Construcción de los instrumentos psicométricos de MentalTestLab

Mental Test Lab representa un grupo de expertos en salud mental e informática que se dedica a desarrollar instrumentos psicométricos y plataformas digitales para la evaluación psicológica y cognitiva. Estas herramientas permiten a los profesionales de la salud mental realizar evaluaciones de manera automatizada y estandarizada. La empresa se enfoca en proporcionar tecnologías avanzadas para administrar y gestionar diferentes tipos de tests, mejorando así la precisión y eficiencia en el diagnóstico y tratamiento de trastornos mentales.

MentalTestLab es una company builder financiada por Square Ventures, un fondo de inversión que aporta recursos en proyectos innovadores. Square Ventures no solo proporciona financiación, sino también apoyo en el desarrollo estratégico de la empresa, ayudando a que MentalTestLab evolucione y expanda sus servicios en el ámbito de la tecnología aplicada a la salud mental.

Construcción de los instrumentos psicométricos

La construcción de instrumentos y pruebas métricas, en el campo de la psicometría, se encuentra guiada por diferentes razones. Responder a una problemática social, fundamentar un marco teórico y funcionar como soporte de investigaciones científicas son algunas de ellas. La problemática a la cual responde el surgimiento de un nuevo instrumento y la serie de decisiones posteriores de quien lleva a cabo el proceso irán dando sentido a cada uno de los pasos que constituyen la producción de tecnología en el campo de la psicología.

MentalTestLab se propone dar respuesta a una serie de problemáticas presentes en el campo de la salud mental, en general, y de la evaluación psicológica, en particular. Estas problemáticas son, por un lado, la necesidad de

la consistencia y la vigencia del marco conceptual psicológico, con los modelos cuantitativos de valoración; y, por otro lado, la construcción de instrumentos de despistaje y diagnóstico para la valoración de los trastornos psiquiátricos.

Estado actual y objetivos

La revisión de la literatura actual, vinculada con los instrumentos psicométricos para evaluar trastornos mentales, pone de manifiesto la necesidad de contar con pruebas calibradas según marcos conceptuales actualizados. En relación con ello, MentalTestLab tiene como objetivo crear modelos tecnológicos de medición en salud mental, que posibiliten operacionalizar los criterios diagnósticos presentes en los manuales internacionales de psiquiatría. Entre las razones que le dan sentido a esto, se puede mencionar que, a partir del marco conceptual construido en el Manual Diagnóstico y Estadístico de los Trastornos Mentales (5° edición) (DSM-V-TR), se otorga rigor y vigencia teórica. El DSM-V-TR representa siete décadas de investigación y metaanálisis en el campo de los trastornos psiquiátricos.

El segundo motivo está estrechamente vinculado con problemas epistemológicos nodales en el campo de la psicología, como la necesidad de alcanzar estándares de mayor cientificidad en el ámbito de aplicación del saber. La clasificación es una de las tareas primeras y principales de la actividad científica.

Haynes solía decir que uno de los problemas más graves de la psicología era que había pasado excesivamente deprisa por el periodo de la clasificación y que, por tanto, los psicólogos se habían puesto a investigar sin, previamente, entender todos lo mismo sobre el manejo de las categorías (Cronbach y Meehl, 1955). Como consecuencia, se puede reconocer la falta de claridad y homogeneidad en los procesos vinculados con la categorización de los fenómenos psicológicos, las dificultades en la comunicación disciplinaria e interdisciplinaria, y por último, aunque no menos importante, la falta de acuerdo

acerca de los procesos de cura, vinculados con los tratamientos de los trastornos mentales.

La propuesta de MentalTestLab busca dar respuesta a esta situación.

Fases en la construcción de los instrumentos

Las diferentes fases llevadas a cabo para la construcción de nuestros instrumentos fueron las siguientes:

1. En una primera fase de diseño, se define la finalidad de las pruebas. El equipo de MentalTestLab llevó a cabo treinta entrevistas a profesionales del campo de la salud mental, de todo Latinoamérica, para indagar cuáles son los desafíos que experimentan en su práctica profesional, vinculados con la evaluación. De allí surgió la problemática de la necesidad de contar con herramientas estructuradas, que posibilitaran una primera aproximación al diagnóstico de ciertos trastornos en salud mental. Se marcó como limitante la complejidad de ciertos instrumentos métricos y el tiempo demandado para poder resolverlos.
2. En un segundo momento, se realizaron encuestas a ciento cincuenta profesionales de salud mental de Latinoamérica, con las que se indagaron la profesión, los años de experiencia, los instrumentos empleados en la evaluación, el tiempo destinado a ella y las limitaciones en su práctica vinculadas con la evaluación de sus pacientes.

En esta última instancia, los profesionales expresaron que, muchas veces, no emplean instrumentos métricos en los procesos de evaluación, debido a su complejidad y al tiempo requerido para hacerlo. Además, marcaron como limitante la dificultad al momento de integrar los diferentes datos recogidos durante el proceso de evaluación, los datos cuantitativos arrojados por los test y los datos cualitativos obtenidos durante las entrevistas.

Para dar respuesta a esta necesidad, MentalTestLab ofrece una serie de

instrumentos psicométricos que operacionalizan los criterios diagnósticos del DSM-V-TR, para evaluar los trastornos mentales. Por otro lado, además, un equipo de programadores trabaja en el desarrollo de soporte informático para la administración de las pruebas, el análisis de los datos y la elaboración automatizada de informes con salida de la información recogida en cada paciente. Para ello, MentalTestLab parte del marco conceptual globalmente aceptado para la clasificación de los trastornos psiquiátricos: el DSM-V-TR. Delimitada la problemática, se continúa con la elección del nivel de medición desde el cual se trabaja, la escala de medida empleada y el tipo de puntaje de los instrumentos.

El nivel de medición que se alcanza es ordinal. Ofrece la posibilidad de determinar la presencia o ausencia de la variable medida, y un escalamiento de ella. Las pruebas permiten determinar la presencia (o no) del trastorno mental en el sujeto evaluado, pero también, el nivel de severidad del cuadro en su vida.

El tipo de puntaje elegido es el puntaje de criterio. En el campo de la medición en psicología, existen diferentes tipos de puntajes, los cuales posibilitan diferentes tratamientos y conclusiones en torno a los datos arrojados por las pruebas. El puntaje de criterio es el que mejor se adecua a la finalidad para la cual fueron pensadas las pruebas de MentalTestLab: determinar la presencia o no de los criterios diagnósticos estipulados por el DSM-V-TR, para el diagnóstico de determinados trastornos mentales.

El puntaje de criterio posibilita constatar la ocurrencia de ciertas conductas, emociones y cogniciones en el sujeto evaluado. Interesa valorar qué síntomas y signos clínicos están presentes y cuál es su intensidad. Los puntajes normativos, en cambio, posibilitan la ubicación de un sujeto en relación con un grupo normativo, su comparación. Los instrumentos aquí tratados no buscan comparar al sujeto con otros, sino constatar si presenta aquellos criterios sintomáticos estipulados por los manuales clasificatorios de psiquiatría para el diagnóstico certero.

Puntaje de criterio y puntaje normativo

En la evaluación en salud mental, especialmente cuando se trata de la identificación de trastornos mentales operacionalizando los criterios del **DSM-V-TR** (Manual Diagnóstico y Estadístico de los Trastornos Mentales, Texto Revisado), la elección entre un **puntaje de criterio** y un **puntaje normativo** tiene implicaciones importantes. Cada uno de estos enfoques ofrece distintas ventajas y limitaciones, dependiendo de los objetivos de la evaluación.

Puntaje de criterio: definición - implicancias

El **puntaje de criterio** es un modo de valorar variables que se basa en determinar si una persona cumple con un conjunto específico de criterios predefinidos para un diagnóstico. En el caso del DSM-V-TR, estos criterios se centran en la presencia, severidad y duración de síntomas clínicamente significativos que son necesarios para diagnosticar un trastorno mental. Este tipo de puntaje no compara al individuo con una norma o un grupo de referencia, sino que se enfoca en si la persona cumple con los umbrales clínicos establecidos para un diagnóstico específico.

Ventajas del puntaje de criterio para la evaluación de trastornos mentales:

Enfoque en la severidad clínica y síntomas específicos: Los puntajes de criterio están diseñados para reflejar directamente la presencia o ausencia de síntomas clínicos tal como se definen en el DSM-V-TR. Este enfoque permite a los clínicos determinar si el individuo presenta suficientes síntomas relevantes para cumplir con los criterios para un diagnóstico determinado. En lugar de medir cómo se compara una persona con el promedio poblacional, este enfoque mide cuán severos o persistentes son los síntomas, lo que es elemental para el diagnóstico y tratamiento.

- **Relevancia clínica**: El puntaje de criterio se alinea con la necesidad de identificar síntomas clínicos que guíen la intervención. En el ámbito

clínico, lo que más importa es si el paciente presenta indicadores claves para un trastorno, como en el caso de la depresión mayor, donde se evalúan aspectos como la tristeza persistente, la pérdida de interés, el insomnio, entre otros, según lo especificado en el DSM-V-TR.

- **Reducción de sesgos culturales y poblacionales**: Los puntajes normativos, al comparar a una persona con una muestra de referencia, pueden verse influenciados por sesgos culturales o contextuales. Las poblaciones normativas a menudo no representan adecuadamente la diversidad cultural, lo que puede llevar a diagnósticos erróneos o inexactos si se trasladan estos valores a otros sujetos. En cambio, el puntaje de criterio se centra en los síntomas clínicamente observables y los compara con los criterios propuestos por el DSM-V-TR, lo que reduce este riesgo de sesgo y mejora la precisión diagnóstica en diferentes contextos culturales.

- **Determinación de intervención clínica**: El puntaje de criterio es especialmente útil para tomar decisiones sobre la necesidad de intervención o tratamiento. Si un paciente cumple con los criterios del DSM-V-TR para un trastorno, eso indica la necesidad de una intervención.

- **Personalización del diagnóstico**: El uso de puntajes de criterio permite una evaluación más individualizada, ya que la atención está en los síntomas y experiencias de cada paciente. Cada individuo es evaluado según los síntomas que presenta, en lugar de ser comparado con una norma estandarizada.

Puntaje normativo: definición y limitaciones

El **puntaje normativo** se refiere a un modo de valorar variables que compara el desempeño o las respuestas de una persona con las de una muestra de

referencia o "norma". En este sistema, el puntaje de un individuo se compara con los resultados de un grupo representativo de la población para determinar cómo se sitúa el individuo en relación con la media poblacional. Los puntajes normativos son comunes en pruebas de inteligencia, rendimiento académico y otras áreas donde el interés está en medir cómo un individuo se compara con un grupo más amplio.

Limitaciones del puntaje normativo para evaluar trastornos mentales:

- **Comparación con la norma en lugar de síntomas clínicos**: los puntajes normativos presentan limitaciones con respecto a valorar diagnósticos de trastornos mentales, ya que la comparación con una norma poblacional no necesariamente refleja la severidad clínica de los síntomas de un individuo. Por ejemplo, una persona con síntomas de ansiedad podría estar dentro del rango normal de la población, pero esos síntomas pueden ser clínicamente significativos y requerir tratamiento.

- **Falta de relevancia clínica directa**: un alto o bajo puntaje en relación con la población general no indica necesariamente la presencia o ausencia de un trastorno mental. Esto puede llevar a la subestimación o sobreestimación de problemas clínicos, especialmente cuando se trata de síntomas de naturaleza subjetiva o difíciles de medir en comparación con una norma general.

- **Sesgos culturales y poblacionales**: Los puntajes normativos se basan en muestras de referencia que pueden no representar adecuadamente la diversidad cultural, socioeconómica o demográfica del individuo evaluado.

- **Menos útiles para decisiones clínicas**: Un puntaje normativo puede indicar una desviación de la norma, pero no informa de manera clara sobre la necesidad de intervención clínica. Si el objetivo de la evaluación es determinar el riesgo de padecer un trastorno mental y decidir sobre un

tratamiento, los puntajes normativos no brindan la información específica necesaria para la toma de decisiones clínicas.

¿Por qué no se usan baremos en los instrumentos de MentalTestLab?

Los baremos son tablas de referencia utilizadas en puntajes normativos para comparar el rendimiento de una persona con una muestra poblacional en un atributo medido. Este enfoque es útil en pruebas de inteligencia o rendimiento académico, donde el objetivo es determinar cómo se posiciona un individuo en relación con un grupo de referencia. En relación a esto, en las evaluaciones que MentalTestLab genera a partir de la operacionalización de los criterios diagnósticos del DSM-V-TR, el uso de baremos no es apropiado debido a que los trastornos se diagnostican en función de criterios clínicos específicos, no de comparaciones normativas.

Los puntajes de criterio se enfocan en determinar si un individuo cumple con los síntomas y condiciones necesarias para un diagnóstico clínico, independientemente de cómo se compare con la norma. Los baremos no reflejan la presencia o ausencia de síntomas clínicamente significativos ni el impacto funcional que estos tienen en la vida de una persona. Además, los baremos pueden estar sujetos a sesgos culturales y poblacionales, lo que limita su utilidad en contextos clínicos donde lo importante es la identificación de síntomas claros según los estándares diagnósticos del DSM-V-TR.

En resumen, los puntajes de criterio son más adecuados para evaluar trastornos mentales dado que se enfocan en la presencia de síntomas específicos, mientras que los baremos, al depender de comparaciones normativas, no informan acerca de la gravedad clínica o la necesidad de intervención.

En la siguiente tabla, se sintetizan las características de cada tipo de puntaje, de criterio y normativo, en relación con el fin que persigue la escala.

Tabla 1. Puntaje de criterio y puntaje normativo

Característica	Puntuación de criterio en psiquiatría	Puntuación normativa
Propósito	Diagnóstico preciso de condiciones psiquiátricas, con base en criterios específicos del DSM-V-TR, para informar tratamientos personalizados.	Comparación del rendimiento o las características de un individuo con los de un grupo normativo, para determinar su posición relativa en una distribución poblacional.
Enfoque	Evaluación detallada para verificar el cumplimiento con criterios diagnósticos detallados, considerando la manifestación única de síntomas en el individuo.	Medición del rendimiento en escalas relativas, que evalúan cómo se sitúa un individuo, en comparación con una norma poblacional.
Resultados	Un diagnóstico nominal, cualitativo, que indica la presencia o ausencia de un trastorno basado en criterios establecidos.	Puntuaciones cuantitativas que indican la posición de un individuo en una curva normativa, como percentiles o desviaciones estándar.
Fundamento	Basado en la evidencia empírica y el consenso clínico reflejado en los criterios del DSM-V-TR, orientado hacia la comprensión de trastornos específicos.	Basado en la teoría de la medición y estadísticas, aplicando normas poblacionales, para evaluar variaciones individuales dentro de un contexto más amplio.
Utilización	Esencial para el diagnóstico clínico y para la planificación del tratamiento en psiquiatría, porque permite intervenciones específicas basadas en el diagnóstico.	Útil para evaluaciones educativas, de desarrollo y de investigación, en las que la comparación poblacional informa intervenciones o necesidades de apoyo.
Complementariedad con otras herramientas	A menudo, se utiliza en combinación con otros cuestionarios y evaluaciones, para obtener una visión holística de la salud mental del paciente.	Puede complementarse con evaluaciones de criterios para obtener una perspectiva más completa del bienestar individual.

Impacto en el tratamiento	Directamente vinculado con las decisiones de tratamiento, ya que los diagnósticos se basan en criterios específicos relacionados con opciones terapéuticas.	Indirectamente, influencia al tratamiento, al identificar necesidades o habilidades relativas que pueden requerir atención.
Flexibilidad en la aplicación	Adaptado para reflejar las revisiones y actualizaciones en la comprensión de trastornos mentales, así como los cambios entre diferentes ediciones del DSM.	Requiere actualización periódica de las normas para reflejar cambios demográficos o culturales en la población de referencia.

Fuente: elaboración propia.

En función de lo dicho, se propone un método estructurado para evaluar la presencia de síntomas de trastornos y su severidad, con base en los criterios diagnósticos del DSM-V-TR:

1. La evaluación comienza con la determinación de la presencia o ausencia de los criterios del DSM-V-TR. Este paso es crucial para establecer un diagnóstico confiable del trastorno.
2. Posteriormente, se introducirá un enfoque criterial, que emplea puntuaciones directas. Esta herramienta, referida como puntuación total de criterio o dimensión (PD10), se utilizará para expresar las puntuaciones de los ítems y establecer criterios para determinar la presencia de síntomas de los trastornos.
3. Además, se abordará la determinación de la severidad del trastorno, con base en la cantidad y la gravedad de los criterios presentes. Distinguir entre diferentes niveles de severidad es fundamental para la planificación del tratamiento.

Para determinar la presencia de cada criterio, se utiliza una puntuación denominada PD10. Esta se concibe como una herramienta esencial para estandarizar las evaluaciones de los criterios del DSM-V-TR en una escala de 0 a 10. Esta escala facilita la interpretación y la comparación de los resultados

obtenidos, a partir de los inventarios.

La razón por la cual se emplea este tipo de puntaje es porque es fundamental valorar las puntuaciones que el sujeto obtiene, de acuerdo con un criterio personal e individual, y no de acuerdo con datos normativos o en comparación con otros. Lo que se pretende es detectar lo que el evaluado piensa, con respecto a la ocurrencia de ciertas conductas y características personales, en cada una de las dimensiones especificadas para el diagnóstico de los trastornos, según el DSM-V-TR. Las puntuaciones directas de los sujetos aparecen expresadas en una escala del 0 al 10, que indica el grado en que el individuo ha marcado los ítems correspondientes a un determinado criterio diagnóstico.

El empleo de la PD10 se basa en varios aspectos claves:

- **Objetividad y precisión**: la PD10 permite transformar las puntuaciones acumuladas de los ítems, que pueden variar en su rango, en una escala uniforme y comparativa. Esto proporciona una base objetiva para evaluar la severidad de los síntomas, crucial para un diagnóstico preciso.
- **Facilidad de interpretación**: una escala de 0 a 10 es ampliamente reconocida y fácil de entender, de manera que facilita la interpretación de los resultados por parte de los profesionales.
- **Consistencia con las prácticas clínicas**: la normalización en una escala de 0 a 10 es coherente con muchas otras herramientas y escalas utilizadas en la psicología clínica y la psiquiatría, lo que garantiza una integración fluida con prácticas clínicas existentes.

La fórmula para detectar la PD10 de cada criterio es la siguiente:

Fórmula de Puntuación Normalizada PD10:

$$\text{PD10} = \left(\frac{\text{Puntuación Total del Criterio o Dimensión}}{\text{Máximo Puntaje Posible}} \right) \times 10$$

Validación de instrumentos

A. Juicio de experto. Validez teórica y de contenido

El juicio de experto es un método utilizado en la evaluación y desarrollo de pruebas psicométricas, con el que expertos en el campo pertinente proporcionan evaluaciones cualitativas sobre la validez y la relevancia de los ítems o afirmaciones que componen el instrumento. Este proceso implica la revisión crítica de los elementos de la prueba, por parte de individuos que tienen un conocimiento profundo y experiencia en el área específica que la prueba pretende medir.

En esta instancia, se convoca a un grupo de quince expertos en psiquiatría, psicología y metodología de evaluación en salud mental. El criterio de inclusión para conformar estos grupos es que los profesionales certifiquen diez años o más de labor clínica en salud mental con adultos. Se les explica la finalidad que persiguen nuestros instrumentos y se les pide que analicen la pertinencia de cada ítem de los inventarios, con las dimensiones propuestas por el DSM-V-TR, para el diagnóstico de los trastornos mentales trabajados. A su vez, se les solicita que, además de esto, propongan observaciones vinculadas a las consignas, giros gramaticales, estilos y formatos de las afirmaciones de cada inventario.

Como resultado de este proceso, se obtienen quince informes que contienen los juicios acerca de la relevancia, claridad y adecuación de la evaluación, en relación con los criterios diagnósticos del DSM-V-TR. Este proceso es llevado a cabo para cada uno de los instrumentos generados.

A partir de esto se calcula el índice kappa, para valorar el grado de acuerdo entre las opiniones emitidas por los expertos, con respecto al inventario en cuestión, a los reactivos que lo conforman y a su adecuación con los criterios diagnósticos propuestos por el DSM-V-TR para el trastorno evaluado.

B. Validez teórica y confiabilidad de las pruebas

La validez teórica de un instrumento psicométrico es crucial para garantizar que las mediciones que proporciona sean significativas y relevantes para los constructos que pretende evaluar. La validez teórica se centra en su alineación directa con los criterios diagnósticos establecidos por el DSM-V-TR. Esta congruencia asegura que el instrumento no solo identifique la presencia de síntomas de cada trastorno evaluado, sino que también los contextualice dentro de un marco diagnóstico reconocido globalmente.

El proceso para establecer la validez teórica de los inventarios implica una revisión minuciosa de los criterios diagnósticos del DSM-V-TR para los trastornos trabajados, asegurando que cada ítem del inventario refleje directamente estos criterios. Esto significa un análisis detallado de los síntomas claves. Además, se presta especial atención a la forma en que se estructuran y redactan las afirmaciones, con el fin de reflejar fielmente el lenguaje y los conceptos utilizados en el DSM-V-TR. Esto incluye la utilización de un lenguaje accesible y comprensible para los pacientes, que mantiene, al mismo tiempo, la precisión conceptual necesaria para una evaluación diagnóstica válida. La adecuación lingüística y conceptual de las afirmaciones es crítica para asegurar que los pacientes puedan reconocer y reportar sus experiencias, de manera que estén alineadas con los criterios diagnósticos reconocidos.

La capacidad del inventario para reflejar los criterios del manual DSM V TR fortalece su validez y utilidad en un espectro amplio de entornos clínicos y culturales.

En resumen, la validez teórica de los instrumentos se establece a través de un proceso riguroso que asegura su alineación con los criterios diagnósticos del DSM-V-TR. Este proceso incluye la revisión experta para evaluar la relevancia y adecuación de los ítems, la evaluación de la congruencia lingüística y conceptual con el manual clasificatorio. A través de este enfoque riguroso, los inventarios se posicionan como una herramienta teóricamente sólida, capaz de

identificar de manera efectiva y precisa los síntomas de los trastornos mentales operacionalizados en línea con los estándares diagnósticos internacionales.

C. Procedimiento de confiabilidad por consistencia interna: alfa de Cronbach

La confiabilidad de un instrumento psicométrico es un aspecto esencial para garantizar la precisión de las mediciones. En términos generales, se refiere a la consistencia de las puntuaciones obtenidas por las mismas personas, en diferentes momentos o con diferentes conjuntos de reactivos equivalentes. En MentalTestLab, el foco está en asegurar que las pruebas midan de forma consistente las variables psicológicas que se desean evaluar, para minimizar al máximo el error aleatorio de medición.

¿Qué es el alfa de Cronbach y para qué sirve?

El coeficiente alfa de Cronbach, desarrollado por Lee J. Cronbach, en 1951, es uno de los métodos más utilizados para evaluar la confiabilidad por consistencia interna de los instrumentos psicométricos. Este indice permite estimar el grado en que las puntuaciones de los reactivos (o ítems) de una prueba están relacionadas entre sí y contribuyen a la medición consistente de la dimensión evaluada.

El alfa de Cronbach se define como la media de todas las posibles correlaciones entre mitades de una prueba, corregidas mediante la fórmula de Spearman-Brown. Es especialmente útil en aquellos instrumentos que contienen más de dos opciones de respuesta, como los inventarios con escalas tipo Likert empleados en las pruebas de MentalTestLab.

Este coeficiente se calcula a partir de la varianza de cada uno de los ítems y la varianza total de la prueba. Su fórmula es la siguiente:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum \sigma_j^2}{\sigma^2}\right)$$

Donde:

- **α**: coeficiente alfa de Cronbach.
- **k**: número de ítems.
- **σ_j²**: varianza de cada ítem individual.
- **σ²**: varianza total de las puntuaciones del test.

El valor del coeficiente **α** oscila entre 0 y 1:

- Un valor cercano a **1** indica una alta consistencia interna, lo que sugiere que los ítems están midiendo el mismo constructo, de manera coherente.
- Un valor cercano a **0** indica baja consistencia interna, lo que sugiere que los ítems podrían estar midiendo diferentes constructos o que no están bien relacionados entre sí.

Procedimiento para obtener el alfa de Cronbach en los instrumentos de MentalTestLab

En MentalTestLab, el cálculo del Alfa de Cronbach se realiza en varias fases, para asegurar que los instrumentos sean precisos y confiables. A continuación, se describe el procedimiento utilizado:

A. **Procedimiento de creación y selección de ítems:** La selección de ítems es un paso crítico en el proceso de desarrollo de cualquier instrumento psicométrico. En MentalTestLab, este proceso comienza con la creación de ítems, con base en el marco teórico que ofrece el DSM-V-TR. A partir

de este cuerpo conceptual, se formulan las directrices que aseguran que los ítems sean relevantes y válidos para medir los constructos asociados con los trastornos mentales, lo que garantiza que la evaluación capture de manera adecuada los síntomas y características del trastorno en estudio.

B. **Revisión teórica:** el equipo de MentalTestLab inicia el proceso, realizando una revisión exhaustiva de la literatura científica relacionada con el trastorno que se desea evaluar. Esta revisión incluye tanto estudios empíricos como guías diagnósticas y manuales (DSM-V-TR), que ofrecen una descripción clara de los síntomas, comportamientos y características claves que deben considerarse, al momento de valorar los cuadros clínicos.

C. **Redacción de ítems iniciales:** a partir de la revisión teórica, se redactan ítems que capturan los aspectos fundamentales del constructo a medir. Cada ítem debe estar alineado con los criterios específicos descritos en el manual DSM-V-TR, y deben estar formulados de manera clara y directa, para que sea comprensible para los usuarios y los sujetos evaluados.
    - Se cuida que los ítems sean relevantes para capturar las dimensiones evaluadas y que aborden de manera precisa los criterios clínicos.
    - Se utilizan escalas de respuesta de tipo Likert, que permiten captar variaciones en la intensidad o frecuencia de los indicadores observados.

D. **Juicio de expertos:** el proceso de juicio de expertos en MentalTestLab es una parte fundamental del desarrollo de los instrumentos psicométricos. MentalTestLab forma parte de una comunidad de expertos de habla hispana, integrada por profesionales de distintos países como Argentina, Chile, México, Colombia, Perú, El Salvador, Cuba y España. Estos expertos no solo participan activamente en las actividades académicas organizadas, sino que también son seguidores de los cursos de

capacitación y usuarios de los instrumentos que MentalTestLab desarrolla.

E. **Participación de los expertos en el proceso de juicio:** Los profesionales que forman parte de esta comunidad han sido seleccionados para colaborar en el proceso de juicio de expertos, debido a su vasta experiencia y su formación académica en el campo de la psicología clínica, la psiquiatría y la psicometría. Estos expertos asisten a los seminarios y talleres organizados por MentalTestLab, en los que se discuten avances en la evaluación y diagnóstico de trastornos mentales, así como el desarrollo de herramientas psicométricas.

Los profesionales convocados para este proceso son cuidadosamente seleccionados bajo los siguientes criterios:

- **Experiencia clínica**: los expertos deben tener al menos **10 años de experiencia** en el ámbito de la salud mental.
- **Diversidad geográfica y cultural**: la selección de expertos de distintos países hispanohablantes garantiza que los ítems sean culturalmente adecuados y puedan ser aplicados en diferentes contextos.
- **Seguimiento de capacitación**: estos profesionales han seguido los cursos de formación y capacitación de MentalTestLab, lo que les otorga un conocimiento profundo de los modelos teóricos y metodológicos utilizados.

Procedimiento de colaboración

El proceso de colaboración se estructura en las siguientes fases:

1. **Convocatoria y entrevistas**: se convoca a los expertos para participar en entrevistas individuales y grupales. Durante estas entrevistas, se les explica en detalle el objetivo del instrumento en cuestión, los constructos que se desean medir y el marco teórico que respalda la creación de los

ítems. En estas reuniones, los expertos también comparten sus experiencias en el uso de instrumentos psicométricos y las necesidades detectadas en sus prácticas clínicas.

2. **Revisión crítica de ítems**: a cada experto se le envía un conjunto preliminar de ítems para su evaluación. Los expertos deben analizar cada ítem en función de su **claridad**, su **relevancia en la dimensión evaluada**, su **pertinencia cultural** y su **adecuación lingüística**. Al respecto, se les solicita que realicen observaciones específicas sobre:
    - **Relevancia**: si el ítem mide adecuadamente el constructo o síntoma que pretende evaluar, de acuerdo con los criterios diagnósticos del DSM-V-TR.
    - **Claridad**: si el ítem está formulado de manera clara y comprensible para la población objetivo, evitando ambigüedades o interpretaciones confusas.
    - **Adecuación cultural**: si el ítem es culturalmente adecuado para la población de habla hispana, considerando las posibles diferencias regionales en el uso del lenguaje o en la interpretación de síntomas.
3. **Retroalimentación y observaciones**: los expertos proporcionan informes detallados con sus evaluaciones, sugerencias y críticas. Estos incluyen propuestas para modificar la redacción de ítems, incorporar elementos adicionales, o eliminar aquellos que no son relevantes o adecuados. En esta fase, los expertos sugieren modificaciones en la terminología.
4. **Análisis del grado de acuerdo**: para evaluar el nivel de acuerdo entre los expertos, se utiliza el **índice kappa**, que permite medir la concordancia entre las opiniones emitidas. Un **índice kappa** superior a **0.70** se considera un indicador de un adecuado acuerdo entre los expertos, lo que refuerza la validez del instrumento. En caso de que el índice sea bajo para algunos ítems, estos se eliminan.

5. **Revisión y ajuste final de ítems**: a partir de la retroalimentación recibida, el equipo de MentalTestLab revisa y ajusta los ítems. Esto puede implicar reformular los enunciados, agregar ejemplos que faciliten la comprensión o adaptar los ítems, para que reflejen de manera más precisa los síntomas y criterios diagnósticos que se quiere medir. El objetivo es asegurar que los ítems finales sean claros, culturalmente apropiados y clínicamente relevantes.

Impacto de la colaboración de los expertos

La participación de esta red de expertos es clave para asegurar que los instrumentos psicométricos de MentalTestLab no solo sean **rigurosos desde el punto de vista métrico**, sino también, **culturalmente adecuados** para su uso en países de habla hispana. Esta colaboración garantiza que los instrumentos puedan ser aplicados en contextos diversos, respetando las particularidades idiomáticas y culturales de cada región, sin perder la validez teórica. Además, muchos de estos profesionales también son **usuarios de los instrumentos** que MentalTestLab ofrece, lo que les otorga una perspectiva única sobre su aplicación práctica. Esto asegura que los instrumentos no solo sean teóricamente sólidos, sino también, útiles y efectivos en el trabajo clínico diario.

En resumen, el proceso de juicio de expertos en MentalTestLab es un enfoque colaborativo, en el que profesionales de toda la comunidad hispanohablante contribuyen a garantizar que los instrumentos cumplan con los estándares de **validez** y **confiabilidad**, que proporcionen herramientas efectivas para el diagnóstico y evaluación de trastornos mentales.

Aplicación del instrumento

Una vez que se ha creado la versión final preliminar del instrumento, el siguiente paso en el proceso es su **aplicación a una muestra representativa**. Esta es una fase fundamental para validar la prueba, ya que la confiabilidad y validez de los ítems deben evaluarse con un grupo de participantes que refleje las características demográficas, culturales y clínicas de la población objetivo.

Procedimiento de aplicación del instrumento

- **Selección de la muestra representativa**: para asegurar la validez de los resultados, el instrumento en estudio se administra a una muestra representativa de la población a la que está destinado. Esta muestra debe incluir individuos que compartan las características relevantes del grupo poblacional en estudio, tales como:
    - Edad y contexto cultural.
    - Presencia o ausencia del trastorno que se desea evaluar (para incluir tanto pacientes diagnosticados como controles sanos).

La selección de la muestra es producto de la colaboración del grupo de expertos que forman parte de la comunidad de MentalTestLab. Cada experto seleccionado del área geográfica a la que pertenece administra el instrumento en cuestión, a un grupo de sujetos, con base en los criterios de inclusión estipulados por el equipo de MentalTestLab.

- **Condiciones estándar de administración**: para minimizar los errores de medición y garantizar que las diferencias en las puntuaciones se deban a diferencias reales en la variable que se está midiendo (y no a factores externos), el inventario se administra bajo **condiciones estandarizadas**:
    - **Ambiente controlado**: se asegura que la aplicación del test se realice en un entorno en el que los participantes no se vean afectados por distracciones o interferencias externas.

  - **Instrucciones claras**: se proporcionan instrucciones detalladas y consistentes sobre cómo responder a los ítems del inventario, asegurando que todos los participantes comprendan lo que se espera de ellos.
  - **Control del tiempo**: según la naturaleza del instrumento, se establecen límites de tiempo, cuando corresponda, o se asegura que todos los participantes tengan el mismo tiempo para responder.

- **Reporte del evaluador**: el evaluador toma nota de cualquier incidente que ocurra durante la administración del test (como dificultades en la comprensión de los ítems o problemas técnicos), para incluir esta información en el análisis de resultados.
- **Recolección de datos**: durante la administración del test, se recopilan los datos de las respuestas de los participantes. Los resultados se registran en bases de datos electrónicas, para su posterior análisis. Se toman medidas para garantizar la confidencialidad y privacidad de los datos de los participantes.
- **Análisis preliminar de consistencia y confiabilidad**: una vez recolectados los datos, se realiza un análisis preliminar de la consistencia interna del instrumento, utilizando el coeficiente de alfa de Cronbach. Este análisis permite identificar ítems problemáticos que no contribuyan de manera adecuada a la medición del constructo, y que podrían requerir ajustes o eliminación.
- **Informe de resultados y ajustes futuros**: tras la aplicación del instrumento y el análisis de resultados, se emite un informe detallado que incluye:
  - Resultados de la confiabilidad (el alfa de Cronbach por dimensión).
  - Posibles ítems que deben ser revisados o eliminados.

- **Evaluación de la confiabilidad:** se interpretan los valores obtenidos del alfa de Cronbach, de acuerdo con las siguientes pautas:
    - **0.90 o más**: excelente consistencia interna.
    - **0.80 - 0.89**: muy buena consistencia interna.
    - **0.70 - 0.79**: aceptable consistencia interna.
    - **0.60 - 0.69**: cuestionable consistencia interna.
    - **0.59 o menos**: inaceptable consistencia interna.

Conservación de Ítems con consistencia interna muy buena

En MentalTestLab, los ítems que se conservan, en la versión final de los instrumentos, son aquellos que logran **índices de alfa de Cronbach muy buenos o excelentes** (superiores a 0.80). Este enfoque asegura que cada dimensión evaluada tenga una consistencia interna de sus puntajes fiable.

A modo ilustrativo, a continuación, se muestran los valores finales del análisis de confiabilidad, basado en el **alfa de Cronbach,** para cada una de las dimensiones del Autoinforme para la Detección Temprana del Trastorno Depresivo Mayor en Adultos (ADTTD - Adultos). Los valores reflejan el coeficiente de confiabilidad de los puntajes obtenidos por los participantes en las diferentes dimensiones del inventario:

Tabla 2. Valores finales de confiabilidad en ADTTD - Adultos

Dimensión	N° de Ítems	Alfa de Cronbach (α)	N (Sujetos Evaluados)
Estado de Ánimo Depresivo	6	0.89	1050
Pérdida de Interés o Placer	6	0.88	1050
Cambios Significativos en el Peso o Apetito	6	0.87	1050
Insomnio o Hipersomnia	6	0.85	1050
Agitación o Retraso Psicomotor	6	0.86	1050
Fatiga o Pérdida de Energía	6	0.90	1050
Sentimientos de Inutilidad o Culpa Excesiva	6	0.88	1050
Capacidad Disminuida para Pensar o Concentrarse	6	0.89	1050
Pensamientos de Muerte o Suicidio	6	0.91	1050

Fuente: elaboración propia.

Interpretación de los coeficientes de confiabilidad

- **Estado de ánimo depresivo**: α = 0.89. Los puntajes obtenidos en los ítems que evalúan el estado de ánimo depresivo son altamente consistentes, lo que indica que los ítems miden de forma fiable el mismo constructo.
- **Pérdida de interés o placer**: α = 0.88. Los puntajes obtenidos en los ítems relacionados con la pérdida de interés muestran una consistencia interna elevada, lo que sugiere una alta confiabilidad en la medición de esta dimensión.
- **Cambios en el peso o apetito**: α = 0.87. Los ítems que miden los cambios en peso y apetito generan puntajes consistentes entre sí, lo que respalda la fiabilidad de los resultados en esta dimensión.

- **Insomnio o hipersomnia**: α = 0.85. Los puntajes obtenidos en los ítems sobre trastornos del sueño son consistentes, lo que asegura una medición confiable.
- **Agitación o retraso psicomotor**: α = 0.86. La consistencia de los puntajes en esta dimensión muestra que los ítems relacionados con la agitación o lentitud psicomotora miden adecuadamente el constructo.
- **Fatiga o pérdida de energía**: α = 0.90. Los puntajes en los ítems sobre fatiga son extremadamente consistentes, lo que indica una medición muy confiable de este aspecto del trastorno.
- **Sentimientos de inutilidad o culpa**: α = 0.88. Los puntajes obtenidos en esta dimensión tienen una alta consistencia, lo que garantiza que los ítems estén midiendo el mismo constructo.
- **Capacidad disminuida para pensar o concentrarse**: α = 0.89. Los puntajes obtenidos en los ítems relacionados con problemas cognitivos son consistentes, lo que asegura la fiabilidad de esta categoría.
- **Pensamientos de muerte o suicidio**: α = 0.91. Los puntajes relacionados con los ítems sobre pensamientos de muerte o suicidio son altamente consistentes, lo que es crucial para garantizar una evaluación precisa en esta dimensión tan sensible.

Estos valores reflejan la **confiabilidad** de los puntajes obtenidos en cada una de las dimensiones. Este nivel de confiabilidad asegura que los puntajes de los participantes reflejan la ocurrencia real de los síntomas y no errores de medición aleatorios.