una cita, describir un hecho o situación para ejemplificar un concepto y la aportación de nueva terminología para referirse a una nueva perspectiva.

¿Cómo saber cuál es la habilidad que una persona posee para comunicarse por escrito?

Con base en el marco anterior, el Ceneval diseñó el EXPRESE con la finalidad de proporcionar información sobre la competencia





que posee una persona para expresarse de forma escrita en español, mediante la construcción de un ensayo o texto argumentativo. Esta tipología ha sido considerada por los autores mencionados como el tipo de secuencia textual que proporciona de forma más amplia datos acerca de la competencia escritural de los sustentantes.

Para este instrumento, la expresión escrita es definida como aquella capacidad que permite al individuo elaborar un ensayo o texto argumentativo que utilice un lenguaje adecuado y apropiado a un tema o propósito determinado, establezca una relación lógica entre sus ideas principales y secundarias, manifieste una postura con razones y ejemplos, distribuya de forma coherente la información de acuerdo con el texto solicitado, utilice de forma adecuada los conectores para relacionar las partes del texto y aplique adecuadamente las normas ortográficas.

El enfoque taxonómico en que se sustenta la prueba parte sobre todo del modelo sociolingüístico de Grabe y Kaplan, el cual propone, como vimos anteriormente, que la comunicación escrita requiere del conocimiento lingüístico (aspectos ortográficos), discursivo (cohesión y coherencia) y sociolingüístico (contexto). Al adaptar esta teoría a los objetivos del instrumento,⁷ más que conocimientos, se evalúan tres grandes habilidades:

1. *Habilidad sociolingüística*. Uso apropiado del lenguaje según el contexto social: usos funcionales de la lengua escrita, parámetros situacionales y de registro (formal). Para evaluar esta habilidad, se consideran los criterios:
 - › Adecuación: grado de adaptación del discurso a la situación comunicativa; claridad y propiedad con que el sustentante emplea el lenguaje, según el tema y propósito de la tarea;

7 Ceneval, 2015a

- › Organización: relación lógica entre ideas principales y secundarias mediante una estructura organizada del discurso, y
 - › Argumentación: forma en que el sustentante manifiesta sus ideas, con una postura clara, acompañada de argumentos y ejemplos.
2. *Habilidad discursiva*. Estructura del texto, coherencia y cohesión, lo cual permite al lector construir un modelo mental de comprensión. Las evidencias que se recaban en este nivel son el grado en que se establecen las relaciones semánticas y el reconocimiento de temas principales mediante los criterios:
- › Coherencia: distribución de la información en un orden lógico de las ideas y el del tipo de texto solicitado;
 - › Cohesión: conjunto de relaciones o vínculos de significado entre distintos elementos o partes del texto.
3. *Habilidad lingüística*. Se refiere a los elementos básicos de una lengua: código escrito, fonología y morfología, vocabulario, sintaxis. Aunque los criterios lingüísticos que podrían evaluarse son numerosos, la delimitación de este instrumento es relevante y pertinente, al evaluar la correcta aplicación de las normas ortográficas, sintácticas y gramaticales (que no estén cubiertos en la habilidad discursiva), de acentuación y puntuación. El criterio se denomina:
- › Convenciones de la lengua.

Habilidad de la escritura en datos concretos

El ensayo o texto argumentativo del EXPRESE se desarrolla a partir de una tarea dada que implica exponer una opinión o una postura acerca de una temática social, lo cual no requiere de un conocimiento especializado por parte del sustentante.

La calificación se basa en un modelo de evaluación criterial y una rúbrica analítica, de acuerdo con el marco teórico del EXPRESE, gracias a lo cual “se puede obtener un dictamen individual para cada una de las dimensiones que componen el





Datos importantes del EXPRESE:

- › Comenzará su operación en **diciembre del año en curso**
- › Puede aplicarse de forma independiente o junto con el Proyecto Egresado

Para mayores informes consulte aquí.

constructo”.⁸ Esto resulta útil “si lo que se requiere es dar retroalimentación detallada con un fin específico”.⁹ De este modo, gracias a este tipo de rúbrica se puede recabar información de forma global y por cada uno de los seis criterios que se evalúan. Los dictámenes están organizados con base en la siguiente escala de cuatro niveles de desempeño: Insuficiente, Suficiente, Bueno y Notable.

Para que la evidencia sobre la competencia comunicativa por escrito sea precisa, el examen es calificado por al menos dos sinodales. Si existen diferencias significativas entre la calificación otorgada por los primeros dos calificadores, se debe realizar una tercera evaluación con un tercer calificador.

La utilidad de estos datos tiene estrecha relación con el tipo de reportes que se generan. En este caso se desprenden reportes individuales e institucionales. Los primeros detallan los criterios mencionados más el puntaje global; en el segundo caso se ofrece una lista con los dictámenes de cada uno de los participantes de una determinada institución.

Un punto por destacar del instrumento EXPRESE es que, dadas sus características de medición por criterios evaluados y por cuatro niveles de desempeño, permitirá ofrecer sugerencias de mejora generales a cada uno de los sustentantes, según los resultados obtenidos. Pocos instrumentos de evaluación presentan estas características de devolución formativa de resultados.

Conclusiones

La importancia de la escritura radica en su inevitable uso en la vida cotidiana, laboral, social y académica. Por eso es necesario evaluarla. Comunicar de manera efectiva implica desarrollar

⁸ Quinlan (2006)

⁹ Wilkerson (2012)

las habilidades necesarias para adecuarse y cumplir el propósito comunicativo. Gracias a la evaluación de las habilidades para comunicarnos de forma escrita en español, se pueden identificar las áreas de oportunidad e implementar técnicas didácticas que permitan un buen desarrollo de dichas habilidades.

“El pensamiento –de acuerdo con la Unidad de Apoyo para Aprendizaje de la UNAM– crea y determina al lenguaje; a su vez, el lenguaje enriquece y revoluciona al pensamiento. De tal forma, ambos mantienen una relación mutua, interdependiente y dinámica. La escritura activa posibilita el desarrollo del pensamiento en funciones como la percepción, la atención y la memoria”, por lo que, si se desarrolla la habilidad de la escritura, se desarrolla el pensamiento y se abren las posibilidades para la adquisición de nuevas competencias para el siglo XXI.

Lecturas recomendadas

- Bassols, M. y Torrent, A. M. (1996). *Modelos textuales. Teoría y práctica*. Eumo Editorial.
- Bereiter, C. y Scardamalia, M. (1987). *The Psychology of Written Composition*. Lawrence Erlbaum Associates Publishers.
- Ceneval (2015). *Marco teórico. Examen de Expresión Escrita en Español. Expresse*. Dirección Técnica y de Investigación.
- Ceneval (2015a). *Guía para el Sinodal. Examen de Expresión Escrita en Español. Expresse*. Dirección Técnica y de Investigación.
- Grabe, W., y Kaplan, R. (1996). *Theory and Practice of Writing. An Applied Linguistic Perspective*. Longman.
- Rodríguez, C., González, P., Álvarez, L., Cerezo, R., García, J. N., González, J. A., Alvarez, D., Bernardo, A. B., (2010). *Un análisis de los modelos teóricos actuales de escritura y su relación con el trastorno por déficit de atención e hiperactividad (TDAH)*. Papeles del Psicólogo, núm. 2, mayo-agosto, pp. 211-220. Consejo General de Colegios Oficiales de Psicólogos.
- Salvador, F. (2021). *Habilidad de expresarse por escrito*. Docer Argentina. <https://docer.com.ar/doc/x01v51v>
- Unidad de Apoyo para el Aprendizaje (s/f). Pensamiento, lenguaje y escritura. Coordinación de Universidad Abierta y Educación a Distancia-UNAM. <https://proyectos.cuaed.unam.mx/uapa/derecho/pensamiento-lenguaje/>
- Vähäpassi, A. (1988). The Domain of School Writing and Development of the Writing Tasks. En T. P. Gorman, A. C. Purves y R. E. Degenhart (Eds.), *The IEA Study of Written Composition I: The International Writing Tasks and Scoring Scales* (pp. 15-39). Pergamon Press





Hacia un modelo único para la medición de habilidades socioemocionales en el Ceneval

Ingrid García Pinzón

Jefa del Departamento de Investigación del Ceneval

Actualmente, el objetivo general de los cuestionarios de contexto está enfocado en cubrir las necesidades de información propias del Ceneval, las instituciones usuarias y las autoridades educativas. Por ello se requiere emprender cursos de acción muy claros, que enfatizan la necesidad de establecer bases teóricas sólidas y un modelo único para la medición de habilidades socioemocionales, ya que es información necesaria para ayudar a la detección y solución de problemas y apoyar en la toma de decisiones. Para ello, en 2018 se probó empíricamente un modelo teórico propuesto por Farrington y sus colegas en 2012, con muestras de estudiantes mexicanos: alumnos de secundaria que solicitan su ingreso al nivel medio superior, estudiantes de educación media superior que solicitan el ingreso a instituciones de nivel

superior y personas que solicitan el ingreso al posgrado, por medio de los Exámenes Nacionales de Ingreso (EXANI-I, II y III).

El objetivo de este estudio es establecer tres grupos con distintos niveles de desarrollo en cada una de las habilidades socioemocionales consideradas: cooperación, compromiso académico, perseverancia, gusto por la escuela y metacognición. Las preguntas sobre habilidades socioemocionales se integran a cuestionarios de contexto que se aplican en formato impreso y en línea. Esto permitirá inferir el nivel de desarrollo de estas habilidades en los sustentantes, independientemente del nivel educativo en el que se encuentren. Esta información, sumada a los datos socioeconómicos de cada alumno y a su desempeño en el examen al que acompaña cada cuestionario de contexto, permitirá tener una visión más integral con respecto a las características de los sustentantes de los instrumentos del Centro.

En la primera parte de este artículo se presenta información sobre la construcción de tres modelos por medio de los cuales se establecerán puntos de corte que definan los niveles de desarrollo de las habilidades socioemocionales. Finalmente se presentan algunas comparaciones entre estos tres modelos, así como recomendaciones sobre el uso de estas propuestas.

Tres modelos

Este estudio se realizó con 1 228 782 sustentantes (37.7% de EXANI-I, 60.6% de EXANI-II y 1.7% de EXANI-III). Los resultados hacen referencia a tres modelos para cada una de las escalas consideradas: análisis factorial, modelo de crédito parcial y regresión lineal. Además, con la finalidad de identificar tres niveles que sean útiles para retroalimentar a las instituciones usuarias y a los propios sustentantes, los estratos fueron contruidos con el método de Dalenius-Hodges, que consiste en la formación de grupos lo más parecidos posible, con la mínima varianza al interior de cada estrato y la máxima varianza entre los diferentes grupos. Para la construcción se siguió la Nota Técnica de Estratificación Unificada del Sistema para la Consulta de Información Censal 2010¹⁰ del Inegi.

A continuación, se describe brevemente esta metodología. Primero, se establecen ponderaciones de cada una de las preguntas para cada escala, según el valor que le corresponde de acuerdo con cada modelo. Para el análisis factorial se utilizó la carga factorial para establecer la ponderación de las preguntas. Para el modelo de crédito parcial se utilizó la dificultad de cada pregunta, aprovechando las ventajas de este análisis que coloca en una escala común el puntaje de las personas y la dificultad de los reactivos. Finalmente, para el modelo de regresión,



10 http://gaia.inegi.org.mx/scince2/documentos/scince/metodo_notatecnica.pdf



se utilizó, como en otras mediciones (por ejemplo la de nivel socioeconómico AMAI¹¹), la beta o efecto de cada pregunta, sin considerar el intercepto del modelo.

A partir de las ponderaciones del modelo factorial, crédito parcial y regresión lineal se calculó el puntaje ponderado de cada constructo, es decir: se multiplicó el valor de respuesta de la pregunta por el valor de ponderación que le corresponde y su resultado se sumó con cada resultado obtenido en las preguntas que forman la escala. Con el resultado se realizó un análisis descriptivo (media, máximo y mínimo) de todas las variables construidas.¹²

Se estableció un conjunto de clases, las cuales se definen multiplicando por 10 el número de grupos que se obtendrán. Para este caso en particular, se establecerán tres grupos; por lo tanto, se consideraron 30 clases (J). Posteriormente, se dividieron los puntajes obtenidos, que van de 100 a 400, en estas 30 clases (J); para ello, en cada una de ellas se estableció un límite inferior y un límite superior que se definen por las siguientes formulas: límite inferior= $100+(J-1)*(400-100)/30$ y límite superior= $100+(J)*(400-100)/30$. Donde 100 y 400 son los valores mínimo y máximo y 30 es el número de clases.

A continuación, se calculó la frecuencia para cada una de las 30 clases, según los límites establecidos, la raíz cuadrada de la frecuencia y la raíz cuadrada acumulada; con esta información se podrán calcular los puntos de corte obteniendo los valores de Q y (h-1)Q.

Donde:

h = grupos (que en este caso son tres)

Q = suma de los valores de la raíz de las frecuencias/3.

11 https://www.amai.org/descargas/REVISION_AMAI_2018_v2.pdf

12 http://gaia.inegi.org.mx/scince2/documentos/scince/metodo_notatécnica.pdf

Los valores obtenidos de Q y (h-1)Q permiten definir entre qué valores de la raíz acumulada están los puntos de corte y dividir los grupos en tres para este caso.

Con estas consideraciones y siguiendo la metodología descrita se establecieron dos puntos de corte y se crearon tres grupos, el primero con niveles más bajos (con respecto a los otros dos grupos) de cada una de las habilidades socioemocionales. El segundo es un grupo intermedio y el tercero tiene los niveles más altos de esta población (sustentantes de EXANI-I, II y III) de cada una de las habilidades socioemocionales descritas anteriormente.

Análisis factoriales

A continuación, se presentan los resultados generales de los análisis factoriales confirmatorios realizados para cada escala. Para este punto se debe considerar que, debido a la naturaleza de las preguntas (con cuatro opciones de respuesta) se utilizó el método adecuado para la estimación (WLSMV). El ajuste de los modelos se evaluó mediante las pruebas de ajuste de χ^2 , RMSEA, SRMR y CFI. Se consideró que valores de RMSEA menores a .05, de SRMR menores a .06 y de CFI mayores a .95 son indicativos de un buen ajuste a los datos; también que el modelo presenta un ajuste adecuado cuando el RMSEA es menor a .08 y el CFI mayor a .90.¹³

De acuerdo con estos análisis, de los resultados asociados al EXANI-I, II y III se puede concluir que existe evidencia de validez de constructo para todas las escalas, ya que los modelos factoriales confirmatorios tuvieron un ajuste adecuado y las cargas factoriales fueron mayores a 0.5. Además, existe

¹³ Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage focus editions*, 154, 136-136; West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. *Handbook of structural equation modeling*, 1, 209-231.





evidencia de validez convergente ya que las escalas correlacionan entre ellas más allá de 0.5. Finalmente, se destaca que para cada escala fue posible identificar ponderadores para cada pregunta considerada.

Análisis de crédito parcial

Los siguientes resultados se obtuvieron por medio del modelo de crédito parcial, el cual pertenece a la familia de los modelos de la teoría de respuesta al ítem. La razón principal para utilizarlo es que resulta el más apropiado para preguntas que pueden tener respuestas parcialmente correctas. El modelo fue desarrollado por Master y Wright y Master en 1982 y comparte las características de la familia de modelos de Rasch.

Para calcular el ajuste al modelo de Rasch se estiman los puntajes de respuesta de los sustentantes en lógits, y se obtienen los valores de “infit” y “outfit”, los cuales son indicadores del ajuste de los datos empíricos al modelo. El ajuste que las preguntas deben tener para poder ser aceptadas debe estar entre 0.7 y 1.3, tanto para el infit como para el outfit.

Se consideró el puntaje en lógits para cada pregunta (que representa a la dificultad) para la extracción de los pesos; sin embargo, como la metodología para el establecimiento de grupos no permite el uso de coeficientes o ponderaciones negativas, y aprovechando que los puntajes generados por medio del modelo de crédito parcial son de naturaleza intervalar, dentro de cada escala se obtuvo el valor absoluto de la dificultad mínima obtenida y se le sumó uno, para trasladar el valor mínimo de la escala a uno. Por ejemplo, en el caso de la escala de cooperación, el valor mínimo de dificultad fue de -1.17, por lo que se sumó 2.17 a todos los valores obtenidos, y así el valor mínimo para la escala fue 1.

En cuanto a la confiabilidad o consistencia interna, todos los valores de alfa de Cronbach fueron mayores a 0.8: el mayor

valor fue para perseverancia (0.87) y el menor (0.82), para las escalas de gusto por la escuela.

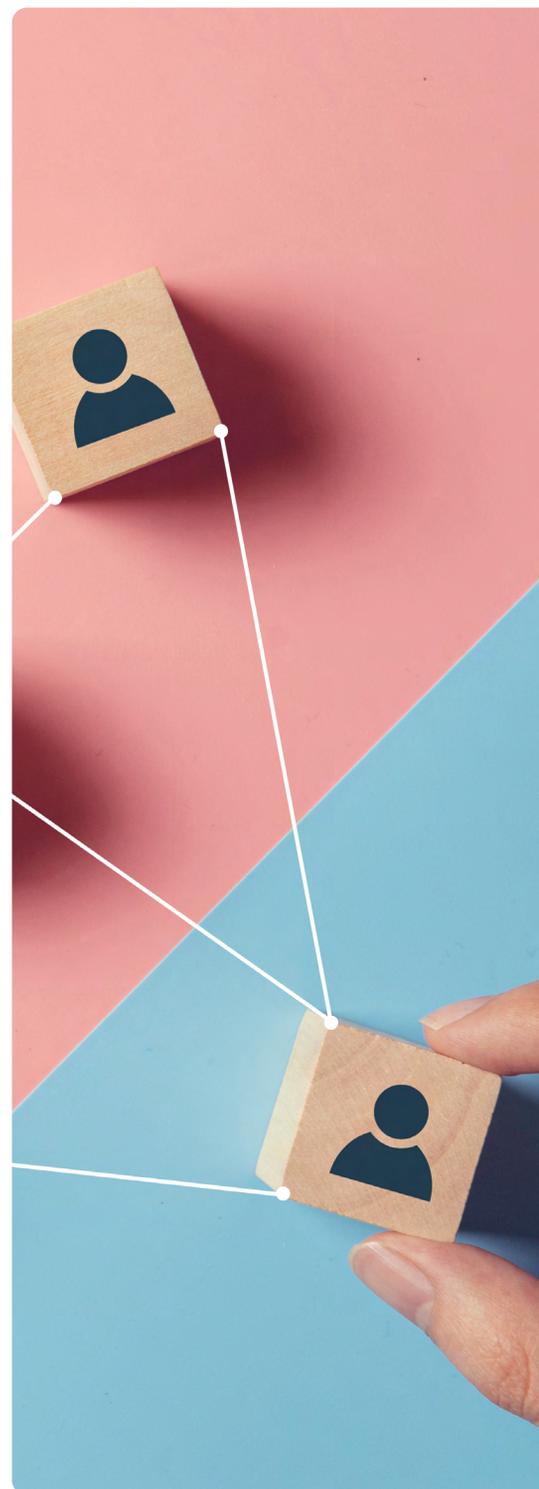
Por lo tanto, los resultados para los EXANI-I, II y III muestran que estas escalas son adecuadas para medir a estudiantes que desean ingresar a la educación media superior y superior (licenciatura y posgrado) considerando los supuestos y parámetros del modelo de crédito parcial.

Análisis de regresión

En el ámbito de la evaluación educativa, los modelos más utilizados son de tipo psicométrico, generalmente relacionados con la teoría de respuesta al ítem. En este estudio, debido a la necesidad de generar una escala única, que además permita su uso práctico, se utilizó como base la metodología indicada en la Nota Técnica de Estratificación Unificada del Sistema para la Consulta de Información Censal 2010¹⁴ del Inegi, en la cual se utiliza un modelo de regresión para la asignación de los pesos. Así, además de las metodologías más comunes en evaluación educativa también se consideró un modelo de regresión para cada escala, en el que se definió como variable dependiente el puntaje obtenido a partir de la suma de las preguntas para cada caso y las variables independientes fueron los puntajes de las mismas preguntas, de las que se extrajo el coeficiente del efecto (β).

Como indicadores de ajuste se consideraron los valores del R cuadrado que es la razón de la varianza explicada y la varianza total, y la significancia de cada una de las variables independientes (todas significativas considerando $p < 0.01$).

Siguiendo la metodología descrita se establecieron tres grupos: el primero, con niveles más bajos (con respecto a los otros



14 http://gaia.inegi.org.mx/scince2/documentos/scince/metodo_notatecnica.pdf



dos grupos) de cada una de las habilidades socioemocionales; el segundo es un grupo intermedio y el tercero, con los niveles más altos de cada una de las habilidades socioemocionales descritas anteriormente. Es importante recordar que esta población incluye a los sustentantes de los EXANI-I, II y III.

Comparación de las diferencias de medias con los grupos generados a partir de los tres modelos

Una vez establecidos los grupos para cada una de las cinco habilidades socioemocionales, a partir de los tres distintos modelos explicados anteriormente, y con la finalidad de identificar cuál de estas metodologías producen resultados más alineados al modelo teórico utilizado, se llevó a cabo una serie de pruebas t para identificar diferencias de medias entre cada grupo, considerando el nivel socioeconómico y el puntaje en la prueba de desempeño académico. Este procedimiento se realizó en primer lugar con la muestra total; posteriormente, se replicó el ejercicio con la muestra aleatoria de 1% de la muestra total. Después, se ejecutó de manera independiente para la población de los EXANI-I, II y III. Finalmente, estas pruebas se replicaron con la muestra aleatoria de 1% de la muestra total. La extracción de la muestra de 1% de los datos se realizó en todos los casos con la finalidad de identificar si los resultados eran distintos, considerando la sensibilidad de la significancia estadística al tamaño de muestra.

En el caso de la población general que sustentó los EXANI-I, II y III hubo diferencias significativas entre los grupos en la mayoría de los casos. No se encontraron diferencias significativas en cuanto al nivel socioeconómico; únicamente para los modelos de crédito parcial de las escalas de compromiso académico y gusto por la escuela. Para el caso de la muestra, las diferencias por nivel socioeconómico desaparecen, con excepción de la escala de perseverancia. En esta comparación se destaca

que la significancia se puede deber en algunos casos a las diferencias de N; con muestras más grandes, la significancia se incrementa. Sin embargo, cabe resaltar que en todos los casos las diferencias por desempeño académico en los exámenes se conservaron. El modelo que se comportó más estable fue el de crédito parcial.

Cuando sólo se consideran los resultados del EXANI-I, los resultados son muy similares a los encontrados en la muestra general del 1%. Se destaca que en cuanto al índice Ceneval únicamente el modelo de regresión lineal no mostró diferencias significativas considerando la variable de gusto por la escuela. Para el caso del nivel socioeconómico, en general no se encontraron diferencias significativas, excepto para las escalas de gusto por la escuela y metacognición, en todos los modelos y para el nivel socioeconómico en el modelo de regresión simple. En la muestra de 1% del EXANI-I, todos los modelos resultaron significativos en cuanto al índice Ceneval y no significativos para el nivel socioeconómico, con excepción del modelo de regresión, que no obtuvo diferencias significativas en el desempeño académico del examen para los constructos de cooperación, perseverancia y gusto por la escuela.

Los datos del EXANI-II arrojan en la muestra general diferencias significativas para casi todos los casos, excepto para el análisis factorial en el constructo de compromiso académico con respecto al nivel socioeconómico. Lo mismo ocurrió para ese modelo en el constructo de perseverancia con respecto al índice Ceneval. En la muestra extraída de la población de este examen no se encontró ninguna diferencia significativa por cuanto hace al nivel socioeconómico. Además, destaca que en la mayoría de los casos se encontraron diferencias significativas en el desempeño académico del examen, con excepción de los constructos de gusto por la escuela y perseverancia, pues los modelos de regresión y análisis factorial no presentaron diferencias significativas entre sus medias establecidas con los grupos.





En lo que corresponde al EXANI-III, las diferencias significativas decaen en general con respecto al resto de las poblaciones. Esto se puede entender en función de que estos sustentantes se encuentran por ingresar ya a estudios de posgrado y son una muestra de que ya ha concluido o está a punto de concluir la educación superior. Teniendo esto en consideración, se observa que para los constructos de metacognición y perseverancia no se encontraron diferencias significativas en ningún caso. La única diferencia significativa encontrada con respecto al nivel socioeconómico fue la de la escala de cooperación. En cuanto al índice Ceneval, para aquellos constructos donde sí se encontraron diferencias significativas (cooperación, compromiso académico y gusto por la escuela), esta significancia desapareció con el modelo de regresión. Para la muestra del 1%, se encontraron muy pocas diferencias de medias significativas; fueron en todos los casos las de la escala de compromiso académico y para el modelo de crédito parcial fue con respecto al desempeño en el examen, mientras que para el modelo de regresión lineal esta diferencia fue significativa para el nivel socioeconómico.

Recomendaciones generales

Dados estos resultados, no se recomienda utilizar el modelo de regresión. Este estudio muestra que este modelo asigna valores a las preguntas que derivan en puntajes que en muchos casos no cumplen con los objetivos del modelo de Farrington, el cual es explicar el desempeño académico. En algunos casos extremos como el de la muestra de 1% del EXANI-III, este modelo parece estar más relacionado con el nivel socioeconómico que con el desempeño académico, mientras que para el Ceneval es relevante generar puntajes que expliquen o den cuenta del desempeño académico, más que del nivel socioeconómico.

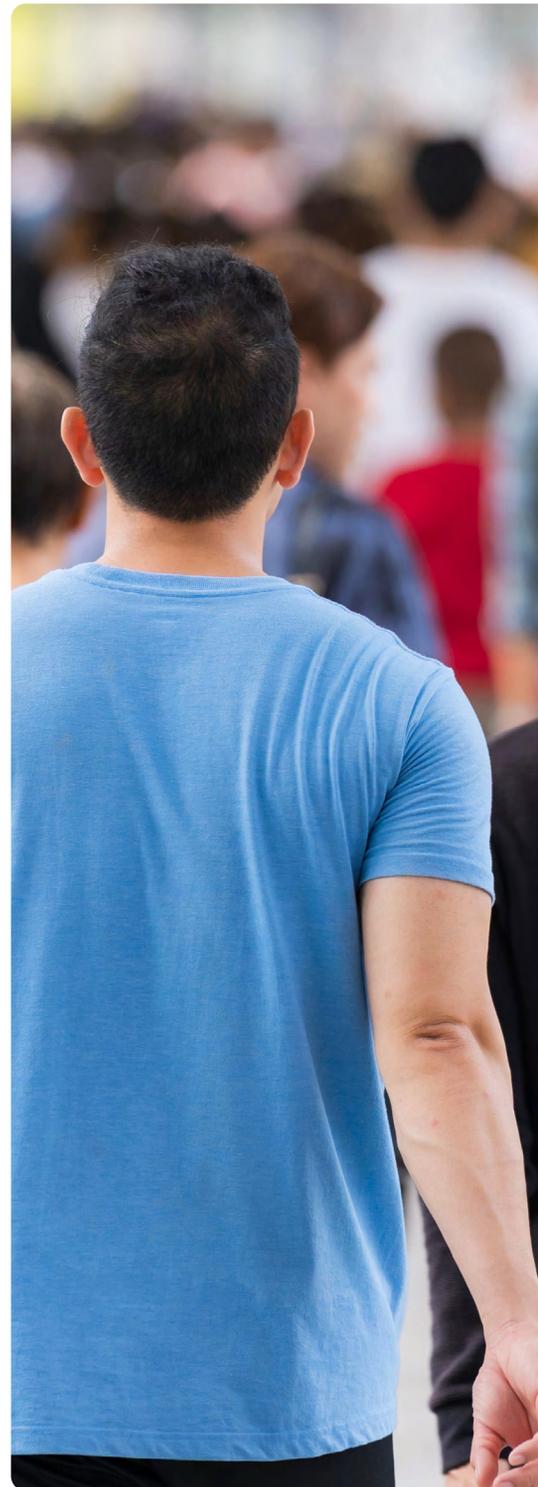
El modelo basado en el análisis factorial parece ser mejor que el de regresión; sin embargo, a mayor nivel educativo, se vuelve más complicado encontrar diferencias significativas relacionadas con las habilidades socioemocionales. La única variable que tuvo diferencias significativas para la población del EXANI-III fue la del compromiso académico, lo cual coincide con el modelo teórico utilizado, ya que se propone como la única variable que tiene un efecto directo en el desempeño académico. En ese sentido, el modelo de análisis factorial pierde consistencia en los resultados obtenidos para cada grupo.

Finalmente, se recomienda utilizar el modelo de crédito parcial para la construcción de los puntajes de las escalas. Con la finalidad de que estos puntajes se vuelvan más ágiles de generar y se fomente su uso, se realizaron estos análisis de datos específicos lo que permitirá una más fácil interpretación de los reportes de resultados.

Con estos resultados se propone construir un modelo único para la medición de habilidades socioemocionales a partir de los cuestionarios de contexto que aplica el Centro, considerando a las diferentes poblaciones que los responden. Aunque actualmente este avance está únicamente desarrollado para el EXANI-I, II y III, se espera incorporar datos derivados de la población que presenta el DOMINA-BACH y el EGEL Plus.

Lecturas recomendadas

- Dalenius, T., & Hodges Jr, J. L. (1959). Minimum variance stratification. *Journal of the American Statistical Association*, 54(285), 88-101.
- Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance--A Critical Literature Review*. Consortium on Chicago School Research. 1313 East 60th Street, Chicago, IL 60637.
- Inegi. *Nota técnica. Estratificación univariada. Censo de Población y Vivienda, 2010. SINCE. Sistema para la Consulta de Información Censal*. México D.F.: Inegi, 2010.





Los reactivos de innovación

Edgar I. Andrade Muñoz

**Jefe del Departamento
de Innovación del Ceneval**

El avance informático observado en las últimas décadas ha cambiado sustancialmente la forma en que la actividad humana se lleva a cabo. Se habla ya de un *mundo digital* en el que estudiantes, profesores, profesionistas, entre otros grupos, construyen su conocimiento y resuelven problemas utilizando herramientas computacionales. La evaluación en la educación no ha sido ajena a este cambio y es previsible que en pocos años se dejen de utilizar pruebas impresas. Estamos en un periodo de transición.

El uso de la computadora como medio para administrar exámenes abre nuevas posibilidades en la evaluación para indagar lo que las personas opinan, conocen, han aprendido, han desarrollado, etcétera, al incorporar nuevos enfoques sobre medición respecto a los utilizados en las tradicionales pruebas impresas. Estos

enfoques aprovechan las funciones y características que tienen las computadoras, en cuanto a *software* y *hardware*, para interactuar con el usuario. Ejemplos de ello son la inclusión de sonido, gráficos, animación o videos; la captura de las respuestas de los examinados mediante la selección de elementos textuales o gráficos con el ratón; la posibilidad de arrastrar u ordenar elementos, escribir respuestas cortas, modificar o redactar textos, manipular figuras, interactuar con las posibles soluciones.

A los reactivos que incorporan estas capacidades de las computadoras –adicionales a las utilizadas ya en pruebas impresas– se les ha denominado reactivos de innovación, reactivos mejorados por la tecnología o reactivos alternativos.

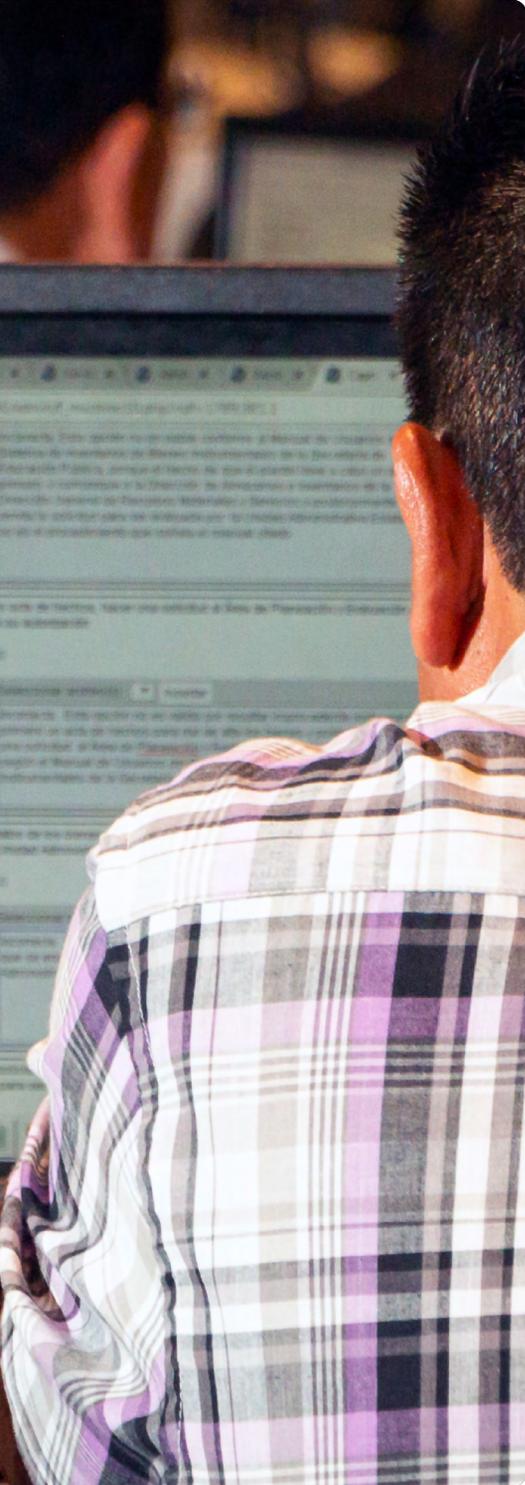
El principal beneficio de incluir reactivos de innovación es su potencial para mejorar la calidad de la medición. Si se diseñan y desarrollan cuidadosamente, pueden aumentar la cobertura del constructo de una prueba, medir procesos cognitivos importantes, ampliar la manera

en que los examinados emiten sus respuestas o recopilar una gama más amplia de comportamiento de los examinados; aumentar la fidelidad a situaciones reales de la práctica profesional; incrementar la oportunidad de medir habilidades cognitivas de nivel superior; disminuir aspectos que contaminan a la medición como la adivinación y posibilitar la calificación de los procesos realizados por los examinados como parte de la respuesta, así como de sus productos. En la evaluación internacional PISA¹⁵, la inclusión de reactivos de innovación para la evaluación de 2015 permitió medir aspectos nuevos y expandir los dominios ya evaluados con anterioridad. Por ejemplo, se incorporaron tareas interactivas en ciencia, lo que permitió a los estudiantes manipular las variables en investigaciones científicas simuladas. Con la simulación de blogs fue posible valorar la capacidad de solucionar problemas de forma colaborativa.

Una ventaja del uso de las computadoras y de los reactivos de innovación es la posibilidad de obtener información más detallada no sólo sobre las respuestas de los estudiantes, sino también sobre el proceso detrás de esas respuestas, como la cantidad de tiempo que tomó completar cada tarea y el número de acciones tomadas por el examinado, lo que contextualiza el proceso de evaluación. Aunque no asociado directamente a los reactivos de innovación sino al uso de las computadoras como medios para responder pruebas, se encuentra la ventaja de generar más fácilmente versiones diferentes de examen al no tener las restricciones que una imprenta impone. Además del correspondiente ahorro de papel en pro de la ecología y de la disminución de los tiempos para dar resultados al captar las respuestas de los examinados directamente, a diferencia de las pruebas impresas en las que las respuestas se codifican en hojas de lectura óptica.



¹⁵ Programa para la Evaluación Internacional de Alumnos de la Organización para la Cooperación y el Desarrollo Económicos (OCDE).



La variedad de tipos de reactivos de innovación que existen puede ser muy grande y con el avance de la tecnología se espera que sea mayor, por lo que se han elaborado diversas clasificaciones. Algunas incluyen tipos de reactivos que nunca se han implementado o que tuvieron problemas para su inclusión en las pruebas. La propuesta de Parshall y sus colaboradores (2010¹⁶) abarca siete dimensiones:

- › Formato del reactivo
- › Acción de respuesta
- › Inclusión de medios
- › Interactividad
- › Complejidad del reactivo
- › Fidelidad del reactivo al mundo real
- › Método de puntuación

Debe tenerse en cuenta que los beneficios que potencialmente ofrecen los reactivos de innovación no están garantizados y que su implementación exitosa representa un desafío. A diferencia de los reactivos de opción múltiple, que se desarrollaron y perfeccionaron durante muchos años en una diversidad enorme de exámenes, los reactivos de innovación aún no cuentan con ese nivel de conocimiento y comprensión, por lo que existe mucho menos información psicométrica sobre ellos, lo que implica que haya menos experiencia por parte de los elaboradores de reactivos para desarrollar reactivos de alta calidad. Además, el riesgo de incorporar nuevas tecnologías es el de modificar el constructo que se está midiendo como por ejemplo al utilizar una interfaz deficiente o que los requisitos de acción sean poco claros. En las pruebas de PISA, por ejemplo, el diseño de las pantallas tuvo que ser estudiado

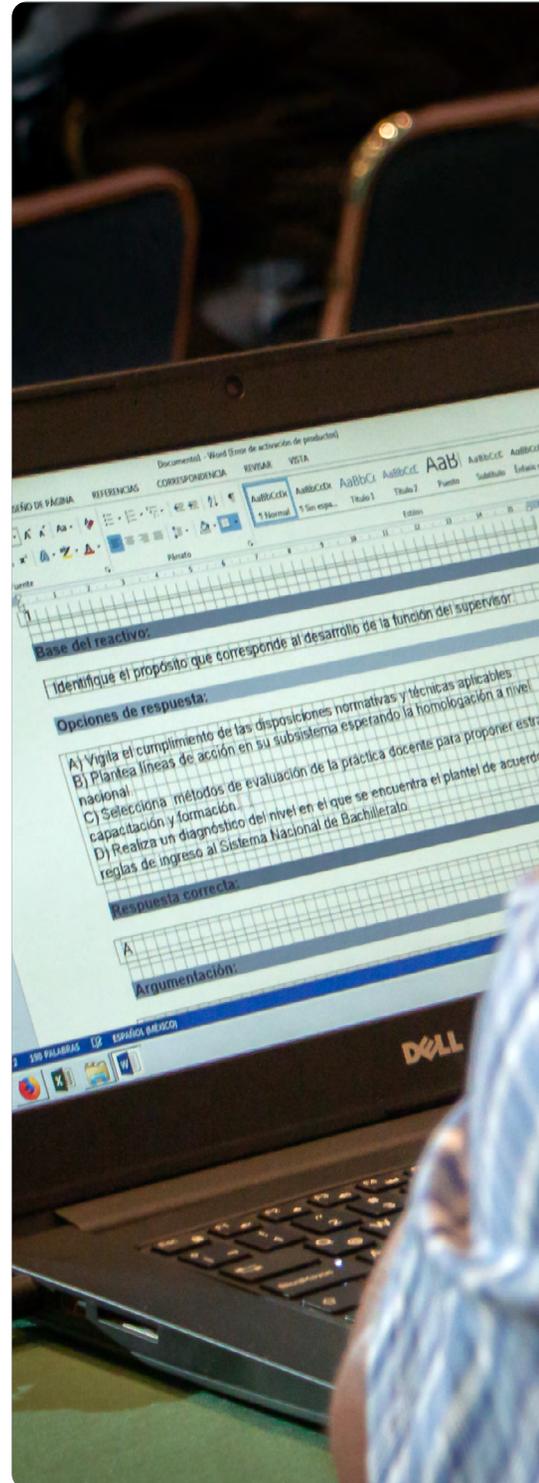
16 Parshall, C. G., Harmes, C. J., Davey, T., y Pashley, P. J. (2010). Innovative Items for Computerized Testing. En *Elements of Adaptive Testing* (pp. 215-245). Springer.

cuidadosamente. Se implementaron varios prototipos con la finalidad de que las nuevas pruebas fueran compatibles con las anteriores y se presentara la información de la mejor forma. En la fase de análisis, se tuvieron que descartar algunos reactivos de innovación con los que hubo problemas en la asignación de las puntuaciones. La experiencia de PISA nos enseña que, aun pasando todos los filtros de diseño, pueden surgir problemas no considerados con estos reactivos.

Debido a lo anterior, vale la pena realizar una evaluación cuidadosa de las necesidades del examen, un análisis exhaustivo de los desafíos inherentes a innovaciones específicas y un enfoque estructurado para el diseño de cualquier tipo de reactivo de innovación. Desafortunadamente, sin la preparación suficiente, se podría disminuir la calidad de los exámenes en vez de incrementarla. Con la finalidad de orientar el diseño de exámenes, Parshall y Harnes han sugerido un proceso de diseño de seis pasos con la convicción de que puede mejorarse la calidad y utilidad de los reactivos de innovación en cualquier programa de exámenes. Recomiendan que este proceso se lleve a cabo a través de múltiples rondas de retroalimentación y revisión antes de su administración en su fase operativa. La intención es minimizar los riesgos y maximizar los beneficios.

Los pasos son los siguientes:

1. Analizar las necesidades de construcción del programa de pruebas.
2. Seleccionar innovaciones específicas.
3. Diseñar prototipos iniciales para su discusión.
4. Refinar iterativamente los diseños de reactivo.
5. Realizar una prueba piloto de los reactivos innovadores.
6. Producir materiales finales.





Respecto al aspecto tecnológico, el *software* juega un papel fundamental y hasta limitante para el uso de los reactivos de innovación. En su inicio, los equipos informáticos de las agencias evaluadoras desarrollaron su propio *software* especializado para que soportara los reactivos tradicionales y al integrar los reactivos innovadores hubo que hacer grandes esfuerzos para su implementación. Desafortunadamente, el proceso ha sido lento y costoso. Esta situación ha ido cambiando paulatinamente al extenderse el uso de la computadora en la evaluación, de tal forma que se crearon empresas de *software* comercial enfocadas en las nuevas necesidades de utilizar la computadora como medio para administrar pruebas y que crearon plataformas para incluir ciertos tipos de reactivos innovadores, lo cual ha facilitado su implementación.

Por otra parte, se han hecho esfuerzos para ofrecer *software* de plataforma abierta creados por instituciones no lucrativas, cuyo objetivo es acelerar la innovación en la evaluación digital. Tal es el caso del Testing Assisté par Ordinateur (TAO), desarrollado por la Universidad de Luxemburgo y el Centro de Investigación Pública Henri Tudor. Si bien en estos momentos ofrece servicios básicos gratis, también ofrece herramientas más sofisticadas y asesoría de distintos niveles por una remuneración siguiendo la estrategia de negocios de muchas de las compañías de *software* en la actualidad. Una característica importante es que se ha adoptado un protocolo basado en especificaciones de interoperabilidad orientadas a la evaluación, el cual constituye el estándar para la organización de la información de reactivos y pruebas. Esto ha permitido una mayor integración y mucho mayor comunicación entre los sistemas de *software*, lo cual permite incorporar estas nuevas herramientas a los sistemas informáticos de las agencias evaluadoras. Dadas sus características de *software* abierto y de interoperabilidad, TAO ha sido utilizado en las pruebas de PISA y de otros organismos de evaluación.

Debido a que las plataformas como TAO ofrecen un espectro finito de los posibles reactivos de innovación, es necesario crear complementos para atender nuevos tipos de reactivos y así poderlos incorporar a las plataformas de administración y calificación. Si bien los adelantos en informática han hecho que el desarrollo de reactivos de innovación sea más factible, quedan desafíos importantes como por ejemplo los costos que en algunos casos pueden ser altos o la adecuación de las reglas para calificar.

Se podría inferir que al incluir más reactivos de innovación en las pruebas que reactivos de opción múltiple, se mejora sustancialmente la calidad de la evaluación en virtud de que los primeros proporcionan más información en todos los niveles de habilidad que los segundos, según algunos estudios realizados, sin embargo, también se ha reportado que los reactivos de opción múltiple brindan más información por unidad de tiempo, es decir que son más eficientes.¹⁷ Este factor –la eficiencia en tiempo– adquiere gran relevancia en los programas de evaluación en los que se le presenta al examinado una cantidad considerable de reactivos los cuáles debe resolver en un tiempo determinado. Por una parte, con los reactivos de innovación se espera obtener mejores medidas y, por otra, con los reactivos de opción múltiple se obtiene más información por unidad de tiempo que se traduce en una mayor precisión en los resultados, lo que influye en las inferencias que se emitirán a partir de los resultados. Por este motivo, en el diseño de una evaluación es necesario hacer un balance entre los factores psicométricos y prácticos acorde con los objetivos de la evaluación. Por muchos años, se han utilizado los reactivos de opción múltiple en los que se selecciona una respuesta entre varias opciones por su facilidad de calificación y administración, lo que los ha hecho idóneos para las pruebas estandarizadas y,



17 Jodoin, M. G. (2003). Measurement Efficiency of Innovative Item Formats in Computer-Based Testing. *Journal of Educational Measurement*, 40(1), 1-15.



por ello, es esperable que se sigan utilizando en el futuro en las evaluaciones en computadora.

En el Ceneval, con el propósito de que las pruebas midan de manera más válida, confiable y eficiente las competencias, conocimientos y habilidades, se han incorporado reactivos de innovación en nueve de los 39 nuevos Exámenes Generales para el Egreso de la Licenciatura (EGEL Plus), que se administrarán exclusivamente por computadora para no causar disparidades con las aplicaciones con pruebas impresas. En la medida en que el uso de la computadora sea común en las aplicaciones será posible incorporar reactivos de innovación en otros exámenes.

Su inclusión inició con una extensa investigación bibliográfica para valorar los formatos de reactivos más prometedores, las capacidades o características de las computadoras susceptibles de incorporarse en los procesos de respuesta, el tipo de interacción que se deseaba tener por parte de los usuarios, el tipo de respuestas que se esperaba, etcétera. En el aspecto tecnológico, se estudiaron y se atendieron las necesidades de *software* y *hardware* para garantizar un uso seguro y confiable a los examinados a través de su interfaz. Se atendió la compatibilidad de estos nuevos reactivos con los generados anteriormente. Asimismo, se establecieron lineamientos específicos dirigidos a los elaboradores de reactivos para garantizar que dichos reactivos de innovación cumplan con su cometido de evaluar mejor y eficientemente, lo cual se agrega a los procedimientos ya establecidos con anterioridad por el Centro para crear evaluaciones de alta calidad técnica.

Por último, debido a que la calidad de las pruebas y específicamente de los reactivos de innovación debe ser una tarea permanente, se recomienda consultar el artículo de Parshall y Harnes: *Improving the Quality of Innovative Item Types: Four Tasks for Design and Development* (2009), que expone un modelo específico para el diseño de reactivos de innovación. Respecto a las estrategias para tener un desarrollo sostenido de las pruebas con reactivos de innovación, se recomienda la lectura del artículo de Parshall y Guille titulado *Managing Ongoing Changes to the Test Agile Strategies for Continuous Innovation* (2016).

Lecturas recomendadas

- Parshall, C. G., y Harnes, C. (2008). The Design of Innovative Item Types: Targeting Constructs, Selecting Innovations, and Refining Prototypes. *CLEAR Exam Review*, XIX(2), 18-25.
- Parshall, C. G., y Harnes, C. J. (2009). Improving the Quality of Innovative Item Types: Four Tasks for Design and Development. *Journal of Applied Testing Technology*, 10(1), 1-20.
- Parshall, C. G., y Guille, R. A. (2016). Managing Ongoing Changes to the Test: Agile Strategies for Continuous Innovation. En F. Drasgow (Ed.), *Technology and testing. Improving Educational and Psychological Measurement* (pp. 1-22). Routledge.



Directorio del Ceneval

Antonio Ávila Díaz
Director General

Lilian Fátima Vidal González
Directora de los Exámenes Nacionales de Ingreso

Alejandra Zúñiga Bohigas
Directora de los Exámenes Generales para el Egreso de la Licenciatura

Ángel Andrade Rodríguez
Director de Acreditación y Certificación del Conocimiento

César Antonio Chávez Álvarez
Director de Investigación, Calidad Técnica e Innovación Académica

María del Socorro Martínez de Luna
Directora de Operación

Ricardo Hernández Muñoz
Director de Calificación

Jorge Tamayo Castroparedes
Director de Administración

Luis Vega García
Abogado General

Pedro Díaz de la Vega García
Director de Vinculación Institucional

María del Consuelo Lima Moreno
Directora de Planeación

Flavio Arturo Sánchez Garfias
Director de Tecnologías de la Información y la Comunicación

CENEVAL[®]



CENEVAL[®]



@SomosCeneval