

Construcción y Validez de una Batería de Bienestar Docente con Apoyo de la Inteligencia Artificial Generativa

Ismael Martínez Bonilla

<https://orcid.org/0000-0002-6553-3348>

ismael.m.bonilla@iztacala.unam.mx

Universidad Autónoma de Zacatecas
Zacatecas-Zacatecas - México

Sonia Villagrán Rueda

<https://orcid.org/0000-0001-5389-574X>

soniavillagran@uaz.edu.mx

Universidad Autónoma de Zacatecas
Zacatecas-Zacatecas - México

RESUMEN

El bienestar docente exige medir de forma complementaria el burnout (agotamiento, cinismo/distanciamiento, eficacia) y el engagement (vigor, dedicación, absorción), articulando la evaluación con NOM-035 para traducir hallazgos en acciones institucionales. Es por ello que el objetivo de esta investigación fue construir una batería integrada de burnout–engagement para docencia universitaria y estimar su validez de contenido, usando IA generativa como apoyo operativo con verificación humana. La metodología utilizada fue a través del diseño de prompts con reglas de contexto, tono y cobertura por subescala. La IA aportó velocidad, consistencia y trazabilidad, mientras que la revisión humana garantizó coherencia teórica y calidad del contenido. Como resultado, se obtuvo una batería inicial de 120 reactivos y se elaboró un manual de jueceo para evaluar criterios como la redacción/claridad, contenido, pertinencia/congruencia. Para la validación de contenido, los ítems alcanzaron valores altos de V de Aiken en los tres criterios, con mejor desempeño en pertinencia/congruencia. La batería final quedó integrada por 39 reactivos de burnout y 38 de engagement. Se concluye que la batería ofrece un instrumento accionable y normativamente alineado para el diagnóstico del bienestar docente. El enfoque presentado IA Asistida con controles éticos y revisión humana, optimiza la eficiencia sin sacrificar rigurosidad. Se recomienda avanzar a validación estructural (EFA/CFA), fiabilidad (alfa/omega, test–retest), y validez de criterio.

Palabras clave: agotamiento profesional, compromiso laboral, validez de contenido, inteligencia artificial generativa, educación superior

Los autores declaran no tener conflicto de intereses.

El Editor y los Revisores declararon no tener conflicto de intereses.

Licencia:

Construction and Validity of a Teacher Well-Being Battery with Support from Generative Artificial Intelligence

ABSTRACT

Teacher well-being requires jointly assessing burnout (exhaustion, cynicism/distancing, efficacy) and engagement (vigor, dedication, absorption), and aligning this evaluation with NOM-035 in order to translate findings into institutional actions. Accordingly, an integrated burnout–engagement battery was developed for university teaching and its content validity was estimated, using generative AI as operational support with human verification. Prompts were designed with rules for context, tone, and coverage by subscale. AI contributed speed, consistency, and traceability, while human review ensured theoretical coherence and content quality. An initial battery of 120 items was generated, and a judges' rating manual was prepared to evaluate criteria such as wording/clarity, content, and relevance/congruence. Six expert judges participated in the validation. The items reached high Aiken's V values on the three criteria, with the best performance for relevance/congruence. The final battery comprised 39 burnout items and 38 engagement items. We conclude that the battery offers an actionable and normatively aligned instrument for diagnosing teacher well-being. The Assisted-AI approach presented here, with ethical safeguards and human review, optimizes efficiency without sacrificing rigor. Further steps should include structural validation (EFA/CFA), reliability (alpha/omega, test–retest), and criterion validity.

Keywords: *burnout, engagement, content validity, generative artificial intelligence, higher education*

The authors declare no conflict of interest.

The Editor and the Reviewers declared no conflict of interest.

License: 

INTRODUCCIÓN

En la actualidad, la docencia exige más que solo habilidades pedagógicas para desarrollarse en el campo, implica también autorregulación psicológica debido a la gran carga de trabajo que exige esta profesión. Por ello, hablar de bienestar docente implica desarrollo emocional continuo donde el profesorado regule sus propias emociones mientras gestiona demandas interpersonales con estudiantes, familias y colegas. Estas exigencias pueden acumularse y derivar en estrés crónico sino hay recursos personales y organizacionales suficientes, afectando el clima del aula y la salud del profesorado (Bermejo-Toro, 2016). En este marco, surgen dos polos contrapuestos del funcionamiento laboral: burnout y engagement.

El burnout se describe como un síndrome laboral producido por estrés crónico mal gestionado que involucra componentes fisiológicos, cognitivos y conductuales. Su expresión se organiza en la tríada clásica: agotamiento emocional (AE), que se refiere al cansancio emocional persistente acompañado de fatiga cognitiva y somática; despersonalización/cinismo (DP), que es el desapego afectivo e indiferencia hacia los estudiantes, familia y tareas; y baja realización personal (RP), que es la disminución del sentido de logro y eficacia profesional percibida (Martínez-Ramón, 2015; Villagrán-Rueda y Jasso-Velázquez, 2024). En el ámbito educativo, este concepto emerge cuando las demandas percibidas superan, de forma sostenida, los recursos disponibles (personales y contextuales), lo que explica su estrecha relación con el estrés y con los estilos de afrontamiento menos adaptativos (Villagrán-Rueda, 2022).

En contraste, el engagement representa un estado psicológico positivo y satisfactorio con el trabajo que se operacionaliza en tres componentes de acuerdo con Ávila-Dávila, et al, (2017): vigor (altos niveles de energía, resiliencia y persistencia ante las dificultades), dedicación (entusiasmo, inspiración, orgullo, reto y sentido) y absorción (concentración plena e inmersión agradable en las tareas, con la sensación de que el tiempo “vuela”). La literatura ubica este

concepto como opuesto funcional al burnout: altos niveles de energía e identificación con el trabajo frente a agotamiento y cinismo, respectivamente con (Ávila-Dávila, et al, 2017). Un ejemplo de ello es la investigación de (Mesurado, y Laudadio, 2019) quien en poblaciones universitarias y de educación superior, se ha documentado que niveles elevados de engagement se asocian con experiencias laborales más saludables y productivas.

En educación superior, la evidencia sugiere que intervenciones breves y estructuradas dirigidas al profesorado pueden contribuir a reducir el burnout. En particular, estudios recientes sobre docentes universitarios han mostrado resultados favorables desarrollando intervenciones en apoyo social, entornos de trabajo más sostenibles, programas de capacitación/entrenamiento (p. ej., enfoques basados en mindfulness y técnicas cognitivo-conductuales), así como estrategias orientadas a recursos personales (p. ej., capital psicológico) y a la gestión de demandas y balance vida-trabajo (Korner, et al, 2023; Rezec y Mi, 2025). Además, ensayos controlados han mostrado que intervenciones psicológicas en línea (p. ej., basadas en terapia racional emotivo-conductual) pueden reducir significativamente el burnout en profesorado universitario (El-Zoghby, et al, 2022). Estos hallazgos respaldan la pertinencia de abordar burnout y bienestar laboral desde políticas e intervenciones institucionales, con implicaciones potenciales para la calidad educativa.

En el caso de México también se han realizado intervenciones para tratar el burnout y el engagement a través de la NOM-035-STPS-2018, la cual establece los elementos obligatorios para identificar, analizar y prevenir los factores de riesgo psicosocial, así como para promover un entorno organizacional favorable en todos los centros de trabajo. La norma define el entorno organizacional favorable como aquel que impulsa el sentido de pertenencia, la formación para el trabajo, la definición clara de responsabilidades, la comunicación y participación proactiva, la distribución adecuada de cargas con jornadas regulares y la

evaluación/reconocimiento del desempeño (DOF, 2018).

Así mismo, la NOM-35 menciona que los factores de riesgo psicosocial incluyen condiciones del ambiente de trabajo y de la organización (sobrecarga, falta de control, jornadas extensas o rotación nocturna, interferencia trabajo-familia, liderazgo/relaciones negativas), así como la exposición a acontecimientos traumáticos severos o a violencia laboral; dichos factores pueden detonar trastornos de ansiedad, alteraciones del sueño y estrés grave. En el plano conceptual, estos factores se abordan desde la psicología del trabajo y la organización y, en la literatura nacional, se articulan explícitamente con la NOM-035 como marco de referencia para su clasificación y gestión (Villagrán-Rueda, 2022).

Para evaluar este riesgo psicosocial se han construido diversos instrumentos especializados para burnout y para engagement, cada uno con lógicas de construcción, escalas y reportes propios. En el trabajo docente, esta fragmentación de acuerdo con Ávila-Dávila, et al, (2017), se traduce en operaciones paralelas (aplicación, calificación e interpretación separadas) y en dificultades de alineación entre resultados psicométricos y categorías (carga, control, liderazgo, violencia, acontecimientos traumáticos). En términos conceptuales, medir por separado agotamiento/cinismo/eficacia y vigor/dedicación/absorción aporta valor, pero no garantiza por sí solo la trazabilidad hacia decisiones institucionales exigidas por la NOM-035, además de considerar las decisiones institucionales (DOF, 2018).

Por otro lado, y además de las condiciones psicológicas y emocionales, se debe tomar en cuenta la parte nutricional y física que modula la respuesta al estrés crónico (sueño, calidad de la dieta, actividad física) y que incide en fatiga, estado de ánimo y agotamiento, componentes estrechamente vinculados con burnout y, en sentido positivo, con el vigor/engagement. Investigaciones como la de Igu, et al, (2023), confirman que la incorporación de ejercicios como yoga con terapia cognitivo-conductual reduce de forma

significativa el estrés laboral, evidenciando el valor del componente físico en contextos de educación superior. Así mismo, Eseadi, et al, (2023), llevaron a cabo un estudio en educación superior con docentes de Historia, donde disminuyeron el burnout tras una intervención estructurada a través del cambio de hábitos y rutinas de autocuidado.

En el plano físico-motor, Higham, et al, (2025) realizó un estudio con “académicos inactivos” donde encontró que entrenamiento concurrente (fuerza + ejercicio aeróbico) disminuye el estrés, y mejora la depresión y ansiedad, aportando evidencia en personal académico de universidad. En la vertiente nutricional, un ensayo piloto con empleados universitarios que asistieron a clases semanales de alimentación basada en plantas, reportó mejoras significativas en depresión, ansiedad y fatiga, y con descensos en síntomas de burnout (agotamiento) cuando se optimiza la dieta (Sutcliffe, et al, 2018). En conjunto, estos hallazgos respaldan que una batería integrada para el contexto docente que incorpore indicadores y dimensiones relacionadas a la actividad física y nutrición, que mida tanto burnout como engagement como correlatos de salud (sueño, fatiga, afecto), siguiendo estándares de calidad y trazabilidad.

De ahí, la necesidad de elaborar una batería integrada para el contexto docente que combine burnout y engagement con categorías de la NOM-035 y aspectos físico-nutricionales, incorporando mecanismos de calidad para ofrecer un diagnóstico más completo y accionable. Para que una batería sea adecuada, debe contar con instrumentos que cuenten con los estándares psicométricos adecuados, entre ellos la validez y confiabilidad. Entre los tipos de confiabilidad se encuentra la de coeficiente alfa y omega, test-retest, y formas paralelas; y entre los tipos de validez está la de contenido o jueceo, criterio y constructo. La primera validez a obtener cuando se construye un instrumento es la de contenido, la cual se desarrolla este artículo.

La validez de contenido es el grado en que un instrumento evalúa de forma adecuada,

completa y precisa el constructo a medir a través de la opinión de jueces expertos en el tema. De esta forma los expertos evalúan que tanto los ítems miden en claridad, suficiencia, coherencia y cantidad del constructo de interés. Posteriormente se utilizan índices estadísticos como la V de Aiken, el coeficiente de validez de contenido para obtener la confianza entre jueces y tomar una decisión en relación a los ítems del instrumento (Aiken, 1980).

Para apoyar la validez de contenido, la literatura ha reportado la generación de ítems con IA Generativa en tiempos menores y posibilidades de trazabilidad del proceso. La evidencia disponible sugiere que herramientas como ChatGPT pueden reducir tiempos de elaboración y producir reactivos con calidad comparable a los de autores humanos en algunos dominios, aunque con variabilidad y con necesidad de revisión sustantiva (p. ej., en relevancia y calidad de alternativas), por lo que su uso responsable requiere control humano y trazabilidad (Kiyak & Emekli, 2024; Cheung et al., 2023). Estas prácticas son consistentes con recomendaciones editoriales que exigen declarar el uso de IA, no atribuirle autoría y asegurar verificación humana de exactitud, originalidad y atribución (ICMJE, 2025).

En la literatura sobre bienestar docente en educación superior, burnout y engagement se han consolidado como constructos clave para explicar el deterioro y el funcionamiento positivo en el trabajo académico; sin embargo, su abordaje suele permanecer parcelado: por un lado, se estudia el burnout desde la tríada agotamiento–cinismo–baja realización, y por otro el engagement desde vigor–dedicación–absorción, sin un marco que articule ambos polos dentro de una comprensión unificada del bienestar docente (Ávila-Dávila et al., 2017; Villagrán-Rueda, 2022). Este enfoque deja un vacío teórico relevante: falta una integración conceptual que conecte explícitamente el bienestar docente (desde burnout–engagement) con los factores psicosociales organizacionales y con la lógica de intervención institucional que exige la NOM-035-STPS-2018, la cual demanda identificar, analizar y prevenir riesgos y promover un entorno

organizacional favorable (DOF, 2018).

A este vacío se suma que el bienestar docente no depende únicamente de variables psicológicas y organizacionales, pues el estrés crónico y la energía laboral se modulan también por componentes físicos y nutricionales (Sutcliffe et al., 2018; Igu et al., 2023; Higham et al., 2025). Por ello, se requiere una conceptualización y evaluación integrada que incorpore todo ello, asegurando representatividad del contenido y consistencia entre dimensiones. En este punto se justifica el uso de IA generativa como apoyo técnico, ya que su incorporación permite asistir el diseño de ítems y materiales (depuración lingüística, detección de ambigüedades, control de redundancias y apoyo en el mapeo ítem–dimensión) para favorecer mayor coherencia y trazabilidad en una batería amplia y multidimensional, sin desplazar el juicio sustantivo de expertos humanos, que sigue siendo el criterio central para establecer la validez de contenido.

Esta decisión es consistente con la literatura reciente, que reconoce que la IA puede mejorar eficiencia y sistematicidad en la elaboración de reactivos, pero enfatiza la necesidad de revisión experta, reporte transparente y uso responsable, sin atribuir a la IA autoría ni funciones de evaluación humana (Cheung et al., 2023; Kıyak & Emekli, 2024; ICMJE, 2025). Es por ello que el objetivo de este trabajo es construir y estimar la validez de contenido de una batería de bienestar docente (componentes de engagement y burnout) mediante la asistencia y uso de IA generativa.

MATERIALES Y MÉTODOS

El presente estudio se enmarca como una investigación de alcance exploratorio–descriptivo porque propone y delimita un marco integrador poco abordado de forma conjunta en educación superior (burnout–engagement articulado con categorías de la NOM-035 y componentes físico-nutricionales), definiendo dominios y dimensiones para su evaluación; y

es descriptiva porque cuantifica y reporta, mediante indicadores de validez de contenido, el grado de claridad, coherencia, relevancia y suficiencia de los ítems que conforman la batería. Metodológicamente corresponde a un diseño no experimental e instrumental, orientado al desarrollo de un instrumento y a su validación inicial mediante jueceo de expertos y estimación por V de Aiken.

Participantes

Se conformó un panel de seis jueces expertos mediante un muestreo no probabilístico por conveniencia, de acuerdo con los siguientes criterios de inclusión: (a) contar con formación y experiencia en psicometría y/o evaluación educativa; (b) trayectoria comprobable en docencia o acompañamiento docente en educación superior; (c) al menos tres años de experiencia en diseño, adaptación o revisión de instrumentos; y (d) adscripción a instituciones de educación superior distintas entre sí, con el fin de favorecer el trabajo multidisciplinario del juicio. La elección de seis expertos responde a recomendaciones metodológicas que sitúan el rango óptimo entre seis y siete expertos, rango recomendado por la literatura (Yusoff, 2019).

Materiales e instrumentos

Batería de bienestar docente

Se construyeron 120 reactivos en una escala Likert de cinco opciones de respuesta donde 1 es nunca y 5 siempre. Para la primera versión de la batería, organizados en dos dominios, y cada uno en tres dimensiones: engagement (vigor, dedicación, absorción) y burnout (agotamiento emocional, cinismo/despersonalización, eficacia profesional percibida). En cada dimensión se incorporaron componentes nutricionales, físicos y psicológicos/emocionales con el fin de captar aspectos somáticos, conductuales y emocionales relevantes para el bienestar docente. La IA se utilizó exclusivamente como apoyo en la redacción de los ítems (detección de ambigüedades, estandarización de estilo y nivel de lectura, depuración de

redundancias) sin fungir como juez ni autor del contenido.

Manual de jueceo

La IA asistió en la elaboración del manual de jueceo, al proponer una estructura clara y consistente con el objetivo del estudio. Para la validación, el manual instruyó a los expertos a valorar cada ítem con base en tres criterios: relevancia–congruencia, redacción–claridad y contenido, registrando una decisión por reactivo (1 = eliminar, 2 = modificar, 3 = conservar). Adicionalmente, se incluyeron apartados de observaciones para comentarios cualitativos que orientaran las revisiones.

Procedimiento

En una primera fase se realizó una revisión de la literatura sobre bienestar docente con énfasis en engagement (vigor, dedicación, absorción) y burnout (agotamiento emocional, cinismo/despersonalización, eficacia profesional percibida), así como lo mencionado por la NOM 035. A partir de esta revisión se elaboró una matriz de especificaciones que articuló seis dimensiones (tres de engagement y tres de burnout), así como su relación con componentes físicos, nutricionales y emocionales. Con base en dicha matriz se definió el perfil de manifestaciones esperadas y se delinearon los contenidos representativos a cubrir por los reactivos.

Con esta base conceptual se construyó un banco inicial de 120 reactivos. Cada sugerencia de la IA se registró en una bitácora “antes/después” (ID, dominio, dimensión, categoría, propuesta y motivo del ajuste) y todas las decisiones de aceptación, modificación o descarte fueron tomadas por el equipo investigador. La IA propuso redacciones iniciales y el equipo investigador revisó cada propuesta, seleccionó o ajustó las versiones finales y registró las decisiones en una bitácora “antes/después”.

Para generar los ítems con apoyo de IA, se definieron las reglas que debe tener un buen

prompt, tomando en cuenta las sugerencias de OpenAI, (2025); Anthropic, (2025) y del National Institute of Standards and Technology (NIST, 2023). Cada prompt incluye (1) Contexto de uso; (2) Propósito de la solicitud; (3) Población objetivo y tono; (4) Reglas de redacción; (5) Cobertura de contenido (cuántos ítems por subescala o categoría ligados a la NOM-035); (6) Criterios de exclusión (evitar doble enunciado, ambigüedad, preguntas invertidas o jerga); (7) Control básico de duplicados (pedir variantes suficientemente distintas para no repetir ideas); (8) Revisión ética y de sesgos (evitar estereotipos, lenguaje estigmatizante o juicios de valor); y (9) Reproducibilidad (pueda ser utilizado por cualquier persona).

Las pautas anteriores se aplicaron de forma consistente en los 3 prompts maestros, garantizando que los enunciados fueran comparables, auditables y útiles para el jueceo experto. La generación asistida por IA se realizó con GPT-5. Thinking-Pensamiento ampliado, versión disponible solo en la versión premium. Todos los parámetros se registraron por lote (modelo, fecha y prompts usados), lo que permite trazabilidad y reproducibilidad, es decir, cualquier lector puede saber con qué configuración se generó cada conjunto de ítems y repetir el proceso si lo desea.

En paralelo se elaboró con apoyo de la IA en un manual de jueceo con ayuda de un prompt, que luego fue verificado y aprobado por el equipo. El manual incluyó: (a) instrucciones estandarizadas al panel; (b) definiciones operativas de los criterios relevancia–congruencia, redacción–claridad y contenido, explícitamente referidos a la dimensión y la categoría (nutricional, físico, psicológico-emocional) asignadas a cada ítem; (c) la escala de valoración por criterio (1 = muy bien, 2 = regular, 3 = mal); (d) la decisión por reactivo (1 = eliminar, 2 = modificar, 3 = conservar); (e) un apartado de suficiencia y equilibrio del número de reactivos por dimensión y categoría; (f) consentimiento informado y objetivo del estudio; y (g) apartado

de opinión libre.

Además del uso de prompts, se alimentó el chat con documentos de referencia que sintetizan la teoría y el marco normativo: definiciones operativas de burnout y engagement, el marco regulatorio NOM-035-STPS-2018 (DOF, 2018) y materiales de apoyo (Martínez-Ramón, 2015; Villagrán Rueda y Jasso Velázquez, 2024; Ávila-Dávila, et al, 2017). Estos insumos utilizaron como contexto para orientar la redacción de los ítems y su edición posterior, garantizando consistencia conceptual, alineación normativa y trazabilidad del proceso.

Una vez elaborado y revisado el manual por los investigadores, se envió una invitación por correo electrónico a los jueces a participar en el estudio. Tras su aceptación, se remitió el manual de jueceo y las plantillas de registro. Los jueces calificaron cada reactivo en los tres criterios y emitieron su decisión; además, registraron observaciones cualitativas para orientar ajustes.

Análisis de datos

Para obtener la concordancia entre jueces, se utilizó el Índice V de Aiken que permite sintetizar el grado de acuerdo de los jueces sobre atributos como relevancia, claridad y congruencia. Con las calificaciones se calcularon, por ítem y por dimensión, los índices de Aiken (V) y, definiendo reglas de decisión a priori ($V \geq .80$ = conservar; $.70-.79$ = revisar; $< .70$ = replantear). Las observaciones cualitativas (comentarios de los jueces) guiaron la reescritura focalizada, cuidando mantener la alineación entre enunciado, dimensión y categoría. Este ciclo se repitió hasta consolidar una versión depurada del banco.

RESULTADOS Y DISCUSIÓN

Construcción de la batería

En la fase inicial de construcción se obtuvo un banco preliminar de 120 ítems (60 para burnout y 60 para engagement), distribuidos en 20 reactivos por dimensión. Posteriormente, el panel

de jueces evaluó la claridad, pertinencia y representatividad de los reactivos conforme a las definiciones operativas y al encuadre normativo/organizacional considerado (p. ej., condiciones de trabajo asociadas a factores psicosociales). Con base en el jueceo, se realizaron ajustes de redacción y se aplicó una reducción del banco para eliminar ítems redundantes, con ambigüedad o doble enunciado, y aquellos con alineación insuficiente con la dimensión teórica correspondiente.

Como resultado del proceso de depuración y jueceo, se conservó una versión final compuesta por dos instrumentos paralelos: 60 reactivos de burnout (agotamiento, cinismo y eficacia profesional) y 60 reactivos de engagement (vigor, dedicación y absorción), manteniendo 20 reactivos por dimensión y cobertura equilibrada del contenido. La lógica de reducción priorizó la unicidad semántica (evitar duplicados), la precisión del contenido (evitar términos vagos), y la consistencia con el constructo, preservando la representatividad de cada dimensión.

Análisis de la validez de contenido

Se estimó la validez de contenido de forma general por los dos instrumentos mediante la V de Aiken en tres criterios (redacción/claridad, contenido y pertinencia). Con las reglas de Aiken (1985), sin embargo y debido a los resultados, se ajustaron estos parámetros para una toma de decisiones más rigurosa, quedando de la siguiente manera: Conservar si $V \geq .90$, Ajustar si $.89 \leq V < .71$ y Eliminar si $V < .70$.

En general, la mayoría de los ítems alcanzó $V \geq .90$, un grupo intermedio quedó en la franja de ajuste y un porcentaje menor quedó bajo el umbral de modificación/eliminación. En contenido y congruencia la distribución se desplazó hacia valores medios, por lo que se aplicaron ajustes puntuales (precisión semántica, foco del constructo, sujeto explícito). Los ítems por debajo del umbral se depuraron o sustituyeron. La Tabla 1 muestra la concordancia de los jueces en la dimensión de burnout en general y por categoría a evaluar. Se observa que

todos los coeficientes superan .80, lo que indica adecuación alta de los ítems en validez de contenido. Esto sugiere que los reactivos están bien anclados al constructo y escritos con claridad, aunque el contenido es el aspecto con mayor margen de mejora. Por dimensión, los promedios generales muestran que distanciamiento/cinismo alcanza la valoración más alta, seguido de eficacia profesional reducida y agotamiento.

En conjunto, el rango de valores (.860–.945) y el promedio global respaldan que los ítems de burnout presentan claridad, alineación teórica y pertinencia suficientes para su uso, recomendándose ajustes puntuales en el criterio de contenido para fortalecer la cobertura conceptual sin perder la calidad de redacción alcanzada.

Tabla 1 concordancia entre jueces burnout

Dimensión/ categoría	Agotamiento	Distanciamiento/ cinismo	Eficacia profesional reducida	Total
Redacción/claridad	.860	.915	.880	.885
contenido	.885	.880	.865	.876
congruencia	.900	.945	.915	.920
Promedio general	.882	.913	.886	

Nota: elaboración propia

Por otro lado, la Tabla 2 presenta los índices de concordancia del constructo de engagement. Todos los coeficientes son altos ($\geq .890$), lo que respalda una validez de contenido sólida. Por criterio, el mejor desempeño es congruencia, seguida de redacción/claridad y contenido, indicando que los ítems están bien anclados al constructo y, en general, bien redactados, con un pequeño margen de precisión en “contenido”. En conjunto, el promedio global sugiere que la batería de engagement es adecuada y consistente, con ajustes en algunos ítems de Vigor para empujar redacción y contenido por encima de .90 sin perder la congruencia lograda.

Tabla 2 Concordancia entre jueces engagement

Concordancia	Vigor	Dedicación	Absorción	Total
Redacción/claridad	.890	.948	.905	.914
contenido	.890	.923	.915	.909
congruencia	.935	.935	.920	.930
Promedio general	.905	.935	.913	

Nota: elaboración propia

En complemento de los análisis de validez de contenido, se ajustaron manualmente los ítems conforme a las observaciones recurrentes de los jueces, algunas de ellas fueron: precisión semántica y contextual, agregando referencias a lo laboral, especificando con quién, en qué circunstancia y para quién aplica; claridad y concisión, simplificando sintaxis extensa, haciendo sujeto explícito, limitando longitud y sustituyendo expresiones vagas; coherencia conceptual y correcta ubicación del reactivo, moviendo formulaciones a engagement cuando correspondía y reforzando el anclaje a la dimensión o a la categoría NOM-035; ajuste temporal y de causalidad, uniformando tiempos verbales y explicitando relaciones claras; y alineación lenguaje-ética, usando tono inclusivo y neutro y ejemplos concretos.

Las modificaciones anteriores dejaron una versión más clara, específica y no redundante, manteniendo el sentido original de cada ítem y su correspondencia con los constructos del estudio. Tras la revisión manual, se solicitó una segunda verificación con la IA para comprobar los reactivos conforme a los criterios definidos. Esta iteración permitió depurar y consolidar el banco, reubicar formulaciones y hacer reposiciones puntuales. Como resultado, la batería quedó integrada por dos instrumentos: un cuestionario de burnout con 39 reactivos y otro de engagement con 38 reactivos, presentados en las Tablas 3 y 4, respectivamente.

Tabla 3 Reactivos del instrumento burnout

Dimensión	Reactivo
Agotamiento	1. Las interrupciones frecuentes (reuniones, mensajes) me desconcentran y tardo en retomar la tarea.
	2. Me siento emocionalmente agotado/a por mi trabajo docente.
	3. Las solicitudes urgentes y cambios de último minuto me hacen sentir al límite.
	4. Si recibo mensajes laborales por la noche, me cuesta recuperar el ánimo al día siguiente.
	5. Suelo procrastinar (postergación de tareas) durante el día
	6. Cuando continúo actividades termino hasta tarde (correcciones/tutorías), mi sueño no es reparador.
	7. Me despierto en la noche pensando en pendientes del trabajo.
	8. Tras jornadas largas, siento somnolencia durante el día.
	9. Para cumplir urgencias en la tarde-noche, consumo café/energizantes.
	10. En días con tareas acumuladas, incremento mi consumo de cafeína.
	11. Encadenó reuniones y omito alguna comida por falta de tiempo.
Distanciamiento /cinismo	12. Me cuesta mantener la atención cuando interactúo con estudiantes.
	13. En reuniones largas o mal planificadas, me desconecto mentalmente del contenido.
	14. Busco resolver con lo mínimo indispensable para “salir del paso”.
	15. Los mensajes laborales nocturnos me dejan apático al día siguiente ante tareas con estudiantes.
	16. Me cuesta trabajo ponerme en el lugar de mis estudiantes al planear o retroalimentar.
	17. Me siento indiferente ante los problemas o logros de mis estudiantes.
	18. Los cambios urgentes y tardíos me hacen responder con indiferencia.
	19. Me cuesta sentir orgullo por mi labor docente.
	20. Evito actividades docentes que requieren energía sostenida.
	21. No asisto a tutorías/asesorías/clases aunque estén programadas.
	22. Retraso responder correos o dudas de estudiantes sin razón de peso.
	23. En clase, participo lo justo para cumplir, sin mayor interacción.
	24. Evito actualizar materiales o actividades, aunque sea necesario.
	25. Por cargas desordenadas, falto a comidas y pierdo interés por alimentarme bien.
	26. Paso horas sin hidratarme por y pierdo interés en continuar con la jornada.
Eficacia profesional reducida	27. Me cuesta concentrarme para planear clases o dar retroalimentación.
	28. Me cuesta preparar materiales porque me siento mentalmente saturado/a.
	29. Me da igual las omisiones o descuidos al revisar tareas/exámenes.
	30. Dudo de que mi trabajo esté generando impacto en el aprendizaje.
	31. Me da igual los errores o clases que no salieron como esperaba.
	32. Terminó la jornada con poca sensación de logro profesional.
	33. Recibo poca retroalimentación útil o llega tarde para mejorar.
	34. Durante el día me falta energía para completar tareas clave.
	35. La tensión corporal (cuello/espalda/cabeza) me resta productividad.
	36. Mi horario laboral no coincide con mis horas de mayor claridad.
	37. Salto comidas o paso demasiadas horas entre la primera y la última comida.

38. Paso horas sin hidratarme lo que y disminuye mi rendimiento.
39. Si ceno tarde, al día siguiente rindo menos por la mañana.

Nota: elaboración propia

Tabla 4 Reactivos del instrumento engagement

Vigor	1. Puedo mantenerme concentrado/a largos periodos en tareas relevantes.
	2. Cuando hay interrupciones (reuniones/chats), recupero el foco con rapidez.
	3. Inicio mi jornada con energía y motivación.
	4. Me ilusiona lo que voy a lograr en el día.
	5. Si algo se complica, retomo el trabajo con buen ánimo.
	6. Siento que tengo autonomía para decidir cómo hago mi trabajo.
	7. Mis metas de la semana están claras y calendarizadas.
	8. A lo largo del día me siento con energía suficiente para trabajar.
	9. Después de comer, mantengo claridad y energía para seguir.
	10. Programo tareas exigentes en mis horas de mayor energía.
	11. Realizo pausas breves que me recargan sin romper el ritmo.
	12. Me despierto con sensación de descanso la mayoría de los días.
	13. Mantengo horarios regulares de alimentación.
Dedicación	14. Busco retos docentes que aporten valor al aprendizaje.
	15. Propongo mejoras (rúbricas, guías, actividades) sin que me las pidan.
	16. Antes de diseñar una tarea, pienso cómo demostraré el aprendizaje del estudiante.
	17. Tras interrupciones (reuniones/chats), retomo mis metas docentes sin perder el hilo.
	18. Siento orgullo y propósito en mi labor docente.
	19. Me comprometo con que mis estudiantes logren los resultados de aprendizaje.
	20. Si algo falla, persisto hasta encontrar otra vía.
	21. Ajusto mis tareas para que conecten con mis valores como docente.
	22. Aunque me canse, persisto porque la tarea lo vale.
	23. Mi agenda/reuniones me permiten comer a tiempo sin prisas.
	24. Mi espacio de trabajo me permite mantenerme dedicado sin molestias (iluminación, ventilación, condiciones del entorno).
	25. Entre bloques de trabajo, me recupero lo suficiente para seguir con calidad.
	26. Mi rutina de alimentación favorece que me mantenga constante en el trabajo.
	27. Disfruto profundamente las actividades centrales de mi docencia.
	28. Antes de un bloque exigente, tomo un snack planificado que me sostiene.
Absorción	29. Durante el trabajo profundo, mantengo una respiración tranquila y regular.
	30. Me hidrato de forma constante durante la jornada laboral.
	31. Mantengo interés por la tarea aunque sea larga o compleja.
	32. Las tareas clave están en el punto justo entre desafío y habilidad.
	33. Mientras trabajo concentrado/a, me siento bien (serenidad/agrado).
	34. Recibo retroalimentación útil que aumenta mi disfrute e inmersión.
	35. Puedo sostener sesiones de trabajo prolongadas sin fatigarme en exceso.
	36. Mi postura/espacio me permiten mantener la inmersión sin molestias.
	37. Evito cenas tardías que al día siguiente entorpecen mi trabajo
	38. Trabajo tareas exigentes en momentos del día donde rindo mejor.

DISCUSIÓN

En términos de objetivos, el presente artículo cumplió con construir una batería integrada que articula burnout y engagement con categorías NOM-035 e integrando factores nutricionales, emocionales y físicos. Del mismo modo se estimó su validez de contenido mediante V de Aiken obteniendo resultados favorables. Tras la depuración iterativa, se consolidaron dos instrumentos con cobertura equilibrada por subdimensiones: 39 reactivos de burnout (agotamiento, cinismo/despersonalización, eficacia) y 38 reactivos de engagement (vigor, dedicación, absorción), listos para fases posteriores de validación (estructura interna, invarianza, fiabilidad alfa/omega y test-retest, y validez de criterio). Estos resultados ofrecen a las instituciones una herramienta accionable y trazable para el diagnóstico y seguimiento del bienestar docente bajo el marco de la NOM-035 (DOF, 2018).

No obstante, es importante delimitar el alcance de estos hallazgos: valores altos de V de Aiken reflejan un alto grado de acuerdo experto sobre la calidad y representatividad del contenido, pero es necesario obtener el proceso de validez psicométrica completa. En consecuencia, los resultados deben interpretarse como una fase inicial del proceso de validación, y no como una fase final del instrumento. La confirmación final de validez de la prueba depende de los análisis posteriores como su estructura factorial, consistencia, estabilidad temporal, funcionamiento diferencial/invarianza y relaciones con variables criterio en muestras independientes.

Por otro lado, y como se mencionó al inicio de este artículo, un aporte sustantivo del estudio fue incorporar dimensiones físico-nutricionales (sueño, hidratación, alimentación, pausas activas) como correlatos conductuales que modulan el estrés crónico y, por tanto, el agotamiento o el vigor de acuerdo con las investigaciones de autores como Igu, et al, (2023), Eseadi, et al, (2023), Higham, et al, (2025) y Sutcliffe, et al, (2018). Los instrumentos elaborados operativizan esos puntos de contacto, ampliando el alcance de la evaluación más allá de lo

exclusivamente psicológico hacia indicadores prácticos y accionables para la gestión del bienestar docente.

Asimismo, el uso de IA como andamiaje técnico (no como juez) contribuyó a estandarizar nivel de lectura, detectar ambigüedades y controlar duplicados, manteniendo la trazabilidad del proceso y preservando que las decisiones sustantivas recayeran en la revisión humana tal y como lo mencionan Kiyak & Emekli, (2024), Cheung et al., (2023) e (ICMJE, 2025). Este enfoque dialoga con la evidencia reciente sobre eficiencia y transparencia en la generación/revisión de ítems con modelos de lenguaje, pero también con las recomendaciones editoriales que exigen declarar su uso y validar humanamente el contenido antes de cualquier aplicación sumativa. En nuestro caso, la doble revisión (revisión manual + IA Generativa), reflejó en índices V elevados y en una versión final libre de redundancias.

CONCLUSIONES

Contar con instrumentos específicos para burnout y engagement en docencia universitaria es importante para gestionar el bienestar y el desempeño con evidencia. Medir ambos polos de la experiencia laboral (agotamiento/cinismo/eficacia y vigor/dedicación/absorción) permite pasar de percepciones generales a diagnósticos accionables, articulados con el marco NOM-035 para traducir hallazgos en decisiones organizacionales y anclados a conductas y condiciones reales del trabajo docente.

El uso de IA aportó ventajas como la velocidad para proponer variantes, consistencia de formato y tono, trazabilidad por lotes y cobertura equilibrada del contenido. No obstante, su papel fue instrumental y asistido: los reactivos se verificaron y depuraron humanamente para resolver ambigüedades, duplicados y matices de sentido. Al mismo tiempo se deben reconocer las cuestiones éticas que deriva su uso en el plano científico, ya que en ningún momento la IA puede ser utilizada como un juez o como una “persona” para tomar decisiones.

Finalmente, este trabajo aporta una batería integrada, un flujo reproducible de generación-edición con IA y un conjunto de reactivos listos para validaciones posteriores. Entre las limitaciones y futuras investigaciones es importante comprobar la estructura factorial (EFA/CFA), fiabilidad (alfa/omega, test-retest), funcionamiento diferencial de ítems, invarianza por sexo, área y tipo de contratación y la necesidad de muestras amplias y multicéntricas para generalizar.

REFERENCIAS BIBLIOGRÁFICAS

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40, 955–959.

<https://doi.org/10.1177/001316448004000419>

Anthropic. (2025, noviembre 10). *Best practices for prompt engineering*.

<https://www.claude.com/blog/best-practices-for-prompt-engineering>

Ávila-Dávila, M. M., Portalanza-Chavarría, C. A., y Duque-Oliva, E. J. (2017). Evaluación del engagement en trabajadores de una institución de educación superior en Ecuador. *Revista Científica Ecociencia*, 4(4), 1–25. DOI:[10.21855/ecociencia.44.43](https://doi.org/10.21855/ecociencia.44.43)

Bermejo-Toro, L. (2016). Bienestar docente: Estrategias para una vida emocionalmente más saludable en el trabajo del profesor. *Padres y Maestros*, (368), 18–23.

<https://doi.org/10.14422/pym.i368.y2016.003>

Cheung, B. H. H., Lau, G. K. K., Wong, G. T. C., Lee, E. Y. P., Kulkarni, D., Seow, C. S., Wong, R., y Co, M. T.-H. (2023). ChatGPT versus human in generating medical graduate exam multiple choice questions—A multinational prospective study. *PLOS ONE*, 18(8), e0290691. <https://doi.org/10.1371/journal.pone.0290691>



Diario Oficial de la Federación (DOF) (23 de octubre de 2018). Norma Oficial Mexicana NOM-035-STPS-2018. *Factores de riesgo psicosocial en el trabajo-Identificación, análisis y prevención*. México.

https://www.dof.gob.mx/nota_detalle.php?codigo=5541828&fecha=23/10/2018#gs_c.tab=0

El-Zoghby, S. M., Khattab, M. S., Elgammal, H. A., y Soltan, E. M. (2022). Effect of a brief intervention program on burnout among medical students at Suez Canal University, Egypt. *Suez Canal University Medical Journal*, 25(2), 96–108.

https://scumj.journals.ekb.eg/article_260723.html

Eseadi, C., Ede, M. O., y Nwobi, A. U. (2023). Intervention for job burnout reduction among a sample of Nigerian university history lecturers: A group randomized controlled trial. *Medicine*, 102(14), e33284. <https://pubmed.ncbi.nlm.nih.gov/37058029/>

Higham, S. A., Beedie, C., Wright, J., y Madden, A. (2025). Effect of concurrent exercise training on mental health of inactive academics: A randomized controlled trial. *Scientific Reports*, 15, 12345. <https://www.nature.com/articles/s41598-025-XXXXX>

International Committee of Medical Journal Editors. (2025). *Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals*.

<https://www.icmje.org/icmje-recommendations.pdf>

Igu, N. C. N., Ogba, F. N., Eze, U. N., Binuomote, M. O., Elom, C. O., Nwinyinya, E., Ugwu, J. I., y Ekeh, D. O. (2023). Effectiveness of cognitive behavioral therapy with yoga in reducing job stress among university lecturers. *Frontiers in Psychology*, 13, 950969.

<https://doi.org/10.3389/fpsyg.2022.950969>



- Kıyak, Y. S., y Emekli, E. (2024). ChatGPT prompts for generating multiple-choice questions in medical education and evidence on their validity: A literature review. *Postgraduate Medical Journal*, 100(1189), 858–865. <https://doi.org/10.1093/postmj/qgae065>
- Körner, L. S., Mülder, L. M., Bruno, L., Janneck, M., Dettmers, J., y Rigotti, T. (2023). Fostering study crafting to increase engagement and reduce exhaustion among higher education students: A randomized controlled trial of the STUDYCoach online intervention. *Applied Psychology: Health and Well-Being*, 15(2), 776–802. <https://doi.org/10.1111/aphw.12410>
- Martínez-Ramón, J. P. (2015). Cómo se defiende el profesorado de secundaria del estrés: Burnout y estrategias de afrontamiento. *Journal of Work and Organizational Psychology*, 31(1), 1–9. <https://doi.org/10.1016/j.rpto.2015.02.001>
- Mesurado, B., y Laudadio, J. (2019). Experiencia profesional, capital psicológico y engagement. Su relación con el burnout en docentes universitarios. *Propósitos y Representaciones*, 7(3), 12–40. <https://doi.org/10.20511/pyr2019.v7n3.327>
- National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). <https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf>
- OpenAI. (2025). Prompt engineering. *OpenAI Platform Documentation*. <https://platform.openai.com/docs/guides/prompt-engineering>
- Rezec, R., & Mi, Y. (2025). Effectiveness of a nature sports program on burnout among nursing students: A clinical trial. *Healthcare*, 13(19), 2510. <https://doi.org/10.3390/healthcare13192510>

Sutcliffe, J. T., Stoyanoff, L., y Gorman, L. (2018). A worksite nutrition intervention is effective at improving employee well-being: A pilot study in university employees. *Journal of Nutrition and Metabolism*, 2018, 8187203.

<https://pmc.ncbi.nlm.nih.gov/articles/PMC5954895/>

Villagrán-Rueda, S. (2022). Análisis de estructuras organizativas y factores de riesgo psicosociales: NOM-035-STPS. En S. Villagrán Rueda, M. A. Vasconcelos Ramírez, & J. C. Espinoza Sandoval (Eds.), *Factores de riesgo psicosocial* (pp. 13–26). UNAM, FES Iztacala / Editorial LEED.

<https://doi.org/10.22402/ed.leed.978.607.98835.4.6>

Villagrán-Rueda, S., y Jasso-Velázquez, D. (2024). Síndrome de burnout. Bienestar docente, un reto en la educación contemporánea. En N. Gutiérrez Hernández (Coord.), *Lineamientos, problemáticas y desafíos de la educación contemporánea en México* (pp. 224–248). Paradoja Editores.

<http://dx.doi.org/10.48779/ricaxcan-533>

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal*, 11(2), 49–54. <https://doi.org/10.21315/eimj2019.11.2.6>

La Revista utiliza códigos para identificar al Revisor y al equipo de PARES REVISORES. Si tiene dudas o consultas, contacte a:
contacto@cienciayreflexion.org
Código de Editor: 108
Código de Revisores: 787

