



Lineamientos para la revisión de sesgo de contenido en las preguntas de pruebas estandarizadas



Presidente de la República

Iván Duque Márquez

Ministra de Educación Nacional

María Victoria Angulo González

Publicación del Instituto Colombiano
para la Evaluación de la Educación
(Icfes)

© Icfes, 2021.

Todos los derechos de autor
reservados.

Se agradece la colaboración y
participación en la elaboración de
este documento a:

María Camila Devia Cortés

María del Pilar Soler Parra

Manuel Alejandro Amado González

Miguel Fernando Moreno Franco

Roger Camilo Alfonso Leal

Edición

Juan Camilo Gómez-Barrera

Diseño de portada y diagramación

Linda Nathaly Sarmiento Olaya

Fotografía portada

Freepik (2021)

Directora General

Mónica Patricia Ospina Londoño

Secretario General

Ciro González Ramírez

Directora de Evaluación

Natalia González Gómez

Director de Producción y Operaciones

Oscar Orlando Ortega Mantilla

Director de Tecnología e Información

Sergio Andrés Soler Rosas

Subdirector de Diseño de Instrumentos

Luis Javier Toro Baquero

Subdirectora de Estadísticas

Jeimy Paola Aristizábal Rodríguez

Subdirectora de Análisis y Divulgación

Mara Brigitte Bravo Osorio

ISBN de la versión digital:

En trámite

Bogotá, D. C., julio de 2021

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.

TABLA DE CONTENIDO

Presentación	4
Introducción	6
Alcance de los lineamientos	8
Conceptos básicos: constructo, validez, equidad de la prueba y sesgo de contenido	10
¿Cómo usar estos lineamientos?	14
1. Principios generales para la revisión del sesgo de contenido	16
1.1 Barreras cognitivas como fuente de sesgo	16
1.2 Barreras afectivas como fuente de sesgo	22
1.3 Barreras físicas como fuente de sesgo	28
2. Consideraciones especiales	31
2.1 Representación de la diversidad ...	31
2.1.1 <i>La diversidad en el contenido de las evaluaciones</i>	32
2.1.2 <i>Pertinencia de la diversidad en el contenido de los ítems</i> ...	35
2.2 Contenido apropiado para la evaluación de menores de edad	37
Conclusión	40
Referencias	41

Presentación

Este texto hace parte de los documentos de trabajo con los que el Icfes cuenta para asegurar la calidad de sus contenidos y, en especial, de las preguntas que conforman sus pruebas. En ese sentido, los lineamientos que aquí se presentan son una herramienta dirigida, principalmente, a diseñadores de pruebas (gestores de prueba, constructores de marcos de referencia y de ítems, revisores técnicos, asesores) y demás agentes que participan en el proceso de construcción de ítems (validadores, correctores de estilo y revisores en las diferentes etapas de construcción de la prueba).

Este documento se estructura en tres partes. En la primera se presentan el objetivo, el alcance, el público a quien va dirigido y las recomendaciones de uso de estos lineamientos, así como algunos conceptos básicos para su comprensión, como **constructo**, **equidad de la prueba**, **sesgo de contenido** y **validez**. En la segunda parte se exponen los principios generales de la revisión de sesgo en el contenido de los ítems, los cuales aseguran lo que en la literatura especializada se conoce como equidad de la prueba (en inglés, *test fairness*)¹. Estos principios

¹ El concepto de *test fairness*, así como los de *sesgo* y *validez*, se extraen principalmente de: American Educational Research Association, American Psychological Association, y National Council on Measurement in Education (AERA, APA y NCME; 2014, pp. 49-50), Educational Testing Service (2015, pp. 4-6), Messick (1989, p. 13) y Zieky (2013, p. 293-295).

se asocian a las tres fuentes de sesgo que comprometen la equidad y, por tanto, la validez de las pruebas. Estas fuentes son: las cognitivas, las afectivas y las físicas. La exposición de cada uno de los principios generales viene acompañada de recomendaciones sobre el abordaje de temas que pueden generar, en quien responde, efectos emocionales negativos. Finalmente, en la tercera parte, se aborda el tratamiento y representación de la diversidad en las pruebas y se brindan algunas recomendaciones para la selección de material en evaluaciones de menores de edad.

Introducción

Para garantizar la calidad de las pruebas y de las preguntas es necesario llevar a cabo revisiones cuantitativas y cualitativas. Las primeras son revisiones de los parámetros estadísticos de los ítems (como la dificultad y la discriminación, el funcionamiento diferencial de los ítems, el análisis de proporción de los distractores, entre otros); mientras que las segundas están relacionadas con el contenido de los ítems. En este último tipo de revisión se busca garantizar que los ítems cumplan con una serie de requerimientos técnicos: por ejemplo, que los de selección múltiple con única respuesta solo tengan una opción correcta, que el contenido refleje las especificaciones de la prueba y que el lenguaje y contenido sea el apropiado para la población en la que se busca medir los atributos o características. Junto a esta revisión técnica del contenido, también se efectúa una revisión editorial, de estilo y diagramación, en la que se inspeccionan aspectos gramaticales, la claridad del lenguaje, la consistencia gramatical, la disposición espacial de los ítems y los gráficos que los acompañan.

Además de las revisiones cuantitativas y cualitativas de los ítems, la literatura recomienda una revisión del contenido, denominada **revisión de equidad** (en inglés, *fairness revision*). El propósito de esta revisión es identificar factores que no están relacionados con lo que se pretende medir con una prueba y que pueden afectar el modo como los miembros de un determinado subgrupo de la población evaluada responde a un ítem. A menudo, estos factores pasan desapercibidos en las revisiones técnicas y estadísticas debido a que son factores de carácter cognitivo, afectivo o físico que no afectan a la mayoría

de la población, sino a grupos demográficos específicos. Por esto, se hace necesario llevar una revisión de equidad usando lineamientos que provean una descripción detallada de los procedimientos y el fundamento de este tipo de revisión. En el presente documento se recogen los lineamientos que el Icfes ha aplicado con el objetivo de ejecutar la revisión de sesgo en conjunto con las recomendaciones de la literatura técnica (ETS 2009, 2015; Ziecky 2013, Wendler y Burrus 2013).

Estos lineamientos sirven de guía para la comprensión del concepto de **equidad** y su relación con la **validez** de las pruebas, así como la preservación de la equidad de las pruebas del Icfes a través de la identificación de potenciales **fuentes de sesgo** en el contenido de los ítems. En esa medida, estos lineamientos no son normas rígidas que se aplican a todas las pruebas, independientemente de lo que estas evalúen. Por el contrario, son recomendaciones sobre el uso de criterios que pueden variar dependiendo de la prueba; sirven para identificar rasgos en el contenido de los ítems que constituyan potenciales barreras cognitivas, afectivas o físicas para medir de forma apropiada los conocimientos, habilidades o destrezas que evalúan las pruebas. En otras palabras, estos lineamientos son una herramienta que permiten señalar aspectos en los ítems que pueden causar diferencias injustas en los puntajes de los evaluados de diferentes grupos demográficos; es decir, diferencias en los puntajes por factores que no están relacionados con lo que se pretende medir (por ejemplo, factores como conocimientos de temas técnicos o tecnológicos, o temas que puedan generar reacciones emotivas fuertes como ansiedad) y que pueden producir una desventaja, o una ventaja, no deseada a un determinado grupo demográfico.

Alcance de los lineamientos

Los lineamientos para la revisión de sesgo se enfocan en el contenido de los ítems que componen las pruebas del Icfes, tanto las que evalúan habilidades genéricas como las que evalúan competencias específicas. Sin embargo, los lineamientos también pueden ser usados en la revisión de otros documentos: comunicaciones, guías de orientación, presentaciones, informes, entre otros.

Con respecto a la revisión del contenido de los ítems, aunque en este documento se hace énfasis en los de selección múltiple con única respuesta (contexto, enunciado, opciones de respuesta, clave), la revisión de sesgo no se limita a este tipo de ítems, por lo que puede aplicarse a otros tipos de pregunta cerrada y de respuesta construida.

Cabe anotar que estos lineamientos deben ser implementados antes de aplicar las pruebas a los evaluados para reducir los efectos de contenidos potencialmente sesgados. Dado que la evidencia empírica de los efectos del sesgo se encuentra normalmente luego de la aplicación de los ítems (al llevar a cabo estudios sobre el comportamiento diferencial de los ítems), la revisión del sesgo a partir de los lineamientos es necesariamente una identificación de fuentes **plausibles** de ventajas, o desventajas, injustas que afecten los puntajes de los evaluados. Por esta misma razón, estos lineamientos no versarán sobre las estimaciones estadísticas de las diferencias en puntaje, pues no se trata de una revisión cuantitativa, sino de una revisión cualitativa adicional a las revisiones técnicas y editoriales o de estilo.

Por supuesto, que la revisión de sesgo se realice antes de tener datos estadísticos de las preguntas de la prueba no implica que, durante el análisis estadístico de ítems, las medidas del funcionamiento diferencial (DIF) no puedan ser tenidas en cuenta para hacer juicios sobre la equidad de la prueba. Esto porque el DIF ocurre cuando un grupo de personas obtiene puntajes sustancialmente diferentes a los de otro grupo, a pesar de que los dos grupos de personas sean similares en términos de los conocimientos, habilidades y otras destrezas evaluados en la prueba. Así, si los datos del DIF se encuentran disponibles, se recomienda que aquellos ítems con DIF alto sean revisados para considerar su inclusión en la calificación (ETS 2015, p. 47).

Conceptos básicos: constructo, validez, equidad de la prueba y sesgo de contenido

En términos generales, una prueba es un instrumento de medición de un constructo, el cual se define como un conjunto de conocimientos, habilidades, destrezas, rasgos u otros atributos que no son directamente observables. Algunos ejemplos de constructo son: las habilidades de diseñar estrategias para resolver problemas que involucren las matemáticas, la comprensión y evaluación de textos, la capacidad de usar conceptos de las ciencias naturales para la resolución de problemas, las actitudes de identidad racial y rasgos como la autoestima o el perfeccionismo (AERA, APA, y NCME, 2014, p. 11, lcfes 2018, 2020a).

La calidad con que una prueba mide un constructo está determinada por el grado de validez que exhiba la prueba. La **validez** está relacionada con la medida en que las interpretaciones de los puntajes o resultados de una prueba son adecuadas y apropiadas para los usos propuestos para la prueba; en efecto, esta noción se define como el grado en el cual tanto la teoría sobre el constructo por medir (que en el lcfes se consigna en los marcos de referencia) como la evidencia empírica soportan o apoyan las interpretaciones o acciones que se tomen a partir de los puntajes de una prueba (Messick 1989, 1995). En consecuencia, una prueba tendrá mayor validez entre más apropiadas o adecuadas sean las interpretaciones o acciones que se realicen a partir de los puntajes obtenidos por los evaluados. Así, por ejemplo, los puntajes de una prueba pueden ser usados para afirmar que un estudiante tiene determinado nivel de comprensión lectora.

También pueden ser usados para ubicar a un estudiante en un determinado programa educativo o para permitirle ingresar a la educación superior. En definitiva, la validez es una propiedad de los usos o interpretaciones que es posible hacer de los puntajes de una prueba.

Para sustentar la validez de las interpretaciones y usos de una prueba, es necesario contar con evidencias suficientes. Las fuentes de dichas evidencias son diversas: algunas tienen que ver con el contenido de las preguntas, otras con el proceso de responder las preguntas o con la estructura interna de la prueba, entre otras fuentes de evidencia. La revisión de sesgo, como se explicará, se propone recoger evidencias relacionadas con el contenido, en la medida que suministra evidencias de que las pruebas reflejan aquello que define el constructo de la prueba (Messick 1995).

La validez y la calidad de la medición de una prueba se ven afectadas por **factores irrelevantes al constructo**. Estos factores pueden ser vistos como contaminantes que impiden la interpretación apropiada de los puntajes de una prueba. Por ejemplo, un lenguaje innecesariamente complejo (como el uso de metáforas, símiles o descripciones físicas sofisticadas o adornadas) en las preguntas de una prueba de matemáticas es un factor irrelevante al constructo. Esto se debe a que la comprensión de dicho lenguaje no pertenece a los conocimientos que constituyen el constructo y puede afectar negativamente los puntajes de los estudiantes, quienes, si no apareciera dicho lenguaje, podrían exhibir las habilidades que la prueba pretende medir. Por tanto, reducir al mínimo la

ocurrencia de factores irrelevantes al constructo evita que las interpretaciones de los resultados de la prueba pierdan validez.

Cuando un factor irrelevante al constructo produce diferencias en los puntajes de un subgrupo de evaluados, pero no en todos los evaluados, se produce una **varianza irrelevante al constructo** (Ziecky 2013, p. 294-295). Por ejemplo, si para responder correctamente una pregunta de comprensión lectora se requiere un conocimiento o familiaridad con un aspecto irrelevante al constructo (como conocer un lugar de la capital del país o los datos biográficos de una persona), probablemente se generará una varianza irrelevante al constructo. Esto ocurre porque un grupo de estudiantes que no estén familiarizados con dicho aspecto (que no conozcan la capital o a la persona mencionada) pueden ver afectados sus puntajes. Como consecuencia de esta varianza, las interpretaciones de los puntajes para este subgrupo de evaluados serán menos válidas.

Cuando una fuente de varianza irrelevante para el constructo afecta más a los puntajes de un subgrupo de evaluados (por ejemplo, mujeres, estudiantes de zona rural) que a otros (hombres, estudiantes de zona urbana), la prueba carece de **equidad** y, en este sentido, la interpretación de los resultados pierde validez.

La revisión de sesgo en el contenido juega un papel central en el contexto de la equidad de las pruebas. En efecto, el **sesgo** en el contenido se define como “cualquier factor irrelevante para el constructo que genere de forma sistemática una disminución o aumento en los puntajes de un subgrupo identificable de evaluados” (Ziecky 2013, p. 294). Puesto que los sesgos

conducen a una disminución en la equidad, es deseable, para contribuir a la equidad, contar con un procedimiento para evitar o minimizar los sesgos.

Cabe notar que no toda variación de puntajes en subgrupos implica inequidad en la prueba. Si las variaciones en puntajes se deben a diferencias en la adquisición o desarrollo de conocimientos o habilidades que sí son centrales al constructo, las variaciones de puntajes en subgrupos serán válidas y la equidad de la prueba no estará comprometida. Por ejemplo, si las variaciones se deben a la comprensión y uso de un concepto que evalúa la prueba (como la relación entre conceptos fisicoquímicos en una prueba de ciencias naturales), la equidad no se afecta.

El propósito de la revisión de sesgo es identificar los factores irrelevantes para el constructo que plausiblemente pueden afectar los puntajes de un subgrupo de evaluados. Una vez identificadas las fuentes de sesgo, es preciso minimizarlas para evitar que se conviertan en barreras que impidan que los evaluados exhiban las habilidades, conocimientos u otras destrezas que se pretendan medir. Los presentes lineamientos ofrecen una serie de recomendaciones para llevar a cabo la revisión de sesgo en el contenido, de modo que se puedan identificar las fuentes potenciales de sesgo y minimizar su presencia. En la sección 2 de este documento se explica cuáles son estas fuentes y cuáles son las recomendaciones para identificarlas y reducirlas.

¿Cómo usar estos lineamientos?

Dado que el propósito de los lineamientos es reducir las fuentes de sesgo y, así, contribuir a las evidencias de la validez de las pruebas, estos no deben ser entendidos como normas de redacción de textos políticamente correctos o como parte de una censura de contenidos. Los lineamientos no son prohibiciones. Por el contrario, se basan en la noción de contenido relevante para un constructo, lo que implica que los contenidos que son innecesarios, o potencialmente inapropiados en una prueba, pueden ser relevantes y aceptables para otra. Por ejemplo, la descripción de los síntomas de una enfermedad de transmisión sexual puede ser relevante para una pregunta que mide el constructo de una prueba de las ciencias de la salud, pero la misma descripción puede ser innecesaria e inapropiada (o incluso ofensiva) para una prueba que evalúe habilidades genéricas, como las evaluadas en las pruebas de Competencias Ciudadanas. En consecuencia, los lineamientos deben considerarse teniendo en cuenta lo que se evalúa en cada prueba.

Dado que los juicios sobre lo que se considera relevante para un constructo pueden variar para diferentes expertos de contenido en una prueba, se recomienda que los lineamientos sean revisados por un conjunto de expertos en los diversos momentos del proceso de construcción, revisión y validación de ítems. Adicionalmente, con el fin de enriquecer los juicios acerca de los materiales potencialmente sesgados y considerando que los lineamientos deben ser contextualizados de acuerdo con la prueba revisada y el público evaluado, se recomienda tomar estos lineamientos en conjunto con los

documentos teóricos que sustentan las pruebas del Icfes, como los que versan sobre el Diseño Centrado en Evidencias, la definición de niveles de desempeño, el proyecto de inclusión, las reglas de construcción de preguntas, así como los marcos de referencia y (en el caso de gestores, constructores y revisores) las especificaciones de prueba.

Idealmente, la revisión de sesgo debería considerar cualquier subgrupo demográfico de evaluados. Sin embargo, en un país tan diverso como el nuestro, este es un ideal inalcanzable. Por ello, se recomienda dar especial atención a las posibles fuentes de sesgo que afecten a grupos históricamente discriminados por presuntas diferencias en edad, género, etnia, discapacidad, religión, ubicación geográfica o nivel socioeconómico (Ziecky 2013, p. 296). Por último, debido a que los materiales que pueden ser fuente potencial de sesgo son numerosos y pueden variar en el tiempo por cambios en el contexto social o ajustes en las especificaciones de prueba (cambio del constructo), estos lineamientos requieren actualización y revisión periódica.

1. Principios generales para la revisión del sesgo de contenido

De acuerdo con el Educational Testing Service (ETS) (2009, 2015), existen al menos tres principios generales para la revisión de contenido que permiten reducir el sesgo y, así, garantizar la equidad de una prueba. Estos principios se corresponden con las fuentes potenciales de sesgo o variación irrelevante al constructo. Los principios generales de revisión de sesgo que se derivan de sus principales fuentes son:

- a) Evitar barreras cognitivas irrelevantes para el constructo.
- b) Evitar barreras afectivas irrelevantes para el constructo.
- c) Evitar barreras físicas irrelevantes para el constructo.

A continuación, se explican los tres tipos de fuentes y se indica cómo aplicar los principios generales en la revisión de sesgo de contenido.

1.1 Barreras cognitivas como fuente de sesgo

Las fuentes cognitivas de sesgo son todos aquellos conocimientos o habilidades irrelevantes al constructo que, a pesar de no ser parte de lo que se pretende medir, se requieren para contestar una o más preguntas de una prueba. Estos conocimientos o habilidades irrelevantes al constructo pueden ser desarrollados o adquiridos de manera diferente por subgrupos de evaluados

y, por tanto, pueden generar variaciones en los puntajes. Esto puede comprometer la equidad y validez de los resultados de la prueba al actuar como barreras, pues hacen que aquellos grupos de estudiantes que no tienen estos conocimientos o habilidades vean afectado su desempeño en la prueba.

La equidad y, por consiguiente, la validez de la prueba se comprometen cuando las variaciones de puntajes se deben a habilidades o conocimientos irrelevantes al constructo. De este modo, si para responder correctamente una pregunta de una prueba se requiere conocer las reglas de un deporte, pero el conocimiento de estas reglas es irrelevante para el constructo, las diferencias en puntajes que esto ocasione en diferentes subgrupos sí afectarán la equidad de la prueba. Como consecuencia, dicho conocimiento se convertirá en una barrera, una fuente de un sesgo cognitivo (ETS 2015, p. 14). Para evitar o minimizar las barreras cognitivas en las pruebas y reducir el sesgo que de estas se derivan, la literatura recomienda atender a las siguientes consideraciones:

- ▶ Evitar el lenguaje innecesariamente complejo. Se sugiere evitar tecnicismos que describan conceptos no evaluados en la prueba, y que puedan transmitirse de forma accesible en un lenguaje de uso corriente o que puedan suprimirse sin afectar el constructo. En pruebas genéricas, se debe evitar el uso innecesario de terminología especializada de áreas como las finanzas, el derecho, las ciencias o las tecnologías (computación, medios de transporte, herramientas), así como temas deportivos, médicos, militares y políticos especializados. No es apropiado suponer que todos los

evaluados saben, por ejemplo, qué es un fideicomiso, una resonancia magnética funcional, cómo se usa una caladora, cuáles son los tipos de agravaciones punitivas, cuáles son los grados militares, cuántos jugadores deben conformar un equipo de baloncesto o qué es una reducción al absurdo.

- ▶ Reducir el uso de metáforas o analogías, sobre todo si la prueba no mide la comprensión de estas (como en la prueba de Lectura Crítica). La comprensión del lenguaje figurado impone una carga cognitiva innecesaria a diversos grupos poblacionales.
- ▶ No emplear regionalismos (expresiones de uso frecuente en una región del país). Por ejemplo, evitar palabras como “chicanero”, “mañoso”, “vaina”, “caco”, “boleta”, “cachumbo”, etc. Los regionalismos son fuentes potenciales de sesgo, pues el significado de estas expresiones puede variar en diferentes grupos demográficos y, por tanto, alterar el sentido de una instrucción o dificultar la comprensión de las opciones de respuesta. Igualmente, se debe evitar el uso de dichos populares (por ejemplo, “a caballo regalado no se le mira el colmillo”, “matar dos pájaros de un solo tiro”), pues la comprensión de estos puede requerir conocimientos culturales específicos de una región o imponer una carga cognitiva innecesaria a diversas poblaciones, como aquellas con discapacidad.

- ▶ En lo posible, no se deben mencionar explícita o implícitamente personajes o entidades de la realidad nacional o internacional, de programas de televisión o de redes sociales, si el conocimiento de estos o de sus funciones no es central al constructo. Cuando una pregunta hace referencia a un individuo o a una institución (entidad, país, organizaciones, partidos políticos, empresas), aun si esto hace parte de una cita o si se presenta como la posición de una persona, se genera tácitamente un compromiso con una visión particular (positiva o negativa) de la actuación de esa persona o entidad. Esto puede llevar a pensar equivocadamente que el Icfes aboga por tal o cual posición política, ética, estética, etc. Además, mencionar a individuos o entidades específicos resulta en una distracción de la que se puede prescindir para los objetivos de evaluación que persigue el instrumento, ya que la pregunta podría requerir que se conozcan detalles biográficos o datos históricos que son irrelevantes al constructo.
- ▶ En caso de emplear nombres propios, deben ser fácilmente legibles y comunes entre la comunidad de evaluados. Si es posible, reemplazar nombres o apellidos extranjeros, como “Koothrappali”, por versiones cortas y familiares, como “García”.
- ▶ De ser irrelevante para el constructo, se recomienda no suponer conocimientos contextuales de, por ejemplo, ubicaciones o lugares geográficos específicos de zonas rurales o urbanas (por ejemplo, la avenida Caracas en Bogotá o la Ensenada de Utría en el Chocó). Similarmente, no

se debe suponer que los estudiantes poseen conocimientos culturales o de grupos sociales específicos de una o varias regiones del país, como conocimientos sobre festividades, usos, costumbres, platos típicos o marcas de productos. Por ejemplo, es inapropiado suponer que todos los evaluados conocen el origen y la manera como se celebra el Carnaval de Negros y Blancos.

- ▶ No requerir conocimientos en temas problemáticos o controversiales, como religiosos o filosóficos (metafísica, epistemología, ética, estética), a menos que el propósito de la pregunta sea medir el conocimiento de dichos temas. Si la prueba no mide esos conocimientos, es inapropiado suponer, por ejemplo, que todos los evaluados conocen la doctrina social de la Iglesia Católica, la moral cristiana o las diferencias entre religiones.

El siguiente ejemplo ilustra la aplicación de algunos de los criterios mencionados para reducir el contenido que genera barreras cognitivas:

Enunciado original	Enunciado revisado
<p>Una trabajadora en estado de embarazo que se desempeña en el sector privado fue despedida por su empleador sin la correspondiente autorización del Ministerio de Protección Social, en razón de que durante los últimos 100 días había estado realizando deficientemente sus labores.</p> <p>En este caso, el mecanismo idóneo para solicitar la acción de reintegro es presentar una:</p>	<p>Un empleador despide a una mujer en estado de embarazo, sin una autorización legal. El empleador argumenta que la mujer no estaba cumpliendo bien con su trabajo.</p> <p>De acuerdo con la Constitución, ¿qué mecanismo puede emplear la mujer para proteger su derecho al trabajo?</p>
<p>Comentario:</p> <p>El ítem original de una prueba de Competencias Ciudadanas es innecesariamente extenso, pues emplea términos cuya comprensión no es indispensable para responder correctamente el ítem (“sector privado”, “acción de reintegro”). Adicionalmente, emplea oraciones largas en voz pasiva (“Una trabajadora... fue despedida por su empleador...”) y el enunciado de la pregunta no es interrogativo, lo que dificulta la accesibilidad para la población con discapacidad.</p> <p>Finalmente, “mecanismo idóneo” es un término relativo al contexto, por lo que es preferible hacer explícito de que se trata de un mecanismo constitucional.</p>	

1.2 Barreras afectivas como fuente de sesgo

Las reacciones afectivas producidas por los estímulos de una prueba pueden impedir un buen desempeño de los estudiantes, o puede llevarlos a pensar que su desempeño fue obstaculizado. Adicionalmente, si estas reacciones son más probables para unos grupos que para otros, la equidad de la prueba se reducirá. Por tanto, ningún grupo de evaluados debe afrontar un lenguaje o imágenes que sean insultantes, excluyentes y ofensivos por dos razones. Primero, los estudiantes pueden distraerse si piensan que la prueba se basa en posiciones que son contrarias a las que ellos sostienen o que pretende adoctrinarlos sobre dichas posiciones. Segundo, los estudiantes pueden reaccionar de una manera más emocional que crítica, al enfrentarse a un contenido excesivamente controversial que, no obstante, mida habilidades ligadas a la reflexión y el pensamiento crítico.

En consecuencia, deben evitarse los temas que tienen una alta probabilidad de generar reacciones emocionales negativas, a menos que sean fundamentales para la prueba. Si es necesario tratar estos temas porque son centrales al constructo, deben abordarse de manera balanceada, sensible y neutral para que el efecto emocional se reduzca al mínimo. A continuación, se presentan algunos temas que pueden generar reacciones emocionales innecesarias para la medición de un constructo (ETS 2015, p. 20-27):

- ▶ Detalles sobre desastres, enfermedades, accidentes, pobreza extrema o lujo excesivo. Así mismo, eventos recientes que puedan haber generado experiencias negativas en los estudiantes, referencias negativas a población con bajos ingresos o a población cuyos gastos puedan ser considerados como un lujo excesivo.
- ▶ Defensa o toma de partido por causas sociales, ideologías, estilos de vida o juicios (generalizaciones, acusaciones, puntos de vista) injustificados como el racismo o el machismo. Las preguntas deben ser neutrales y balanceadas: las pruebas no se usan para favorecer, defender o tomar partido por ninguna ideología o posición específica en un asunto controversial.
- ▶ Tratamiento de temas controversiales u ofensivos. Algunos temas generan reacciones emocionales negativas, de manera que se deben evitar siempre que sea posible. Algunos de estos son: aborto, abuso de personas o animales, genocidio, anticoncepción, eutanasia, experimentación dolorosa o dañina con humanos y animales, caza de animales, abuso sexual, satanismo y tortura.
- ▶ Uso de nombres propios de figuras, entidades o personalidades públicas. En general, es mejor evitar el uso de nombres propios de individuos o entidades, especialmente, cuando están asociados a temas controversiales u ofensivos. El asunto crítico es que el nombre de una entidad o persona impacte emocionalmente a los evaluados, impidiéndoles responder correctamente una pregunta.

- ▶ Si se precisa el uso de datos biográficos para la construcción de una pregunta, es preferible mencionar personas o eventos del pasado distante, pues los actuales pueden cambiar sus acciones y, con ello, modificar lo que un evaluado pueda conocer de su biografía. Adicionalmente, se sugiere no mencionar, en las citas bibliográficas, los nombres de personajes o entidades. Por ejemplo, no usar la referencia cuando se menciona a un personaje reconocido.
- ▶ Referencias ambiguas, enigmáticas o desconocidas en torno a temas como el sexo, las drogas, la brujería o fenómenos paranormales. Dado que los materiales usados en una prueba provienen de diversas fuentes, es probable que algunas de ellas contengan referencias arcanas. Siempre que sea posible, se deben evitar esas referencias examinando palabras, nombres, números o imágenes que parezcan arbitrarias o extrañas, por ejemplo, el número 666 suele asociarse con el satanismo y, en redes sociales, el número 420 se asocia con el consumo de sustancias psicoactivas. Similarmente, el uso de siglas cuyo uso pretende hacer una mención despectiva a un grupo (por ejemplo, ACAB), deben ser evitadas.
- ▶ Detalles innecesariamente gráficos sobre violencia, sexo o muerte (suicidios o asesinatos). Esto no excluye que pueda indicarse el año o la causa del deceso de una persona. En cuanto al suicidio, es aceptable mencionar que alguien lo cometió, pero se debe evitar entrar en detalles o describirlo como una forma de solucionar un problema.

- ▶ Diferencias de género. Dado el creciente número de categorías con las que las personas se identifican, se recomienda no asumir que las indicaciones de género se refieren únicamente a hombres y mujeres.
- ▶ Diferencias de grupo. No es recomendable generalizar la existencia de diferencias entre grupos y sus causas. No se debe sugerir que un grupo es superior o inferior a otro con respecto a rasgos como el coraje, la honestidad, el atractivo físico y otras cualidades cuya atribución sea subjetiva o depende de la adopción de estereotipos culturales.
- ▶ Sarcasmo e ironía. Estas expresiones se deben tratar con mucho cuidado porque es posible que no sean comprendidas y, en cambio, resulten ofensivas para algunos grupos de la población que toma la prueba. Así, por ejemplo, las exageraciones, las burlas acerca de un grupo de personas, sus creencias o su cultura deben, en lo posible, evitarse.
- ▶ Insultos (blasfemia, obscenidad, groserías), sobre todo para menores de edad. En lo posible, se sugiere revisar las preguntas para detectar asuntos que puedan ser ofensivos para algunas personas o grupos, porque esto podría generar que algunos evaluados reaccionen emocionalmente de manera negativa. Sin embargo, si el constructo implica identificar un insulto u ofensa como una vulneración de un derecho (como es el caso en algunas de las preguntas de la pruebas de Competencias Ciudadanas), podría ser relevante incluirlo, siempre y cuando se mitigue el efecto

negativo que pueda tener sobre los evaluados. Para estos casos, se recomienda buscar fuentes externas, de manera que sea claro para los evaluados que lo que se evalúa no es una posición del Icfes ni de sus colaboradores.

- ▶ A menos que el constructo lo requiera, deben evitarse estereotipos (de género, religiosos, de grupo social, entre otros) tanto positivos como negativos. En algunas pruebas, como es el caso de la prueba de Competencias Ciudadanas, se puede hacer mención de estereotipos dado que una de sus tareas indaga por la identificación de prejuicios y estereotipos. No se debe asumir que todos los miembros de un grupo comparten las mismas actitudes o creencias, a menos que el grupo se organice alrededor de dichos atributos. Por ejemplo, sería incorrecto asumir que todas las personas que se denominan veganas son antisistema.

El siguiente ejemplo ilustra la aplicación de algunos de los criterios mencionados para reducir el contenido que pueda generar barreras afectivas:

Enunciado original	Enunciado revisado
<p>Como argumento a favor del vegetarianismo, una persona afirma que consumir carne es inmoral, porque el animal del que se obtiene sufre en el proceso: muchas veces el animal es desmembrado vivo y muere desangrado, luego de un largo suplicio. Otra persona, igualmente a favor del vegetarianismo, sostiene que el consumo de productos de origen animal es nocivo para la salud humana, pues puede producir enfermedades cardíacas; además, cree que las centrales de carnes, como la de Guadaxxx, deberían ser condenadas por eso.</p> <p>¿Cuál es la relación entre las posiciones de las dos personas?</p>	<p>María opina que consumir carne no es bueno, pues el animal del que se obtiene es sometido a un sufrimiento excesivo. Por otro parte, Camilo opina que el consumo de productos de origen animal es nocivo para la salud humana, pues puede producir enfermedades cardíacas.</p> <p>¿Cuál es la relación entre las opiniones de María y de Camilo?</p>
<p>Comentario:</p> <p>En esta pregunta se evalúa una competencia argumentativa; particularmente, la capacidad de establecer la relación entre dos posturas u opiniones. En el enunciado original hay una descripción gráfica del sufrimiento animal que es innecesaria para comprender la postura de la primera persona (identificada como María en el enunciado revisado para facilitar el acceso). Además de que la descripción detallada es innecesaria, puede generar un impacto afectivo negativo. Por otro lado, el enunciado original menciona el nombre de una entidad, lo cual también es innecesario para comprender la postura de la segunda persona (Camilo, en la versión revisada) y también puede generar un impacto afectivo no deseado en aquellas personas que conocen la entidad en mención o experiencias previas que la involucran.</p>	

1.3 Barreras físicas como fuente de sesgo

Las barreras físicas son cualquier elemento material, o conjunto de estos elementos, que restringen o dificultan el acceso perceptual a la prueba y por tanto pueden generar sesgo. Por ejemplo, los espacios que no están adecuados para personas con discapacidad motriz, o los formatos de prueba que no permiten acomodaciones para estudiantes ciegos o sordos, imponen barreras físicas.

Se sugiere evitar cualquier barrera física innecesaria para evaluar el desempeño. Esto no siempre es posible, por ejemplo, cuando se evalúan habilidades desarrolladas por los estudiantes en la pronunciación de un idioma extranjero, ya que el constructo impide que algunas personas con discapacidad puedan acceder a la prueba. Sin embargo, se deben evitar las siguientes barreras:

- a) Uso de imágenes con función decorativa o que dificulten la lectura.
- b) Imágenes y tamaños de fuente que no se adapten a las necesidades específicas de los evaluados. Así, se recomienda emplear una fuente con un tamaño mínimo de 14 puntos y, en caso de pruebas por computador, brindar la posibilidad al evaluado de aumentar este tamaño. De este modo, se facilitará el acceso a la prueba para aquellos estudiantes con baja visión.
- c) Impresión de texto vertical o entremezclado con imágenes que limiten al software que traduce el texto a lenguaje verbal.

Para ampliar las consideraciones sobre barreras físicas, se sugiere consultar el texto *Referentes teóricos del Proyecto de Inclusión de la evaluación estandarizada* (Icfes, 2020).

El siguiente ejemplo ilustra la aplicación de algunos de los criterios mencionados para reducir el contenido que genere barreras físicas:

Enunciado original	Enunciado revisado
 <p>Las industrias de la construcción y manufactura aportaron el 46 por ciento del Producto Interno Bruto (PIB) de China en 2012.</p> <p>Pekín y otras partes de China sufren de elevados niveles de contaminación. Algunos días, las autoridades cancelan las actividades deportivas en el exterior en colegios y guarderías y recomiendan a niños y ancianos que no salgan a la calle. Aunque se han cerrado centenares de fábricas en la capital y en otros puntos del país, la situación no mejora. Los expertos consideran que la única solución es que China reduzca de forma paulatina su dependencia de las industrias de la construcción y la manufactura.</p> <p>En esta situación, ¿cuáles dimensiones están en conflicto?</p>	<p>China sufre elevados niveles de contaminación. Algunos días, las autoridades cancelan las actividades deportivas en el exterior y recomiendan a niños y ancianos que no salgan a la calle. Aunque se han cerrado centenares de fábricas en varias partes del país, la situación no mejora. Los expertos consideran que la única solución es que China reduzca de forma paulatina su dependencia de las industrias de la construcción y la manufactura. La dependencia es tan grande que estas industrias aportaron el 46 por ciento del Producto Interno Bruto (PIB) de China en 2012.</p> <p>En esta situación, ¿cuáles dimensiones están en conflicto?</p>

Continúa

Comentario:

Esta pregunta evalúa una competencia de pensamiento sistémico al indagar por las diferentes dimensiones (política, económica, social, cultural, ambiental, entre otras) que se encuentran en conflicto en la situación descrita. En el enunciado original se hace un uso decorativo y, por tanto, innecesario de una imagen. Además, la información del pie de foto, aunque puede ser relevante para responder correctamente la pregunta, se encuentra en un tamaño de fuente que dificultaría el acceso a población con baja visión, lo que resulta en una barrea física. En el ítem revisado, se suprime la imagen decorativa y se incluye en el cuerpo del texto la información del pie de imagen.

2. Consideraciones especiales

En esta sección se exponen algunas consideraciones adicionales relacionadas con el tratamiento y representación de la diversidad, relevantes dada la gran diversidad étnica y cultural del país y la importancia de fomentar la inclusión en las pruebas del Icfes. Adicionalmente, se mencionan aspectos de la revisión de sesgo que se aplican especialmente a las evaluaciones de menores de edad.

2.1 Representación de la diversidad

Las recomendaciones de este apartado van en dos vías: a) lograr evaluaciones equitativas en las que se tenga en cuenta la diversidad de la población a la que va dirigida y b) incluir en las evaluaciones contenido relacionado con grupos específicos de nuestra sociedad; por ejemplo, la representación de género, etnia y población con discapacidad.

La diversidad puede entenderse como los diferentes valores, actitudes, perspectivas culturales, creencias, origen étnico, nacionalidad, orientación sexual, identidad de género, aptitud, salud, estatus social, habilidades y otras características personales específicas (ACNUR, 2018). Aunque desarrollar evaluaciones que apunten a un reconocimiento de la diversidad o a la inclusión de los múltiples grupos puede resultar técnicamente difícil dado su gran número, es necesario diseñar evaluaciones que apunten a una equidad en su contenido para que la población que las responde se enfrente a un instrumento adecuado, equilibrado e incluyente.

2.1.1 La diversidad en el contenido de las evaluaciones

Para las evaluaciones de carácter nacional, tener en cuenta la diversidad de la población implica considerar varios aspectos que van desde el diseño, construcción, revisión, hasta la validación de las preguntas. Con respecto al diseño, desde la definición del constructo de evaluación y de la estructura de la prueba, se procura que los conocimientos, habilidades, destrezas y otros rasgos sean apropiados para obtener una información válida de toda la población, sin privilegiar grupos particulares por aspectos como tipo de educación, acceso a ciertos recursos económicos o tecnológicos, experiencia en contextos particulares, etc.

Un aspecto importante para lograr instrumentos de evaluación que tengan en cuenta la diversidad de la población es contar con constructores de diferentes partes del país y que pertenezcan a diferentes grupos demográficos.

Con respecto a la revisión y validación de preguntas, se espera que participen personas pertenecientes a diversos grupos demográficos con amplio conocimiento tanto en el tema que se está evaluando como en la población a la que va dirigida la prueba. Dentro de los criterios de revisión está que ninguna de las preguntas lleve a una interpretación incorrecta sobre un grupo de personas, que se refuerce algún estereotipo, se atribuya características de manera equivocada, se mencione un grupo particular sin ser relevante para el constructo que se evalúa, etc. En esta parte del proceso de construcción, de

ser posible, se pueden implementar estrategias de validación con grupos pequeños poblacionales del grado evaluado. Para ello, se pueden realizar entrevistas con protocolos dirigidos a reconocer qué piensan o saben los evaluados del constructo evaluado y qué procesos de pensamiento ejecutan para responder las preguntas. Algunos de los elementos que permiten una adecuada representación de la diversidad en una prueba son:

- ▶ **Etnia.** De acuerdo con el Ministerio de Salud y Protección Social (s.f.) los grupos étnicos son poblaciones cuyas condiciones y prácticas sociales, culturales y económicas, los distinguen del resto de la sociedad, ya que han mantenido su identidad a lo largo de la historia. Como sujetos colectivos que aducen un origen, su historia y características culturales están dadas en sus cosmovisiones, costumbres y tradiciones. Dentro de los grupos étnicos colombianos se encuentran los pueblos indígenas, población negra o afrocolombiana, población raizal, pueblo palenquero y pueblo Rom (gitano). Se debe procurar incluir contenido relacionado con estos grupos siempre y cuando la temática lo permita, evitando atribuir características que resulten imprecisas o que generen estereotipos. Por la especificidad y variedad de los grupos étnicos se recomienda evitar construir preguntas en las que los evaluados deban tener un conocimiento sobre estos y que no sea necesario para el constructo que se está evaluando.
- ▶ **Orientación sexual y diversidad de género.** Si es relevante para lo que está evaluando una prueba, es válido incluir

contenidos en los que se haga referencia a la orientación sexual de las personas (bisexuales, gays o lesbianas, entre otros). En caso de hacer referencia a la orientación sexual o diversidad de género, es importante revisar que los contextos utilizados no refuercen estereotipos de género ni validen conductas discriminatorias, basadas en el género o la orientación sexual.

- ▶ **Personas con discapacidad.** Si se considera que mencionar la discapacidad o a personas con discapacidad es relevante para el constructo, se debe evitar reforzar estereotipos. Se sugiere no enfocar los contextos relacionados con esta población únicamente a temas como la limitación, necesidad de ayuda, enfermedad, rechazo, rezago escolar, etc.

Un aspecto fundamental de la representación de la diversidad es usar la terminología adecuada para los distintos grupos a los que se decida mencionar. Si se tienen dudas sobre la denominación sobre un grupo particular o aspectos asociados a estos, se recomienda consultar documentos oficiales o de fuentes confiables, relacionados con el tema. Por ejemplo, para referirse a personas con discapacidad se hablaba de personas en situación o en condición de discapacidad, de persona ciega, sorda, etc. A partir del año 2006, con lo establecido en la convención sobre los derechos de las personas con discapacidad, en la cual se regulan y hacen explícitos los derechos de las **personas con discapacidad**, se reconoce como denominación común y universal personas con discapacidad (visual, auditiva, intelectual, etc.). Conocer

esto resulta importante para evitar que la población a la que se haga referencia considere que existe un desconocimiento sobre ellos y que, en general, se brinda información errónea sobre estos grupos.

2.1.2 Pertinencia de la diversidad en el contenido de los ítems

Es necesario tener en cuenta la diversidad en el contenido de las preguntas, por un lado, para generar reconocimiento y representatividad de todas las poblaciones y grupos sociales y, por otro lado, para no generar ventajas ni desventajas injustas derivadas del conocimiento irrelevante para el constructo que puedan tener los estudiantes dada su pertenencia a un grupo poblacional específico.

Respecto al reconocimiento y la representatividad debe considerarse que, si una prueba incluye referencias a personas o grupos, los examinados deberían, idealmente, poder ver a personas similares a ellos o personas de su mismo grupo (Zieky 2013, p. 299). Los estudiantes no deben sentirse excluidos de la prueba por falta de inclusión de ningún miembro de su grupo. Por tanto, una prueba ideal debería incluir diversos grupos relevantes en la población que realiza el examen. Si bien no es factible este tipo de inclusión en su totalidad, es necesario procurar representar la diversidad cuando se mencionan o muestran personas. Por ejemplo, las mujeres pueden sentirse excluidas si todas las personas mencionadas son hombres. En consecuencia, se recomienda tener en cuenta los siguientes aspectos:

- ▶ En cuanto lo permita el tema, los miembros de todos los grupos incluidos en las pruebas deben representar una variedad de roles sociales. Por ejemplo, en el contenido de las preguntas debe evitarse que todos los gerentes sean hombres y todas las asistentes mujeres. Similarmente, debe evitarse la redacción de preguntas en la que las personas pertenecientes a un grupo étnico aparezcan con roles exclusivos, por ejemplo, asociados siempre a labores u oficios propios de la ruralidad.
- ▶ Evitar ventajas y desventajas injustas (ver página 8). Se sugiere reducir la posibilidad de que algún grupo poblacional tenga mayor conocimiento sobre algún tema específico, irrelevante para el constructo. Así mismo, debe evitarse que una pregunta pueda ser más difícil para un grupo poblacional porque este carece de un conocimiento puntual. En el caso de utilizar contextos, situaciones, usos o costumbres de una comunidad específica, los conceptos deben ser explicados de tal manera que se impida una interpretación basada en una visión, ritual, creencia o valores estéticos propios de un grupo.
- ▶ Revisar que la representación de la diversidad no esté sesgada, sea paternalista o condescendiente sobre algún grupo poblacional a pesar de que esta incluya cosmovisiones, roles sociales, ideologías, perfiles socioeconómicos o creencias. Es decir, la inclusión de grupos diversos no debería estar cargada de connotaciones negativas sobre las condiciones o los contextos particulares. Adicionalmente, debe contemplarse que estas connotaciones pueden

convertirse en barreras afectivas. Así, por ejemplo, se debe evitar encasillar o asociar de forma generalizada a las comunidades y contextos rurales con condiciones de pobreza extrema, inequidad de género o violencia.

2.2 Contenido apropiado para la evaluación de menores de edad

Las pautas para esta población se centran en evitar contenidos que puedan molestar o asustar a las niñas, los niños y adolescentes (NNA), o que puedan ofender o excluir a alguno de estos grupos o fomentar un modelo de comportamiento inapropiado. Las consideraciones adicionales para las niñas, los niños, los adolescentes y los jóvenes son, mayormente, extensiones del segundo principio general de revisión de la equidad: evitar las fuentes afectivas de sesgo. A pesar de que algunas de estas pautas se aplican también a pruebas presentadas por adultos, se recomienda una revisión más estricta en el caso de pruebas aplicadas a menores de edad.

Aunque las pautas para evaluar a los menores de edad son estrictas, se debe contemplar cualquier material necesario para una medición válida. Por ejemplo, las pruebas de estudios sociales pueden incluir el tema de la esclavitud, a pesar de que este podría excluirse de una prueba de lectura y comprensión (Zieky 2013, p. 301). Algunos de los temas que se sugieren evitar, especialmente en pruebas aplicadas a menores de edad, son:

- ▶ Materiales perturbadores, como enfermedades graves, divorcio, violencia doméstica, pérdida de empleo de los padres, disputas entre familiares, peleas, acoso o matoneo, suicidio, violencia sexual, muertes violentas, maltrato animal y aborto.
- ▶ Plagas (por ejemplo, ratas, piojos) y animales peligrosos (como serpientes venenosas, escorpiones). Estos temas son generalmente inapropiados en las pruebas para los estudiantes porque no son centrales al constructo y los acerca a sus posibles temores personales.
- ▶ Comportamientos inapropiados. Evitar temas y acciones como groserías, insultos, juegos de azar y apuestas, fumar, beber alcohol y consumir drogas ilegales. Así mismo, se sugiere excluir contenidos que validen modelos de comportamiento como mentir, hacer trampa, robar, evadir clases o hacer cosas peligrosas que puedan atentar contra su integridad o la de otros.
- ▶ Celebraciones que pueden ser lejanas de su contexto social y cultural por incluir creencias personales, usos y costumbres, creencias religiosas, o festividades comerciales. Por ejemplo, festividades como el Día de las Brujas, Semana Santa, Navidad, Día del Amor y la Amistad, entre otros.
- ▶ También deben evitarse constructos irrelevantes que referencien cualquier deidad o religión.

- ▶ Las pruebas no pueden tener una visión cínica o sarcástica sobre los valores que podrían considerarse socialmente aceptados como la honestidad, solidaridad, entre otros (ETS 2015, p. 42).

Conclusión

Los lineamientos para la revisión de equidad, como se indicó en la introducción, no son reglas fijas de construcción de ítems, pues su aplicación depende de al menos dos factores: la prueba revisada y el contexto o tiempo de revisión. Con respecto al primer factor, como se sugirió, los lineamientos deben ser aplicados considerando el marco de referencia de la prueba, que recoge la definición del constructo de cada prueba. Con respecto al segundo factor, el contexto y tiempo son determinantes para la elaboración de los lineamientos, pues cambios contextuales (culturales, sociales, políticos) pueden obligar una revisión de los lineamientos que implique matizarlos, cambiarlos o agregar nuevos. A pesar de que no son reglas fijas, la observación de estos lineamientos permitirá reducir las fuentes de sesgos que afectan negativamente a los puntajes de la población evaluada y la validez de las interpretaciones de dichos puntajes.

Referencias

Alto Comisionado de las Naciones Unidas para los Refugiados (ACNUR). (2018). *Política de edad, género y diversidad*. Recuperado de: <https://www.refworld.org/es/pdfid/5b48f83c4.pdf>

American Educational Research Association, American Psychological Association y National Council on Measurement in Education (AERA, APA, & NCME) (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Educational Testing Service (2009). ETS international principles for fairness review of assessments: A manual for developing locally appropriate fairness review guidelines in various countries. Princeton, NJ. Recuperado de: <http://www.ets.org/about/fairness>

Educational Testing Service (ETS) (2015). ETS Guidelines for Fair Tests and Communications. Princeton, NJ. Recuperado de: https://www.ets.org/s/about/pdf/ets_guidelines_for_fair_tests_and_communications.pdf

Freepik (2021). *Manos de personas sosteniendo maqueta de papel rasgado gris*. Recuperado de: https://www.freepik.es/foto-gratis/manos-personas-sosteniendo-maqueta-papel-rasgado-gris_12193232.htm#page=1&query=diversidad&position=40

Icfes (2018). Guía introductoria al Diseño Basado en Evidencias del ICFES, Bogotá: Icfes.

Icfes (2020). *Referentes teóricos del Proyecto de Inclusión de la evaluación estandarizada*. Bogotá: Icfes.

Icfes (2020a). Guía de Orientación Saber 11 2021-1. Bogotá: Icfes.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational Measurement* (pp. 13-103). New York, NY: Macmillan.

Messick, S. (1995). Standards of Validity and the Validity of Standards in Performance Assessment. *Educational Measurement: Issues and Practices*. Vol. 14, 4. pp. 5-8.

Ministerio de Educación Nacional (2017). Documento de orientaciones técnicas, administrativas y pedagógicas para la atención educativa a estudiantes con discapacidad en el marco de la educación inclusiva. Bogotá: Autor.

Ministerio de Salud y Protección Social. (s.f.). Promoción Social. Grupos étnicos. <https://www.minsalud.gov.co/proteccionsocial/promocion-social/Paginas/grupos-etnicos.aspx>

Wendler, C. y Burrus, J. (2013). The Importance of Editorial Reviews in Ensuring Item Quality. En: K. Geisinger (Ed.), *APA handbook of testing and assessment in psychology: Vol. 1. Test theory and Testing and assessment in industrial and organizational psychology* (pp. 293-302). Washington, DC: American Psychological Association.

Zieky, M. J. (2013). Fairness review in assessments. En: K. Geisinger (Ed.), *APA handbook of testing and assessment in psychology: Vol. 1. Test theory and Testing and assessment in industrial and organizational psychology* (pp. 293-302). Washington, DC: American Psychological Association.

Calle 26 N.º 69-76, Torre 2, Piso 15, Edificio Elemento, Bogotá, D. C., Colombia • www.icfes.gov.co
Líneas de atención al usuario: Bogotá Tel.: (60+1) 484-1460 PBX: (60+1) 484-1410 -
Gratuita nacional: 018000-519535