

claveo empírico – empirical-criterion keying

Authored by
memjavad

January 22, 2026

RECOMMENDED CITATION

memjavad (2026). *claveo empírico – empirical-criterion keying*. Spanish Psychological Databases. Retrieved from <https://spanish.arabpsychology.com/?p=8509>

Clave de Criterio Empírico (Empirical-Criterion Keying)

Primary Disciplinary Field(s): Psicometría, Evaluación de la Personalidad, Psicología Clínica

1. Definición Central y Propósito

La Clave de Criterio Empírico (CCE), también conocida como método de escalamiento basado en criterios o método de clave de criterio, es una técnica fundamental y rigurosa utilizada en la construcción de instrumentos de evaluación psicológica, especialmente en el ámbito de los inventarios de personalidad. Este enfoque se distingue radicalmente de los métodos de construcción puramente racionales o factoriales, ya que la selección final de los ítems para una escala específica no se basa en su contenido aparente (validez facial) ni en su coherencia interna con una teoría preexistente, sino exclusivamente en su capacidad estadística y demostrada para diferenciar entre dos grupos definidos externamente: un **grupo criterio** y un **grupo de control**. El propósito primordial de la CCE es maximizar la validez predictiva del instrumento para un resultado o diagnóstico específico, permitiendo que la evidencia empírica dicte qué elementos son incluidos en la escala, independientemente de si su contenido parece lógicamente relacionado con el constructo que se pretende medir.

El proceso exige que el desarrollador del test identifique primero un criterio externo observable y verificable, como un diagnóstico clínico (p. ej., esquizofrenia, depresión) o un resultado conductual (p. ej., éxito laboral, riesgo de reincidencia). Una vez establecidos los grupos (pacientes diagnosticados versus población general sana), se administra un vasto conjunto inicial de ítems a ambos. La esencia de la CCE radica en el análisis estadístico subsiguiente: un ítem se incluye en la escala si, y solo si, la respuesta a ese ítem discrimina significativamente a los miembros del grupo criterio de los miembros del grupo de control. Por ejemplo, si el 80% del grupo de control responde "Falso" a la declaración "Me siento feliz la mayor parte del tiempo", mientras que solo el 20% del grupo criterio responde "Falso", esta diferencia es estadísticamente significativa y el ítem se "clava" (se incluye) en la dirección que lo hace distintivo para el grupo criterio. La ponderación de la respuesta (la "clave") se determina empíricamente, de ahí el nombre del método.

Este método, al priorizar la utilidad predictiva sobre la coherencia conceptual inmediata, ha demostrado ser extraordinariamente robusto en la creación de escalas que son difíciles de falsear o manipular intencionalmente por los examinados. Debido a que muchos de los ítems seleccionados mediante CCE pueden tener una baja validez facial--es decir, su conexión con el constructo medido es sutil o indirecta--los sujetos no pueden inferir fácilmente qué respuesta es la socialmente deseable o cuál corresponde a un diagnóstico particular. Esta característica confiere a los instrumentos basados en CCE una ventaja significativa en entornos clínicos o forenses donde la motivación para fingir síntomas (o la ausencia de ellos) es alta. La CCE, por lo tanto, se erige como un pilar de la psicometría que valora la eficacia estadística y la predicción concreta por

encima de la elegancia teórica.

2. Fundamentos Históricos y Desarrollo (El Caso MMPI)

La Clave de Criterio Empírico no es solo una técnica; es un movimiento histórico que surgió como una respuesta directa a las limitaciones percibidas en los métodos de evaluación psicológica predominantes a principios del siglo XX. Antes del desarrollo de la CCE, la mayoría de los inventarios de personalidad se basaban en el método racional o deductivo, donde los constructores de pruebas seleccionaban ítems basándose en su juicio subjetivo y en teorías psicológicas existentes. Esto a menudo resultaba en instrumentos con alta validez facial pero baja validez predictiva, ya que los sujetos podían manipular fácilmente sus respuestas, y la conexión entre la teoría y la conducta real no siempre se sostenía empíricamente.

El hito fundamental en el desarrollo y la popularización de la CCE fue la creación del **Inventario Multifásico de Personalidad de Minnesota (MMPI)**. Desarrollado por el psiquiatra [Starke Hathaway](#) y el psicólogo clínico J. C. McKinley en la Universidad de Minnesota en las décadas de 1930 y 1940, el MMPI fue concebido con el objetivo práctico de proporcionar una herramienta objetiva y estandarizada para el diagnóstico psiquiátrico. Hathaway y McKinley administraron más de 500 ítems a grandes grupos de pacientes psiquiátricos con diagnósticos específicos (los grupos criterio, como hipocondriasis, depresión, histeria, etc.) y a un grupo de control de visitantes "normales" del Hospital de la Universidad de Minnesota. Su metodología fue puramente inductiva: si un ítem era respondido de manera diferente por el grupo deprimido en comparación con el grupo de control, ese ítem se convertía en parte de la escala de Depresión, sin importar lo que el ítem preguntara explícitamente.

Este enfoque empírico marcó una ruptura con la tradición y estableció el estándar para la evaluación objetiva de la personalidad. El éxito del MMPI, y sus revisiones posteriores (MMPI-2, MMPI-3), reside precisamente en esta metodología, proporcionando escalas que no miden necesariamente los constructos teóricos puros, sino más bien patrones de respuesta asociados con estados clínicos concretos. La CCE demostró que la utilidad de un ítem en el contexto de la evaluación de la personalidad se define por su conexión con un resultado observable en el mundo real, no por su coherencia interna o su alineación con una escuela de pensamiento particular. El impacto del MMPI fue tan profundo que la CCE se convirtió en el método preferido para cualquier instrumento de personalidad diseñado para la predicción clínica o la clasificación diagnóstica, influyendo en el desarrollo de otros instrumentos importantes como el Inventario Psicológico de California (CPI).

3. Metodología de Construcción de Ítems

La aplicación de la Clave de Criterio Empírico sigue un proceso estadístico rigurosamente definido

que garantiza que la selección de ítems esté libre de sesgos teóricos o subjetivos. El primer paso crucial es la selección y definición precisa del **criterio externo**. Este criterio debe ser operativo, confiable y medible. Por ejemplo, si se desea crear una escala para predecir la propensión al riesgo financiero, el criterio podría ser un historial documentado de bancarrotas o inversiones fallidas. La calidad del instrumento final depende intrínsecamente de la validez y la estabilidad de este criterio de referencia.

Una vez definido el criterio, se establecen los dos grupos de contraste: el **grupo criterio**, compuesto por individuos que manifiestan claramente el rasgo o el estado de interés (p. ej., veteranos de guerra con diagnóstico de TEPT), y el **grupo de control**, compuesto por individuos representativos de la población general o de una población de referencia relevante que no manifiestan el rasgo. La representatividad y el tamaño de la muestra son esenciales; si el grupo criterio es demasiado pequeño o no es homogéneo, los resultados estadísticos serán inestables y no generalizables. Idealmente, los grupos deben ser emparejados en variables demográficas clave (edad, género, nivel educativo) para aislar el efecto del criterio de interés.

El núcleo metodológico es la fase de análisis de ítems. Se administra el banco inicial de ítems (que a menudo contiene cientos o miles de declaraciones) a ambos grupos. Para cada ítem, se realiza un análisis de contraste estadístico (típicamente una prueba [Chi-cuadrado](#) o un análisis de diferencia de medias) para determinar si la proporción de respuestas a una opción específica difiere significativamente entre el grupo criterio y el grupo de control. Solo aquellos ítems que alcanzan un umbral predefinido de significancia estadística son retenidos. La respuesta que es más característica del grupo criterio se convierte en la respuesta "clave" o puntuable. Es fundamental destacar que, en este método, un ítem puede ser retenido incluso si su correlación con otros ítems dentro de la misma escala es baja o negativa, siempre y cuando su correlación con el criterio externo sea alta. Finalmente, la escala debe ser sometida a un riguroso proceso de **validación cruzada** utilizando nuevas muestras independientes para asegurar que la capacidad predictiva observada en la muestra de desarrollo no fue producto del azar o de características idiosincrásicas de esa muestra específica.

4. Ventajas Metodológicas y Aplicaciones

La Clave de Criterio Empírico ofrece varias ventajas metodológicas que la han mantenido como una técnica viable y poderosa en la psicometría moderna. La principal fortaleza es su capacidad para maximizar la **validez predictiva**. Dado que los ítems son seleccionados precisamente por su demostrada conexión con un resultado externo relevante, el instrumento resultante es un predictor altamente eficiente de ese criterio. Esto es crucial en contextos de toma de decisiones de alto riesgo, como la selección de personal en puestos de seguridad, el diagnóstico clínico diferencial o la evaluación forense, donde la precisión predictiva supera la necesidad de una profunda comprensión teórica del constructo subyacente.

Otra ventaja distintiva, ya mencionada, es la **resistencia al falseamiento**. La naturaleza sutil y, a menudo, contraintuitiva de los ítems seleccionados empíricamente hace que sea extremadamente difícil para los examinados manipular conscientemente sus respuestas para crear una impresión específica. Por ejemplo, un ítem que parece trivial ("Me gusta leer revistas de jardinería") podría estar empíricamente correlacionado con una baja tasa de ansiedad social, y un examinado que intente parecer extrovertido no sabría cómo responder para lograr ese efecto deseado. Esta opacidad protege la integridad de la medición en situaciones donde la deseabilidad social o la simulación son amenazas serias a la validez de la prueba.

En cuanto a las aplicaciones, la CCE domina el campo de la evaluación clínica y de la personalidad. Más allá del MMPI, la técnica ha sido fundamental en la creación de escalas que miden constructos complejos y multifacéticos. Por ejemplo, en el ámbito de la psicología industrial y organizacional, la CCE se utiliza para desarrollar instrumentos de selección de personal donde el criterio externo es el desempeño laboral medido objetivamente. Al permitir que los datos guíen la selección de ítems, el método asegura que el instrumento final esté directamente sintonizado con las realidades conductuales del entorno de aplicación, evitando las trampas de la especulación teórica. La CCE proporciona una base sólida para la inferencia probabilística: si la puntuación de un individuo cae dentro del rango característico del grupo criterio, la probabilidad de que comparta el estado o rasgo definido por ese criterio es significativamente alta.

5. Limitaciones Teóricas y Críticas

A pesar de su utilidad práctica y su solidez estadística, la Clave de Criterio Empírico ha sido objeto de críticas significativas, principalmente relacionadas con su naturaleza **ateórica** y las consecuencias estructurales de su metodología. La crítica más potente es que, al centrarse únicamente en la predicción, la CCE a menudo sacrifica la **validez de constructo** y la capacidad de explicar por qué funciona la escala. Las escalas resultantes pueden ser excelentes predictores, pero no necesariamente contribuyen al avance de la teoría psicológica, ya que no se basan en un modelo coherente de la personalidad o del comportamiento. Esto puede llevar a la "falacia del jingle-jangle," donde escalas con nombres idénticos pueden medir constructos completamente diferentes, o viceversa, porque la etiqueta se basa en el criterio externo y no en una estructura conceptual interna.

Una limitación técnica inherente a la CCE es la **superposición de escalas (Scale Overlap)**. Dado que cada ítem se selecciona independientemente basándose en su capacidad para discriminar un criterio específico, es común que un mismo ítem aparezca en múltiples escalas si demuestra capacidad predictiva para varios criterios (p. ej., un ítem puede diferenciar tanto a pacientes deprimidos como a pacientes ansiosos del grupo de control). Esta redundancia metodológica reduce la independencia de las escalas, haciendo que las puntuaciones entre ellas estén altamente correlacionadas. Cuando dos escalas comparten muchos ítems, es difícil interpretar si

un perfil elevado se debe a la presencia de dos constructos distintos o simplemente a la duplicidad de la medición.

Finalmente, la CCE es altamente vulnerable a la dependencia de la muestra y al **desgaste del criterio (Criterion Drift)**. La capacidad predictiva de una escala construida mediante CCE está intrínsecamente ligada a las características específicas de los grupos criterio utilizados en su desarrollo. Si la naturaleza del criterio cambia con el tiempo --por ejemplo, si los criterios diagnósticos para la depresión evolucionan, o si los estándares culturales de lo que se considera "normal" se modifican--, la clave de puntuación original puede volverse obsoleta. Esto requiere una revisión y una **validación cruzada** constante y costosa para mantener la relevancia y la precisión del instrumento. Si el instrumento se aplica a una población que difiere significativamente en composición demográfica o cultural de la muestra original, su validez predictiva puede caer drásticamente, un problema conocido como especificidad de la muestra.

6. Implicaciones Éticas y Prácticas

El uso de instrumentos construidos mediante Clave de Criterio Empírico conlleva importantes implicaciones éticas y prácticas, especialmente en el ámbito de la salud mental y la justicia. El poder predictivo de estas escalas debe manejarse con cautela. Éticamente, el uso de una escala empírica para clasificar o etiquetar a un individuo exige que el criterio externo utilizado en la construcción de la prueba sea inequívocamente relevante y justo. Si el criterio original refleja sesgos sociales o culturales (por ejemplo, si el grupo de control histórico del MMPI no era representativo de las minorías), la aplicación de la prueba a poblaciones diversas puede perpetuar la injusticia y llevar a diagnósticos erróneos o inapropiados. Por lo tanto, la transparencia sobre los grupos de desarrollo y la realización continua de estudios de sesgo de ítems son imperativos éticos.

En la práctica clínica, la interpretación de los resultados de una escala CCE requiere una formación especializada. A diferencia de las escalas teóricas, donde una puntuación alta puede interpretarse como una manifestación intensa del constructo nominal (p. ej., "extraversión"), una puntuación alta en una escala CCE (p. ej., la escala "Paranoia" del MMPI) solo significa que el patrón de respuestas del individuo es similar al patrón de respuestas de las personas que fueron diagnosticadas con paranoia en la muestra de desarrollo. El clínico debe evitar la reificación (tratar la etiqueta de la escala como una entidad psicológica real) y utilizar la puntuación como una señal de alerta o un indicador probabilístico que debe ser confirmado mediante una evaluación clínica exhaustiva. La utilidad del instrumento es máxima cuando se utiliza como una herramienta de detección y no como la única base para el diagnóstico.

Finalmente, la CCE subraya la importancia de la **estandarización rigurosa**. El valor de la clave de criterio se basa en la comparación de las respuestas de un individuo con las normas

establecidas por los grupos de control y criterio. Cualquier desviación en la administración, puntuación o interpretación de la prueba puede invalidar las inferencias estadísticas. Los profesionales deben asegurarse de utilizar las versiones más recientes y validadas de los instrumentos CCE, reconociendo que la obsolescencia de las normas y la deriva del criterio son amenazas constantes. La responsabilidad ética recae en el psicometrista para garantizar que el poderoso motor predictivo de la CCE se utilice para el beneficio del evaluado, minimizando el riesgo de etiquetado indebido o de toma de decisiones injustificadas basadas únicamente en una puntuación empírica.

7. Lecturas Adicionales

[Minnesota Multiphasic Personality Inventory \(MMPI\)](#)

[Psicometría](#)

[Criterion Validity](#)

[Starke R. Hathaway](#)

ARABPSYCHOLOGY.COM