

# GUÍA INTRODUCTORIA AL DISEÑO CENTRADO EN EVIDENCIAS

Presidente de la República  
Iván Duque Márquez

Ministra de Educación Nacional  
María Victoria Angulo

Viceministra de Educación Superior  
Natalia Ruiz Rodgers

Publicación del Instituto Colombiano para la  
Evaluación de la Educación (Icfes)  
© Icfes, 2018.  
Todos los derechos de autor reservados.

Edición  
Juan Camilo Gómez Barrera

Elaborado por  
Luis Javier Toro Baquero  
Manuel Alejandro Amado

Directora General (E)  
María Sofía Arango Arango

Secretaria General  
María Sofía Arango Arango

Directora de Evaluación  
Natalia González Gómez

Director de Producción y Operaciones  
Giovany Babativa Márquez

Director de Tecnología  
Eliécer Vanegas

Oficina Asesora de Comunicaciones y Mercadeo  
Ilba Janeth Cárdenas Fonseca

Oficina Gestión de Proyectos de Investigación  
Luis Eduardo Jaramillo Flechas

Subdirectora de Producción de Instrumentos  
Nubia Sánchez

Subdirector de Diseño de Instrumentos  
Luis Javier Toro Baquero

Subdirectora de Estadísticas (E)  
Natalia González Gómez

Subdirectora de Análisis y Divulgación  
Silvana Godoy Mateus

Diagramación y diseño  
Diana Téllez Martínez

ISBN de la versión digital: En trámite

## ADVERTENCIA

Todo el contenido es propiedad exclusiva y reservada del Icfes y es el resultado de investigaciones y obras protegidas por la legislación nacional e internacional. No se autoriza su reproducción, utilización ni explotación a ningún tercero. Solo se autoriza su uso para fines exclusivamente académicos. Esta información no podrá ser alterada, modificada o enmendada.

# Contenido

Introducción .....	4
1. Origen y definición del DCE .....	5
2. Los estratos del DCE en el Icfes .....	7
2.1 Dominio .....	8
2.2 Afirmaciones .....	9
2.3 Evidencias .....	10
2.4 Tareas .....	12
3. Ventajas de la implementación del DCE .....	15
Referencias .....	16

# INTRODUCCIÓN



El objetivo del presente texto es presentar los rasgos generales del marco sistemático, conocido como **diseño centrado en evidencias** (DCE), empleado actualmente por el Instituto Colombiano para la Evaluación de la Educación (Icfes) en la creación, aplicación y uso de instrumentos de evaluación o pruebas<sup>1</sup>. Para ello, en primer lugar, se expone el origen del DCE y sus rasgos básicos, a partir de sus principales proponentes; en segundo lugar, se explican las capas o estratos definidos por el DCE (adoptados por el Icfes) para el diseño de las especificaciones de sus pruebas. Finalmente, se señalan las ventajas que conlleva adoptar el DCE, las cuales giran en torno, principalmente, a la validez de las inferencias que concluyen afirmaciones sustantivas sobre lo que los estudiantes saben o pueden hacer en la vida cotidiana, a partir de la puntuación que obtienen en una prueba.

<sup>1</sup> “Instrumento de evaluación” y “prueba” se tomarán aquí como conceptos similares.

# 1. ORIGEN Y DEFINICIÓN DEL DCE

Si bien los trabajos de Mislevy, Almond y Steinberg (2003) fueron pioneros en el desarrollo del DCE, de acuerdo con Almond, Steinberg, Mislevy, (2002), varios de los conceptos centrales del DCE pueden encontrarse en las obras de Toulmin (1958) y Messick (1989).

En efecto, según los planteamientos del DCE, una prueba funciona como un *argumento evidencial*, en el sentido que ya lo había planteado Toulmin (1958). Para este autor, cualquier afirmación que se quiera defender exitosamente debe estar basada en datos. Adicionalmente, debe existir una garantía, es decir, una razón, o conjunto de razones, que soporte la conexión lógica entre los datos y la afirmación que se quiere defender. Un *argumento evidencial* es, de esta manera, un argumento cuya conclusión se apoya en datos y para el cual existe una garantía.

De igual manera, para el DCE, una evaluación permite construir un argumento que posibilita defender un conjunto de afirmaciones (sobre la habilidad o rasgo latente que se quieran medir) de un estudiante a partir de datos específicos; a saber, las conductas observables del estudiante en una serie de situaciones cotidianas, conocidas como tareas. Estos datos y las afirmaciones sobre el estudiante deben estar conectados apropiadamente por suficientes garantías, de tal forma que se evidencie cómo las respuestas a las tareas dependen de la habilidad que se quiere medir. La figura 1 representa la aplicación de la noción de argumento evidencial de Toulmin a la evaluación de la habilidad de un estudiante:

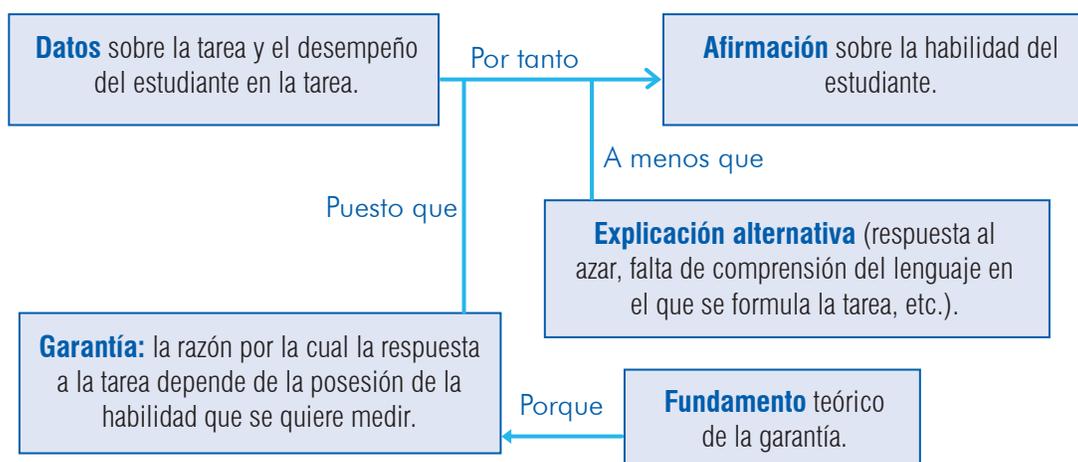


Figura 1. Argumento evidencial y evaluación

Del texto de Messick (1989), el DCE adopta la noción de validez de una prueba. En consonancia, la validez de una prueba está directamente relacionada con el grado en que la evidencia empírica y las bases teóricas soportan la adecuación y pertinencia de las inferencias y acciones hechas a partir de los resultados de la prueba. Así, el modo en que se usa la evidencia para soportar afirmaciones sobre estudiantes, que resuelven una prueba, es fundamental en el DCE.

Conforme con lo anterior, el DCE tiene como objetivos especificar la estructura del argumento evidencial de una prueba y, de esta forma, determinar que el modo en que se recoge y se interpreta la evidencia esté en relación con los propósitos que una prueba pretende cumplir (Mislevy, 2017).

En síntesis, se puede indicar que el DCE es una serie de prácticas que definen los procesos de diseño, desarrollo y uso de un instrumento de evaluación en términos de varios *estratos*, relacionados lógicamente. Estas prácticas ayudan a clarificar y asegurar la validez de las inferencias que se pueden hacer a partir de la puntuación que obtiene un estudiante en una prueba, así como determinar cómo proveer la evidencia para sustentar estas inferencias, dados los constreñimientos y limitaciones propios de la aplicación de una prueba (Mislevy y Riconscente, 2005).

## 2. LOS ESTRATOS DEL DCE EN EL ICFES

Los estratos definidos por el DCE pueden verse como los pasos en una larga cadena de razonamiento (Zieky, 2014, p. 80). Entre estos pasos se encuentran:

- El análisis y modelado del **dominio** de conocimientos, habilidades u otras destrezas (denominados a partir de ahora como CHD) de interés, que incluye una descripción de los CHD que se quieren medir (también llamada modelo de estudiante) y una especificación de la perspectiva psicológica que se asume para la evaluación.
- La especificación de las **afirmaciones** acerca de los CHD de los estudiantes que se harán a partir de los resultados de la prueba.
- La definición de las **evidencias** que se requieren para soportar las afirmaciones sobre los estudiantes.
- El desarrollo de las **tareas** que proveerán las evidencias requeridas, dadas las limitaciones de la aplicación de la prueba.
- El ensamble de las tareas en la prueba, de tal manera que la puntuación o resultados de la prueba respalden las afirmaciones con suficiente evidencia.
- La definición de las reglas que determinen la puntuación de las tareas.
- Por último, la estimación del nivel del estudiante en el dominio evaluado para realizar las afirmaciones o inferencias, definidas en b).

La figura 2 representa los diversos estratos definidos por el DCE:

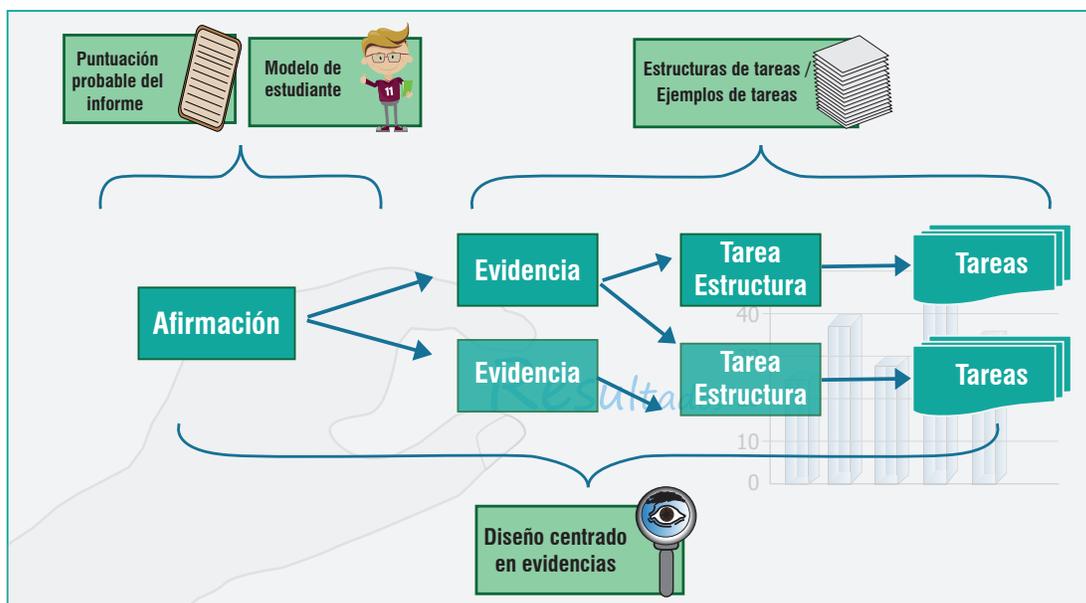


Figura 2. Estratos del DCE

Si bien el diseño de evaluación en el Icfes sigue los pasos o estratos definidos por el DCE, el diseño de las especificaciones de las pruebas de educación en Colombia se enfoca en cuatro estratos fundamentales: el análisis del **dominio**, la especificación de las **afirmaciones**, la definición de las **evidencias** y el desarrollo de **tareas**, a partir de las cuales se construirán las preguntas o ítems de las pruebas. La figura 3 ilustra la jerarquía de estratos del DCE adoptado por el Icfes en el diseño de instrumentos de evaluación de la educación:

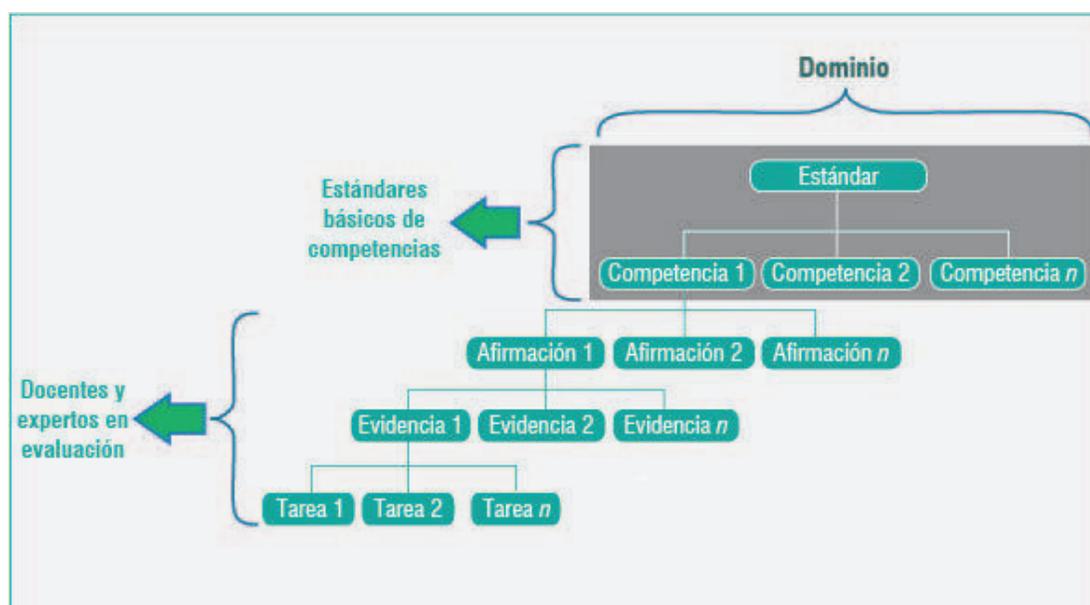


Figura 3. Modelo usado actualmente en el diseño de pruebas en el Icfes

A continuación, se describe con mayor detalle cada uno de los cuatro estratos adoptados, junto con ejemplos concretos de su implementación.

## 2.1 Dominio

El primer estrato en el diseño de evaluaciones es el análisis del dominio, el cual reúne la información que permite identificar el conjunto de CHD que la evaluación pretende medir y, así, determinar el **propósito de la prueba**. Adicionalmente, el análisis de dominio incorpora una investigación de las características de este conjunto de CHD. En esencia, el análisis de dominio provee la fundamentación para el diseño de la prueba, al especificar los CHD relevantes, los modos en

que son representados, la relación entre ellos y la forma general en que son adquiridos. Para llevar a cabo un análisis de dominio se requiere de la ayuda de expertos en el tema que faciliten la adquisición de la información necesaria para soportar el análisis. Dicha información puede encontrarse en diversas fuentes como:

- Comités de expertos en contenido.
- Análisis curricular.
- Análisis de las tareas.
- Encuestas curriculares.
- Estándares nacionales y profesionales.
- Libros de texto.

Los CHD definidos en un dominio pueden ser de diversa índole, por ejemplo, el conjunto de habilidades motoras (flexibilidad, velocidad, fuerza) requeridas para ser un bailarín de *ballet*, las destrezas necesarias para ser un chef, los conocimientos y habilidades para entablar una conversación en mandarín o los conocimientos y habilidades para que un individuo conduzca competentemente un vehículo automotor en un contexto urbano.

## 2.2 Afirmaciones

En este estrato se define aquello que es posible decir acerca de los CHD de los evaluados. Específicamente, las afirmaciones son el modo de comunicar las conclusiones a las que se puede llegar a partir de la puntuación de la prueba. Esto significa que hay un salto inferencial entre, por un lado, la puntuación de la prueba (aspectos observables) y, por otro lado, las afirmaciones sobre CHD; esto es, dado que los CHD no son directamente observables, aunque los resultados de la prueba sí lo sean, es a partir de la observación de estos resultados que se puede inferir que un individuo tiene (y en qué nivel) los CHD que la prueba evalúa. Para aclarar en qué consiste este salto inferencial, considere el siguiente ejemplo de CHD definido en el propósito de una prueba:

**Dominio:** Determinar los CHD que necesita un individuo para conducir competentemente un vehículo automotor (manual) en un contexto urbano.

La habilidad que tiene un individuo de conducir competentemente en un contexto urbano no es, en sí, un objeto que sea directamente observable, debido a que no se tiene una experiencia perceptual (visual, táctil, olfativa, etc.) de esta habilidad, sino de las manifestaciones conductuales o productos físicos de alguien que posea esta habilidad. Por consiguiente, para evaluar si alguien

tiene (y en qué grado) dicha habilidad, y, así, cumplir con el propósito de la prueba, es necesario reunir evidencias a partir de aspectos que sí son directamente observables. En otras palabras, se requiere reunir evidencia a favor de varias afirmaciones que, en suma, sustenten que un individuo tiene la habilidad en cuestión. Estas afirmaciones, para el caso del dominio en consideración, pueden ser las siguientes:

**Afirmación 1:** el individuo tiene la agudeza visual para detectar obstáculos en una vía.

**Afirmación 2:** el individuo conoce las normas de tránsito.

**Afirmación 3:** el individuo es capaz de operar eficazmente los sistemas eléctricos y mecánicos que conforman un vehículo automotor.

Nótese que las afirmaciones pueden ser generales, como la **afirmación 1**, o más específicas, como la **afirmación 2**. No existen criterios invariables o definitivos que definan el grado de generalidad o especificidad que deba tener una afirmación. Sin embargo, dos puntos deben ser considerados a la hora de formular una afirmación. En primer lugar, una afirmación debe ser la respuesta a la pregunta: ¿qué queremos decir sobre los examinados con base en sus respuestas en la prueba? En segundo lugar, una afirmación debe ser **clara**, en el sentido en que debe ser posible, al menos teóricamente, determinar evidencias que permitan mostrar que, efectivamente, un individuo cumple con lo dicho en la afirmación<sup>2</sup>.

El proceso de formulación de afirmaciones, así como el proceso de análisis de dominio, emplea diversas fuentes de información y, a menudo, también requiere la asesoría de un grupo de expertos.

## 2.3 Evidencias

En el tercer estrato se definen las evidencias necesarias para apoyar las afirmaciones descritas en el estrato anterior. Para construir una evidencia es necesario describir los aspectos de una conducta o producto observable que sustentan la conclusión de que un individuo tiene los CHD que constituyen el dominio de una prueba. Por consiguiente, la elaboración de una evidencia es importante por, al menos, dos razones.

Primero, no todos los aspectos de una conducta o producto observable soportan una afirmación sobre la habilidad de un individuo: así, por ejemplo, estar maquillado o usar pantalones cortos mientras se conduce puede ser un aspecto de una conducta de manejo, pero es irrelevante como

<sup>2</sup> Así, para una afirmación como “el individuo es capaz de hacer representaciones mentales en cuatro dimensiones espaciales”, difícilmente se podrá descubrir evidencia que la confirme.

evidencia para establecer alguna de las tres afirmaciones mencionadas arriba. Una descripción detallada de las evidencias permite distinguir los aspectos de una conducta o producto que son **relevantes** como datos para alimentar el argumento evidencial de la prueba.

Segundo, un mismo producto o conducta observable puede, en principio, proveer un sustento a diferentes afirmaciones, por lo que se requiere de una descripción detallada de los aspectos que soportan una afirmación, pero no (necesariamente) a todas las afirmaciones. Por ejemplo, la conducta observable de frenar suavemente ante un semáforo en rojo provee evidencia tanto para la afirmación de que un individuo conoce el significado de las señales de control de tránsito, como para la afirmación de que es capaz de frenar el vehículo en situaciones apropiadas. Así, la descripción de la evidencia tendrá que especificar qué aspectos de frenar suavemente ante un semáforo en rojo proveen un soporte para la primera afirmación (y no necesariamente para la segunda), y qué aspectos de esa misma conducta proveen un soporte para la segunda (pero no necesariamente para la primera).

En términos generales, la construcción de evidencias responde a la pregunta: ¿qué tiene que hacer el evaluado y, más exactamente, qué puede mostrar al hacerlo (dentro de las limitaciones de la aplicación de la prueba) que permita hacer la afirmación deseada?

Aunque no existe un procedimiento establecido para la construcción de evidencias, es recomendable aplicar el siguiente cuestionario para delimitar los aspectos de un producto o conducta observable que pueden contar como evidencias.

- a) Para comenzar, se debe definir una situación hipotética ideal en la cual se podría recolectar evidencias a partir de una conducta o producto.
- b) Luego de definir esta situación, se debería poder responder a las siguientes preguntas: ¿qué se observaría de los evaluados? y ¿cómo se podría dar cuenta de lo que se pretende ver?
- c) Posteriormente, se debe determinar qué aspectos de lo observado no pueden tenerse en cuenta, dadas las restricciones reales de la prueba (límite de tiempo, sustrato en el que se aplica la prueba y lugar de aplicación, entre otras). Para ello, se debe responder a las preguntas: ¿a qué se debe renunciar de la situación ideal, dadas las restricciones concretas? y ¿cuán cercano puede estar el contexto real de observación a la situación de observación ideal?
- d) Finalmente, a partir de los aspectos que se pueden observar en la situación real, se deben escoger aquellos aspectos que son **relevantes** para soportar una afirmación. Para esto, se sugiere dar cuenta de las siguientes cuestiones: ¿cuáles son las características de un producto, o conducta, correcto?, ¿cuáles son las diferencias cruciales entre una conducta o un producto correcto y uno incorrecto? y ¿cuáles de estas diferencias son claras y probablemente observables?

Siguiendo el ejemplo de dominio descrito arriba, se pueden integrar las siguientes evidencias a una de las tres afirmaciones asociadas anteriormente a dicho dominio:

**Dominio.** Determinar los CHD que necesita un individuo para conducir competentemente un vehículo automotor (manual) en un contexto urbano.

**Afirmación 3:** el individuo es capaz de operar eficazmente los sistemas eléctricos y mecánicos que constituyen un vehículo automotor.

**Evidencia 3.1:** el individuo pone en marcha el vehículo en diferentes circunstancias, activando correctamente el sistema de transmisión.

**Evidencia 3.2:** el individuo opera correctamente la caja de cambios con el vehículo en marcha.

**Evidencia 3.3:** el individuo opera correctamente el sistema de dirección del vehículo para dar giros en diferentes contextos (cambiar de carril o de calzada, adelantar, sortear obstáculos, etc.).

A partir de la suma de estas evidencias, el evaluador tendrá información (parcial) **relevante** para decidir si el individuo es capaz de operar eficazmente los sistemas eléctricos y mecánicos que constituyen un vehículo automotor; es decir, tendrá información para apoyar la afirmación 3.

Nótese que las evidencias no son en sí productos o conductas observables, sino descripciones de los aspectos de estos productos o conductas que son **relevantes** para soportar una afirmación sobre los CHD evaluados. Los productos o conductas observables, cuyos aspectos son descritos en el diseño de evidencias, corresponden, como se verá en lo que sigue, al sustrato de las **tareas**.

## 2.4 Tareas

En el cuarto estrato, el de las tareas, se describe cómo estructurar las situaciones que se necesitan para obtener las evidencias definidas en el anterior estrato. En otras palabras, en el estrato de las tareas se define un escenario, o ambiente, normalmente problemático, que requiere de una solución mediante una acción o producto observable que manifiesta la posesión de una habilidad que se quiere medir (Mislevy, 2017, p. 90).

Esquemáticamente, el estrato de las tareas incluye, entre otros, los siguientes elementos:

- a) Una descripción de los CHD que las tareas miden.
- b) Los tipos de estímulo material que pueden emplearse para que el evaluado exhiba esos CHD.
- c) Una descripción de lo que se le pedirá hacer al evaluado.
- e) Ejemplos de varias tareas que pueden generarse a partir de los tres anteriores elementos (Zieky, 2014, p. 83).

El diseño de tareas provee el insumo para la construcción de los diversos tipos de ítems o preguntas que constituyen una prueba y, por esta razón, el diseño de tareas es crucial para la fase de armado de la prueba, estrato en el cual, entre otras cosas, se distribuyen las tareas de acuerdo con su grado de dificultad. Conforme con la anterior, para especificar las tareas es pertinente responder a las siguientes preguntas:

- ¿Qué tareas se deben diseñar para recolectar la suficiente evidencia?
- ¿Qué características de las tareas afectan la validez de la prueba?
- ¿Qué características de las tareas afectan la dificultad de estas?
- ¿Las tareas, en conjunto, se pueden resolver dadas las restricciones de la aplicación (tiempo, sustrato físico de la prueba, etc.)?

La descripción de una tarea puede modificar el tipo de ítems construidos a partir de esta y, consecuentemente, la conducta o producto esperado de un evaluado al resolver dichos ítems. Una descripción inapropiada de una tarea puede, entonces, afectar negativamente la construcción de ítems, ya sea entorpeciendo la comprensión de las instrucciones por parte del estudiante o generando ítems que requieran, para su respuesta, habilidades diferentes (o de mayor o menor complejidad) a las que se intentan medir. Para evitar estos inconvenientes, la descripción de las tareas debe emplear un lenguaje conciso, preciso e inclusivo<sup>3</sup>.

<sup>3</sup> Usar un lenguaje inclusivo implica el uso de enunciados libres de contenidos sesgados ideológicamente, ofensivos o discriminatorios.

Dado el ejemplo de dominio citado anteriormente, se pueden agregar los siguientes ejemplos de tareas para una de sus evidencias:

**Dominio.** Determinar los CHD que necesita un individuo para conducir competentemente un vehículo automotor (manual) en un contexto urbano.

**Afirmación 3:** el individuo es capaz de operar eficazmente los sistemas eléctricos y mecánicos que constituyen un vehículo automotor.

**Evidencia 3.2:** el individuo opera correctamente la caja de cambios con el vehículo en marcha.

**Tarea 3.2.1:** dado un vehículo en marcha, el individuo ejecuta los cambios teniendo en cuenta las revoluciones por minuto del motor.

**Tarea 3.2.2:** dado un vehículo en marcha, el individuo ejecuta los cambios teniendo en cuenta la velocidad alcanzada.

**Tarea 3.2.3:** el individuo reduce o aumenta los cambios apropiadamente de acuerdo con las condiciones de la vía y el tráfico.

Como se puede observar, las tareas funcionan como la estructura o esqueleto de un escenario real al dejar abiertas determinadas variables que en la construcción de ítems específicos se pueden detallar. Así, siguiendo el ejemplo, las tareas no especifican un número de revoluciones por minuto determinado ni una velocidad exacta ni las condiciones precisas de tráfico en una vía. Cada uno de los ítems construidos a partir de estas tareas puede especificar estas variables o pedir al evaluado que lo haga.

Se puede condensar lo dicho hasta acá a través de la tabla 1, que muestra un formato en el cual se pueden construir las especificaciones de una prueba.

Tabla 1. Formato de Especificaciones

Nombre de la competencia					
Definición de la competencia					
Afirmación	Evidencia	Tarea	¿Qué CHD mide?	¿Qué se pide hacer al evaluado?	Ejemplo de ítem
Afirmación 1	Evidencia 1	Tarea 1			
		Tarea 2			
	Evidencia 2	Tarea 1			
		Tarea 2			

### 3. VENTAJAS DE LA IMPLEMENTACIÓN DEL DCE

Los rasgos principales que determinan la calidad de un instrumento de evaluación son dos; a saber, la **confiabilidad** y la **validez**. El primero es un indicador de la consistencia con que una prueba mide un rasgo a través de un periodo de tiempo y diferentes participantes. La validez, como se explicó en la primera sección, hace referencia al grado en el que el instrumento mide los CHD, o rasgos latentes para los que fue diseñado.

La ventaja principal de implementar el DCE, justamente, está vinculada con la validez: el DCE es una herramienta que genera las condiciones para garantizar la validez de una prueba, pues, como se mostró, el DCE articula la línea o cadena de razonamiento que explica por qué los resultados de la prueba se relacionan con los CHD que la prueba pretende medir (Mislevy *et al.*, 2003, p. 2-3).

Adicionalmente, el DCE funciona como un mecanismo para la preservación de una prueba a través del tiempo, ya que permite tanto la construcción de nuevas tareas, similares a tareas preexistentes que han sido calibradas estadísticamente, como el armado de nuevas pruebas paralelas a las ya aplicadas (Mislevy *et al.*, 2017, p.23).

Finalmente, el DCE organiza y unifica conceptos para crear una terminología común a través de las diferentes disciplinas en las que se especializan los diseñadores de los módulos de una prueba, lo cual facilita la comunicación entre estos actores y la incorporación de nuevos actores en el proceso de diseño.

# REFERENCIAS

- Almond, R., Steinberg, L., & Mislevy, R. (2002). "Enhancing the design and delivery of assessment systems: A four-process architecture". *Journal of Technology, Learning, and Assessment*, 1(5). Recuperado de: <https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1671>
- Messick, S. (1989). "Validity". En: R. L. Linn (Ed.). *Educational measurement*. Washington, DC: American Council on Education. (pp.13-104).
- Mislevy, R. et al. (2003). *A brief introduction to evidence-centered design*. Educational Testing Service, Princeton, NJ. July 2003.
- Mislevy, R. J., & Riconscente, M. M. (2005). "Evidence-centered design: Layers, structures, and terminology". Menlo Park, CA: SRI International.
- Mislevy, R. et al (2017). *Assessing Model-Based Reasoning using Evidence-Centered Design: A Suite of Research-Based Design Patterns*. Springer.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Zleky, M. (2014). "An introduction to the use of evidence-centered design in test development". *Psicología Educativa* 20. Doi: <https://doi.org/10.1016/j.pse.2014.11.003>



GOBIERNO DE COLOMBIA