

Encuesta 2014
Observatorio de resultados de
aprendizaje

**MANUAL DE INSTRUCCIONES
PARA CUMPLIMENTAR EL CUESTIONARIO
DEL CATÁLOGO DE EVALUACIONES DE
APRENDIZAJE**

ÍNDICE DE CONTENIDOS

1. INTRODUCCIÓN	4
2. COBERTURA DEL CUESTIONARIO	6
3. CÓMO CUMPLIMENTAR ESTE CUESTIONARIO	7
4. INSTRUCCIONES PARA CADA PREGUNTA.....	8
SECCIÓN 2: LISTA DE EVALUACIONES.....	8
SECCIÓN 3: ALCANCE, OBJETIVOS, FINANCIACIÓN Y PARTES INTERESADAS	12
SECCIÓN 4: DISEÑO Y ADMINISTRACIÓN DE LA PRUEBA	14
SECCIÓN 5: COBERTURA Y DISEÑO MUESTRAL.....	16
SECCIÓN 6: PROCESAMIENTO DE DATOS	18
SECCIÓN 7: MEDICIONES Y RESULTADOS.....	19
SECCIÓN 8: DIVULGACIÓN, PRESENTACIÓN Y ACCESIBILIDAD DE LOS DATOS	22
5. GLOSARIO Y DEFINICIONES	23

Notas técnicas para la cumplimentación del cuestionario

1) Notas sobre la codificación de datos que falten

El correcto uso de los códigos para la información que falta es imprescindible para garantizar la integridad de los datos. En los análisis e informes estadísticos deben señalarse las razones de que haya datos no disponibles.

No deje vacía ninguna celda prevista para ingresar una respuesta. Deberá asignarse uno de los siguientes cinco códigos a las celdas sin datos o con datos no válidos:

i) Categoría no aplicable = **a**

Asigne el código «**a**» a las celdas o categorías que no sean aplicables a una determinada evaluación o examen público. Esto implica que no existen datos correspondientes a estas categorías o celdas.

ii) Valor nulo = **n**

Asigne el código «**n**» a las celdas cuyos valores (o cantidades) sean nulos o insignificantes. De este modo, para indicar un valor nulo, utilice el código «**n**» en lugar de 0 (cero).

iii) Datos incluidos en otras categorías = **x**

Asigne el código «**x**» a las celdas cuyos datos no puedan ser desglosados. Si procede, **indique** en el espacio para metadatos la categoría en la que están recogidos dichos datos.

iv) Datos no disponibles = **m**

Asigne el código «**m**» a las celdas o categorías en las que faltan datos o no están disponibles, cuando tales datos no aparezcan en ninguna otra celda del cuestionario (incluso aunque, en principio, se hayan podido recoger esos datos).

iv) Datos confidenciales = **c**

Asigne el código «**c**» a los datos confidenciales que el país prefiera no hacer públicos.

2) Nota relativa a datos provisionales o estimados = *

Por favor, anote las cifras provisionales o estimadas precedidas de un asterisco (*).

3) Nota relativa a datos parciales = p

Por favor, anote las cifras parciales precedidas del citado código (p).

1. INTRODUCCIÓN

El Instituto de Estadística de la UNESCO (UIS) ha lanzado una nueva iniciativa a nivel mundial para identificar las distintas evaluaciones realizadas por cada país para hacer un seguimiento de los logros de sus estudiantes. Las consultas a nivel mundial en materia de educación han puesto de manifiesto que, además de garantizar un acceso igualitario a una educación de calidad, el aprendizaje debe tener su lugar en la agenda de desarrollo post- 2015. En lo que a esta área del aprendizaje se refiere, se han identificado importantes competencias y destrezas en distintos áreas que todos los niños y jóvenes deberían adquirir con el fin de ser productivos y poder hacer frente a los desafíos que les deparen sus vidas. En la mayoría de los países son comunes las evaluaciones nacionales, que pueden ser de alto o bajo impacto (“high-stake” y “low-stake”)¹. Documentar, analizar y divulgar la gran cantidad de conocimiento derivado de estas evaluaciones es uno de los elementos clave del Catálogo de Evaluaciones de Aprendizaje del UIS, que publicará los contenidos de éste a través de su base de datos internacional. Información adicional sobre esta iniciativa está disponible en [la página web del UIS](#).

El Catálogo constituirá un inventario global de todas las evaluaciones nacionales e internacionales estandarizadas a gran escala y de los exámenes públicos nacionales que los países utilizan para medir y monitorear los resultados de aprendizaje en la educación primaria y la secundaria baja (niveles 1 y 2 de la CINE 2011). El Catálogo se construirá a partir de la información recopilada en este cuestionario, diseñado para capturar las características de las evaluaciones de alumnos mencionadas. El cuestionario recopila los datos sobre las evaluaciones nacionales más recientes administradas en el país entre 2000 y 2014, quedando las implementadas antes de 2000 excluidas.

Productos. El conjunto de datos relativos a las evaluaciones de estudiantes recogido en el cuestionario pasará a formar parte de la base de datos internacional del UIS y quedará, por lo tanto, a disposición pública. La información del Catálogo se combinará con los datos ya disponibles en la base de datos del UIS, lo que dará como resultado informes resúmenes nacionales sobre los esfuerzos del país por medir y monitorear los logros de sus estudiantes. Incluirá una lista de evaluaciones (nacionales e internacionales, de alto y bajo impacto), dominios incluidos y el nivel de aprendizaje que están alcanzando los alumnos en dichos campos. El informe nacional aportará información descriptiva clave sobre estas evaluaciones e indicará las diferencias y similitudes entre determinados elementos de la evaluación sin comparar la calidad de las evaluaciones ni al interior de los países ni entre éstos.

Resultados esperados. Los resúmenes nacionales podrán utilizarse para posibilitar diálogos informados sobre el uso de las evaluaciones entre las distintas partes interesadas. Los gobiernos e instituciones locales podrían usar los resúmenes como información de base relevante para revisar las características de sus evaluaciones en función de objetivos específicos. La comunidad de desarrollo, como las organizaciones bilaterales y multilaterales que aporten fondos, podría ver estos informes nacionales como una herramienta más con la que evaluar las propuestas a favor de evaluar los resultados del aprendizaje. El UIS desempeñará un papel activo promoviendo el uso de estos informes nacionales y la información de la base de datos internacional; por ejemplo, para que sirva como base a los países que quieran revisar la cobertura y el alcance de las evaluaciones de su alumnado.

Beneficios para los países. Las estimaciones preliminares del UIS muestran que, actualmente, más de 150 países efectúan mediciones de los niveles de aprendizaje de sus estudiantes por medio de evaluaciones y exámenes nacionales, o bien participan en iniciativas de evaluación regionales o internacionales. Por el momento no se ha realizado un esfuerzo a nivel mundial para recabar información de manera sistemática sobre las decisiones que toman los países respecto de sus evaluaciones y cómo estas se relacionan con la agenda de los resultados de aprendizaje. El Catálogo resolverá este faltante de información centralizando la enorme cantidad de datos de evaluación de la que ya se dispone y elaborando informes resúmenes nacionales. Además, el UIS llevará a cabo nuevas investigaciones que podrán ser de utilidad a quienes toman las decisiones a nivel nacional; estas consistirán, por ejemplo, en estudiar y documentar nuevas tendencias y patrones en los distintos países, aunque sin valorar su calidad ni comparar los resultados de diversos países.

La utilidad para los gobiernos de revisar las características de las evaluaciones de los alumnos puede explicarse por distintos motivos, como

¹ Los términos “low-stake assessments” y “high-stake assessments” se traducen como “evaluaciones de bajo impacto” y “evaluaciones de alto impacto”.

- (i) por el hecho de que las evaluaciones de calidad suelen ser caras sobre una base per cápita; por ello, podría ser deseable examinar la utilidad y el valor añadido de cada evaluación;
- (ii) que la confianza de la población en las pruebas pudiera verse alterada por la percepción pública de la equidad de éstas, lo que a su vez puede vincularse a la calidad de estas pruebas; y
- (iii) que los países quieran alinear sus evaluaciones con los estándares internacionales. Por lo tanto, los resultados del catálogo pueden ser un recurso útil para tales evaluaciones.

Como parte de su estrategia para difundir los informes y estudios nacionales de su Catálogo, el UIS pretende ofrecer a los países la oportunidad de unirse a comunidades de intercambio de prácticas en las que los gobiernos podrán compartir sus experiencias en el seguimiento de los resultados de aprendizaje. Se espera que las evaluaciones nacionales cobren especial protagonismo. Estas comunidades de intercambio de prácticas podrían asimismo convertirse en foros de debate sobre los sistemas nacionales y los estándares globales emergentes.

El Catálogo no comparará el desempeño de los alumnos. Los resultados del desempeño de los alumnos no son necesariamente comparables, particularmente si se parte de evaluaciones nacionales que no han sido diseñadas para posibilitar una comparación entre países. Sin embargo, la comparabilidad no debería ser tomada como requisito en el análisis de los resultados de las evaluaciones. Los datos no comparables son valiosos para realizar un seguimiento de los resultados del aprendizaje a nivel nacional e internacional.

Independientemente de la comparabilidad, la documentación relativa a las evaluaciones de alumnos extraída del conjunto de bases de datos ha demostrado su capacidad de despertar interés por las causas de las diferencias entre países. Cabe señalar que la causa de estas diferencias podría encontrarse en numerosos factores que no resultan evidentes a primera vista.

Si bien los resultados de desempeño de los alumnos no pueden compararse en todos los casos, sus características incluyen, entre otras cosas: cursos, enfoques, áreas de contenido, idiomas y el modo en que se utilizan los resultados en el diseño de políticas. Los análisis descriptivos de ciertos elementos de las evaluaciones revelan patrones en las decisiones que los gobiernos toman para supervisar los resultados del aprendizaje. Con el fin de documentar estas tendencias y patrones globales, el catálogo realizará comparaciones entre países basándose en los atributos descriptivos de sus evaluaciones.

El presente **manual de instrucciones** se ha redactado para ayudar a los proveedores de datos a cumplimentar el cuestionario del Catálogo de Evaluaciones de Aprendizaje. Para más información sobre cómo cumplimentar este cuestionario, póngase en contacto con: Sr. Georges Boade mediante el correo electrónico g.boade@unesco.org; o por teléfono en el 1-514-343-7845.

2. COBERTURA DEL CUESTIONARIO

Tipo de evaluaciones incluidas en este estudio:

El estudio se centra en las evaluaciones educativas a nivel del sistema escolar dirigidas a estudiantes matriculados en programas de educación formal. En este cuestionario solo se incluirán los datos de evaluaciones nacionales a gran escala, incluidos los exámenes públicos nacionales (de salida o fin de estudios) administrados en los niveles CINE 1 y 2 de programas educativos generales o técnicos/profesionales. Para más información sobre los niveles y cursos de la CINE en su país, consulte [los mapas de la CINE 97 y el marco de la CINE 2011](#). Las evaluaciones internacionales en las que participan los países deberían quedar recogidas en la sección 2.3, sin embargo la información de éstas se obtendrán directamente de las agencias organizadoras de las citadas evaluaciones; en otras palabras, los responsables nacionales no deberán registrar las características de las evaluaciones internacionales en este cuestionario. Algunos ejemplos de evaluaciones internacionales son: LLECE, TIMSS, PIRLS, PASEC y SACMEQ.

Tipos de evaluaciones excluidas de este estudio:

En este estudio no se incluirán tres tipos de evaluaciones:

1. Evaluaciones administradas en las escuelas u organizadas por instituciones educativas descentralizadas cuyo objetivo principal sea la selección de alumnos que pasan de un nivel a otro, o de un programa a otro: Esto incluye:
 - a. pruebas que suelen organizarse al final de la escuela primaria con el objetivo de seleccionar buenos alumnos para que accedan al primer año de educación secundaria general o de un ciclo técnicos/profesional;
 - b. pruebas destinadas a la selección de alumnos en centros de formación profesional;
 - c. evaluaciones organizadas por centros de formación profesional.
2. Pruebas que no están especialmente dirigidas a estudiantes inscritos en programas educativos y que se administran a todos los candidatos que solicitan una certificación profesional y que en general no han estado matriculados previamente en un programa educativo general o profesional equivalente. Este tipo de pruebas está descentralizado y puede ser organizado por sindicatos o asociaciones profesionales. Algunos ejemplos de ello son los exámenes de obtención del permiso de conducir o los que tienen como objetivo convertirse en miembro de una asociación profesional.
3. Las evaluaciones de resultados del aprendizaje efectuadas en los hogares, aunque los niños evaluados estén matriculados en programas de nivel CINE 1 o 2.

3. CÓMO CUMPLIMENTAR ESTE CUESTIONARIO

Rogamos a los participantes que lean con atención las siguientes instrucciones, definiciones y ejemplos para cumplimentar adecuadamente el cuestionario del Catálogo de Evaluaciones de Aprendizaje del UIS. Este manual incluye todas las definiciones de las principales variables utilizadas en el cuestionario y explica qué tipo de datos se solicita en cada ítem o tabla.

El cuestionario cuenta con ocho secciones. Al final de cada sección hay un ítem llamado «metadatos». Este espacio se ha incluido para que los encuestados incluyan explicaciones o aclaraciones específicas relativas a esa sección, lo que permitirá interpretar los datos solicitados.

Si en varias preguntas se solicita la misma información, por favor asegúrese de que los datos facilitados son consistentes entre sí. En los casos en que una contradicción sea inevitable, le rogamos que la explique claramente en el campo para metadatos que hay al final de cada sección.

Por favor, no deje vacía ninguna celda en la que se espere respuesta. Es imprescindible utilizar correctamente los **códigos** para las celdas sin datos con el fin de garantizar la integridad de los mismos. En los análisis e informes estadísticos del UIS se deben indicar los motivos por los que no se han facilitado datos en cada instancia.

El UIS acepta que los países hagan sus propias estimaciones si se enfrentan a datos incompletos o faltantes. Para indicar que una celda contiene datos estimados, coloque un asterisco (*) delante de la cifra. Ej.: *68794. No deje ningún espacio entre el símbolo y la cifra. Si a pesar de los esfuerzos sigue habiendo datos no disponibles o incompletos, incluya por favor una aclaración en una nota aparte o en el campo para metadatos del final de la sección.

Aparte de las explicaciones o aclaraciones que deban incluirse en las secciones de metadatos, le rogamos que use este espacio para citar cualquier material de referencia, manual, publicación o sitio web al que haya hecho alusión y que pueda ser de utilidad para entender los datos de su país.

Acuda al **Glosario** que se encuentra al final de este manual para consultar definiciones y explicaciones de los datos con los que debe cumplimentarse este cuestionario.

Además, tenga en cuenta que el estudio pretende recopilar datos de evaluaciones y exámenes públicos realizados entre 2000 y 2014. Por favor, anote siempre el año en que facilitó los datos han sido reportados.

4. INSTRUCCIONES PARA CADA PREGUNTA

SECCIÓN 2: Lista de evaluaciones

Presente cada evaluación una sola vez, aunque incluya más de un curso o asignatura. En el caso de que existan diferentes pruebas bajo un solo nombre, preséntelas como una única evaluación, independientemente del número de cursos, asignaturas o idiomas que comprenda. A menudo, las evaluaciones abarcan más de una asignatura o constructo por curso. Mientras que las pruebas de los distintos cursos reciban el mismo nombre, estas versiones deberían considerarse como partes integrantes de una misma evaluación. Por ello, si una evaluación se administra un mismo año a más de un curso, deberá mencionarse una sola vez. La información adicional quedará recogida en las siguientes secciones.

Este estudio se centra en evaluaciones a gran escala administradas a alumnos de escuela primaria y de primer ciclo de educación secundaria general/técnica de su país. Abarcan todo el país y se dirigen a todos o a una muestra de alumnos con rasgos comunes, como el curso en el que se encuentran o la edad.



Por favor, diríjase a las páginas anteriores de este manual para obtener más información sobre los **tipos de evaluación en los que se centra este estudio**, y al glosario para ver las definiciones de las **evaluaciones de bajo y alto impacto** antes de cumplimentar esta sección.



2.1 Lista de evaluaciones nacionales a gran escala

Esta tabla pretende recopilar información general sobre todas las evaluaciones nacionales a gran escala.

Introduzca la siguiente información sobre cada evaluación según el orden cronológico de su administración, empezando por la más reciente:

- i) sigla en el **idioma original** si existe (las siglas no deben traducirse nunca). Asígnele el código «a» si tal sigla no existe.
- ii) nombre completo de la evaluación:
 - Junto a la letra N, en su idioma original
 - Junto a la letra I, traducción al inglés disponible en el caso que el idioma original no es inglés.
- iii) el/los programa/s educativo/s que abarca la evaluación. Elija de la siguiente lista:
 - solo educación primaria (CINE 1)
 - solo el primer ciclo de educación secundaria (CINE 2)
 - educación primaria y primer ciclo de secundaria (CINE 1 y 2)
 - primer y segundo ciclo de educación secundaria (CINE 2 y 3)
 - educación primaria, primer y segundo ciclo de la educación secundaria (CINE 1, 2 y 3)
- iv) nombre de las instituciones responsables de la administración de la evaluación.
- v) el impacto de cada evaluación nombrada, ya sea alto o bajo (“high-stake” o “low-stake”), para las siguientes poblaciones: alumnos, profesores y escuelas. Tenga en cuenta que en algunos casos la repercusión de las evaluaciones es informal. De ser posible, incluya aclaraciones en el campo para metadatos.
- vi) el último año (aaaa) en que se administró la evaluación.
- vii) en el caso de que su país tenga una estructura administrativa federal en la que ninguna evaluación nacional abarque a todos los estados, provincias o regiones (dependiendo de la terminología que se use en su país), marque «Sí» en la celda roja que hay bajo esta tabla. Esto implica que la tabla 2.1 quedará vacía, ya que no existe una evaluación nacional a gran escala.

Añada en el campo de observaciones de la sección 2.1. cualquier observación/aclaración que deba tenerse en cuenta junto a los datos brindados con el fin de ofrecer una imagen lo más completa y exacta posible.



2.2 Lista de exámenes públicos nacionales

Esta tabla pretende recopilar información general sobre los exámenes nacionales públicos. Son válidas las instrucciones de 2.1; por ello, consulte las instrucciones detalladas en el apartado 2.1.

En el caso de que su país tenga una estructura administrativa federal en la que ningún examen nacional público abarque a todos los estados, provincias o regiones (dependiendo de la terminología que se use en su país), marque «Sí» en la celda roja que hay bajo esta tabla. Esto implica que la tabla 2.2 quedará vacía, ya que no existe un examen público nacional.

Añada en el campo de metadatos de la sección 2.2. cualquier observación/aclaración que deba tenerse en cuenta junto a los datos brindados con el fin de ofrecer una imagen lo más completa y exacta posible.



2.3 Lista de evaluaciones internacionales

Esta tabla pretende recopilar información sobre las evaluaciones internacionales que se han administrado en su país. Esta categoría incluye evaluaciones a nivel mundial o cualquier otra evaluación administrada en dos o más países.

En el caso de que su país tenga una estructura administrativa federal y solo algunos de los estados, provincias o regiones (dependiendo de la terminología que se use en su país) hayan participado en las evaluaciones internacionales, marque «Sí» en la celda roja que hay bajo esta tabla. Esto implica que la tabla 2.3 quedará vacía, ya que ninguna evaluación internacional habrá sido administrada a nivel nacional, sino únicamente en ciertos estados, provincias o regiones.

Escriba en el campo de observaciones de la sección 2.3. cualquier observación/aclaración que deba tenerse en cuenta junto a los datos que facilite con el fin de ofrecer una imagen lo más completa y exacta posible.

Por favor, tenga en cuenta que los puntos de enlace del país son los responsables de cumplimentar las secciones de la 3 en adelante, que tratan sobre las evaluaciones nacionales y los exámenes públicos de su país. Los datos sobre las evaluaciones internacionales se recabarán directamente en las instituciones encargadas de desarrollar estas evaluaciones.



2.4 Existencia de evaluaciones a gran escala de la Educación de la Primera Infancia (CINE nivel 0)

El objetivo de esta subsección es destacar evaluaciones a gran escala en educación de la primera infancia o preescolar que existan en el país. Si selecciona **Sí**, lo que implicaría que efectivamente se realizan evaluaciones de los resultados del aprendizaje antes de primaria en su país, por favor indique el nombre de cada una de ellas, la edad de los niños evaluados, su cobertura (nacional o regional) y las instituciones encargadas de su administración.

Añada cualquier observación/aclaración que pueda completar los datos facilitados en el campo de observaciones de la sección 2.4.

2.5 Existencia de evaluaciones de los resultados del aprendizaje realizadas en los hogares

El objetivo de esta subsección es destacar las evaluaciones a gran escala de los resultados del aprendizaje realizadas en los hogares de su país.

Las evaluaciones de los resultados del aprendizaje realizadas en los hogares suelen ser realizadas por entrevistadores que visitan los hogares para recopilar información sobre la situación familiar y evaluar las aptitudes cognitivas de uno o varios de sus miembros.

Por lo tanto, no han de incluirse aquí evaluaciones realizadas en las escuelas en las que los niños llevan cuestionarios a casa para sus padres. En esos casos, este cuestionario es un instrumento más para realizar las evaluaciones enumeradas en las subsecciones de la 2.1 a la 2.3.

Las evaluaciones de los resultados del aprendizaje realizadas en los hogares no se centran únicamente en alumnos inscritos en programas formales, sino que suelen incluir a niños no escolarizados y pueden incorporar otros criterios de selección, como la edad de los niños.



A partir de la sección 3 se solicitará información que ayude a caracterizar cada una de las evaluaciones nacionales y exámenes públicos enumerados en la sección 2, solamente en las subsecciones 2.1 y 2.2.

Los datos de las evaluaciones internacionales (enumeradas en la subsección 2.3) serán cumplimentados por la institución responsable de su administración.

Antes de empezar a cumplimentar los datos de estas secciones, recuerde hacer tantas copias de este cuestionario como evaluaciones haya en las secciones 2.1 y 2.2. Para identificar los distintos archivos fácilmente, asegúrese de que cada copia incluye el nombre del país seguido del acrónimo de la evaluación si existe (en caso contrario, elija un nombre con el que distinguir claramente la evaluación). Por ejemplo: Sri Lanka-GCE-O-L o France-CAP.

A continuación empiece a cumplimentar las siguientes secciones para cada evaluación

SECCIÓN 3: Alcance, objetivos, financiación y partes interesadas

Esta sección recoge la información que caracteriza a cada evaluación. Incluye datos sobre su alcance, objetivos, fuentes de financiación y papel de cada parte interesada.

Las celdas verdes ofrecen una lista de opciones entre las que elegir, por lo que no es necesario escribir las respuestas manualmente.

En los casos en los que se solicite especificar una categoría no incluida en la lista <Otro. Especifique>, una **celda amarilla** le indicará donde introducir esa categoría.

Si para comprender estos datos hicieran falta comentarios adicionales de su parte, escríbalos en el campo para metadatos disponible al final de la sección.

Para cumplimentar esta sección es posible que necesite revisar la documentación oficial más reciente al respecto o archivos como los marcos de la evaluación, los informes o las bases de datos.



3.1. Información general sobre la evaluación o examen público

Por favor, anote el nombre completo de la evaluación o examen en inglés y en el idioma nacional u oficial de no ser el inglés, así como su acrónimo, tal y como se nombró en la sección 2. Si no existe un acrónimo, asigne el código «a» al ítem 3.1.3.



Consulte en el glosario las definiciones de las **evaluaciones con repercusiones elevadas y reducidas**.



3.2. Alcance de la evaluación

Es posible que una evaluación haya sido concebida para cubrir varias categorías en cada variable. Por ejemplo, una evaluación podría haber sido concebida para cubrir distintos cursos o alumnos de un tramo de edad concreto. Por favor, marque todas las categorías que corresponda para cada variable.

3.2.1-3.2.2: En cada una de las variables enumeradas a continuación, marque la categoría o categorías que correspondan a la evaluación:

3.2.1. programa(s) educativo(s) / niveles CINE

3.2.2. curso(s)

3.2.3. tramos de edad

3.2.4. currículo evaluado. Para otros, especifíquelos en la **celda amarilla** e incluya las aclaraciones pertinentes en el campo para metadatos al final de la sección.

3.2.5. tipos de centros educativos. Para otros, especifíquelos en la **celda amarilla**.

3.2.6. Por favor, seleccione la orientación del programa (técnica/profesional, general o ambas). Por lo general, una evaluación se centra o bien en un programa educativo general, o bien en uno técnico/profesional; sin embargo, en el caso de que una evaluación cubriera ambos, le rogamos que añada las aclaraciones necesarias en el campo para metadatos del final de la sección.

3.2.7. Por favor, seleccione el formato en el que se administra la evaluación o prueba (escrito, oral, prácticas o una combinación de ellos). Para las evaluaciones en las que los alumnos solo tengan que someterse a pruebas escritas, seleccione «Sí» en «escrito»; si por el contrario el nivel de los alumnos se determina únicamente en base a pruebas orales, marque «Sí» en «oral». Si la evaluación se basa en una actividad práctica, marque «Sí» en «práctica». Si la prueba incluye otros formatos, menciónelos en la **celda amarilla** e incluya las aclaraciones pertinentes en el campo para metadatos al final de la sección.



Consulte en el glosario las definiciones de **curso, niveles CINE y constructo**.



3.3. Objetivos de la evaluación

Elija el objetivo u objetivos de la evaluación. Seleccione todos los que corresponda y utilice la **celda amarilla** para añadir cualquier otro objetivo de la evaluación no incluido en la lista predefinida.



3.4. Fuentes de financiación

Seleccione la(s) fuente(s) de financiación de la evaluación y utilice la **celda amarilla** para añadir cualquier otra fuente de financiación no incluida en la lista predefinida. Podría haber varias fuentes de financiación en las que estuvieran involucradas diversas partes interesadas. Por favor, desglose la información de que disponga en el campo para metadatos al final de la sección. Un ejemplo de desglose podría ser citar la proporción de fondos aportados por cada fuente.

Nota: A efectos del presente cuestionario, las organizaciones extranjeras serán aquellas creadas y financiadas en su mayoría por ciudadanos de otro país, mientras que las organizaciones locales serán aquellas creadas, financiadas y gestionadas por ciudadanos y residentes en su país. Las organizaciones internacionales se refieren a organizaciones financiadas y gestionadas de manera bilateral o multilateral. Por este motivo, las oficinas nacionales de organizaciones multilaterales como el Banco Mundial, la Organización para la Alimentación y la Agricultura, UNICEF o la UNESCO se consideran organizaciones internacionales, aunque su financiación provenga sobre todo del país en el que están establecidas.



3.5. Partes interesadas y su función

Esta sección incluye una lista indicativa con filas de posibles partes interesadas y sus posibles funciones en columnas. Seleccione en ella la combinación que corresponda entre parte interesada y función. Es posible que una sola parte interesada tenga varias funciones. Por otra parte, una función puede ser desempeñada por varias partes interesadas. Seleccione todas las que corresponda y añada cualquier otra parte interesada implicada en la evaluación en la **celda amarilla** de la última fila; anote cualquier función no mencionada en la **celda amarilla** de la última columna.



En países con una **estructura administrativa federal**, la situación podría variar considerablemente entre un país y otro, así como entre las distintas regiones dentro del país. En este caso, le rogamos que incluya en el campo para metadatos toda la información necesaria para aclarar y documentar la situación para los usuarios de los datos.

SECCIÓN 4: Diseño y administración de la prueba



4.1. Diseño de la prueba

4.1.1. En el caso de que se estén evaluando los mismos contenidos, todos los estudiantes recibirán un único cuadernillo. Sin embargo, hay casos en los que los estudiantes que participan en una evaluación concreta reciben contenidos distintos; esto puede ocurrir en el caso de que haya que evaluar demasiados ítems. Dadas las restricciones de tiempo y la carga que la prueba supone a los examinados, cada alumno recibirá un subconjunto del total de ítems. Por lo tanto, los alumnos reciben diferentes cuadernillos con una selección de ítems; los cuadernillos se elaboran según un método llamado muestreo matricial complejo, y todos los cuadernillos tienen una serie de ítems en común.

En algunos casos los cuadernillos se conciben para que distintos grupos de alumnos solo realicen las pruebas de un subconjunto del total de los objetivos de aprendizaje. En ese caso, los cuadernillos no tienen un subconjunto de ítems en común, y se basan en objetivos de aprendizaje diferentes.

Si en esta evaluación o examen público se empleara otro método de selección, descríballo en la **celda amarilla** de la última columna de esta tabla.

Por favor, seleccione el método de selección de las pruebas de la evaluación. Acuda al glosario para consultar la definición de **muestreo matricial** complejo.

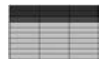
4.1.2. En esta tabla se le pide que indique las modalidades de administración y levantamiento de datos que se han utilizado en la evaluación. Elija el(los) modo(s) de levantamiento de datos pertinente(s) para cada modalidad de administración de la evaluación:

- i. un modo de levantamiento de datos cara a cara se refiere a aquel en el que los alumnos se someten a las pruebas en presencia de supervisores (personal de la evaluación).
- ii. un modo de levantamiento de datos en línea es aquel en el que las pruebas son administradas a los alumnos a través de Internet, sin la necesidad de supervisión por parte del personal de la evaluación. Tenga en cuenta que el tiempo asignado para completar la evaluación podría variar.
- iii. un modo de levantamiento de datos por correo es aquel en el que los alumnos reciben las pruebas por correo postal y las realizan sin la necesidad de supervisión por parte del personal de la evaluación.
- iv. si en la lista faltara algún modo de administración o de levantamiento de datos, por favor anótelos en la **celda amarilla** de la última fila o columna respectivamente, y añada la información adicional que sea necesaria en el campo para metadatos del final de la sección.



Por favor, consulte en el glosario la definición de **prueba fija informatizada (CFT)** y **prueba adaptativa informatizada (CAT)**.

4.1.3. Podría existir un mecanismo de revisión de la evaluación que pueda administrarse en un año concreto, o revisiones posteriores para sacar el máximo partido a las lecciones aprendidas en ediciones anteriores. Uno de sus objetivos podría ser comprobar la adecuación de los instrumentos de la evaluación al currículo o áreas evaluadas. Indique si esta revisión es realizada normalmente por expertos cualificados independientes, si se trata de un proceso interno llevado a cabo de forma periódica por el ministerio o autoridad responsable del diseño de la evaluación, o si el proceso se decide caso por caso. Por favor, indique en la **celda amarilla** si existen otros mecanismos al margen de los mencionados, y añada los detalles necesarios en el espacio para metadatos.



4.2. Características de los ítems/preguntas y de los cuestionarios contextuales

La información solicitada en esta subsección puede ser obtenida de una muestra de cuadernillos de la última evaluación.

4.2.1. Formato de los ítems/preguntas

Por favor, seleccione el formato adecuado:

- i. Dos opciones: si hay ítems o preguntas cuya respuesta debe escogerse entre solamente dos opciones predefinidas.
- ii. Tres opciones o más: si hay ítems o preguntas cuya respuesta debe escogerse de una lista de tres opciones predefinidas o más.
- iii. Respuesta abierta breve: si hay ítems o preguntas en las que se puntúa a los alumnos (mediante una escala con un máximo y un mínimo) en función de una respuesta breve redactada con sus propias palabras.
- iv. Redacción: si los alumnos tienen que redactar una respuesta estructurada compuesta por varios párrafos.
- v. Proyecto individual: las evaluaciones técnicas/profesionales suelen incluir un proyecto práctico concreto mediante el que se evalúa la aplicación práctica de conocimientos y competencias. Marque esta opción si los alumnos deben realizar una presentación del proyecto ante un jurado o el personal de la evaluación, o si deben entregar un resultado concreto que demuestre su capacidad para aplicar cierta técnica.
- vi. Proyecto grupal: si el proyecto es realizado por un grupo de alumnos.
- vii. Otro: por favor, en la **celda amarilla** añada cualquier otro formato de ítem o pregunta que forme parte de la evaluación y no aparezca en la lista de posibles formatos.

4.2.2. Formato del estímulo

El estímulo aporta el contexto sobre el que se basará el alumno para responder a las preguntas o realizar algunas tareas. Los estímulos pueden tener diversos formatos. Seleccione todos los que corresponda:

- i. Texto: si se utiliza texto para comunicar el contexto al alumno, independientemente del soporte empleado. Por ejemplo, el texto podría entregársele en papel o proyectado sobre una pantalla. Además, podría adoptar diversos formatos y presentarse como un texto corrido en prosa, o un texto discontinuo (listas, tablas, horarios, mapas, etc.). Consulte en el glosario la definición de textos corridos y discontinuos, mixtos y múltiples.
- ii. Audio: si el estímulo es únicamente sonoro. Por ejemplo, podría pedirse a los alumnos evaluados que escucharan un breve relato y contestaran algunas preguntas sin contar con el texto correspondiente por escrito. Los contenidos podrían reproducirse a partir de una grabación o enunciarse en directo.
- iii. Vídeo: si el estímulo empleado es una combinación de audio e imágenes.
- iv. Otros: Por favor, describa cualquier otro formato incluido en la evaluación que no haya sido mencionado entre los formatos de estímulo posibles. Descríbalos brevemente en la **celda amarilla** e incluya cualquier otro detalle en el campo para metadatos al final de la sección.



Consulte el glosario para conocer la definición de **texto corrido, discontinuo, mixto y múltiple**.

4.2.3 - 4.2.4.

Alumnos con necesidades especiales: Consulte en el glosario la definición de autismo, deficiencia auditiva, sordoceguera, sordera, trastorno emocional, discapacidad física, deficiencia visual y deficiencia del habla.

4.2.4 (i) Marque la casilla del tipo de discapacidad para el que se ha adaptado esta evaluación.

4.2.4 (ii) Describa las adaptación(es) ofrecidas para cada tipo de discapacidad. Por ejemplo, si la evaluación responde a las necesidades de los discapacitados visuales, podría estar disponible en braille, o el formato del estímulo podría variar (audio en vez de texto).

Tenga en cuenta que si la opción en 4.2.4.i es "No", las celdas correspondientes en 4.2.4.ii se completarán automáticamente con el código "a" para categoría no aplicable.

4.2.5. Si las personas evaluadas tienen conocimientos limitados de la lengua de la evaluación, describa las adaptaciones realizadas en la celda correspondiente. Por ejemplo, podría permitirse a los alumnos evaluados que presentaran sus respuestas en un idioma distinto. Esta situación podría ser frecuente en contextos bilingües. También podría concedérseles tiempo adicional para terminar la evaluación.

4.2.6. Cuestionarios contextuales

Le rogamos que cumplimente la siguiente tabla con información general relativa a los distintos cuestionarios de contexto que forman parte de la evaluación general. En cada cuestionario:

- Escriba su nombre en la primera columna.

- Elija el grupo encargado de cumplimentar el cuestionario. Para otro distinto, añada los detalles necesarios en el campo para metadatos al final de la sección.
- Aporte más detalles sobre el tiempo asignado oficialmente para terminarlo (si se trata de estimaciones, recuerde colocar un asterisco delante de la cifra).
- Explique el objetivo del cuestionario. Por ejemplo, el cuestionario contextual podría ser cumplimentado por los padres y versar sobre la participación de los niños en actividades de alfabetización en el hogar.
- Indique la naturaleza de los datos recopilados. Algunos de los conjuntos de variables utilizados con más frecuencia para levantar datos contextuales son los datos demográficos, la estructura del hogar y familiar, la educación y el lugar de residencia.

Ejemplo: en PISA 2012, la principal prueba de la evaluación incluía cuestionarios sobre matemáticas, lectura, ciencia, resolución de problemas, conocimientos financieros y TIC. Por otra parte, los **cuestionarios** contextuales que debían cumplimentar alumnos, padres y directores de los centros educativos recopilaban información sobre el entorno. Este tipo de información suele emplearse en los análisis.



Por favor, envíe al UIS copias electrónicas de los cuestionarios contextuales o formularios de las pruebas de cada evaluación incluida en este estudio si fueran públicos. Tenga en cuenta que estos cuestionarios contextuales y los formularios de las pruebas se pondrán a disposición pública salvo que se indique lo contrario.

4.2.7. Marcos de la evaluación

En diversas evaluaciones nacionales, las herramientas de levantamiento de datos se elaboran con base en marcos específicos desarrollados por el equipo nacional de la evaluación. Estos marcos explican la racionalidad de la estructura de la evaluación. Por ejemplo, los marcos de la evaluación de PISA 2012 fueron: marco de evaluación de matemáticas, marco de evaluación de lectura, marco de evaluación de ciencias, marco de evaluación de la resolución de problemas y marco de evaluación de los conocimientos financieros.

4.2.8. Si está disponible, le rogamos que facilite la lista de constructos que deberían medirse en la evaluación. Suelen figurar en los marcos de la evaluación.



Por favor, envíe al UIS copias electrónicas de los marcos de la evaluación o la lista de constructos cubiertos por la evaluación (si existen). Tenga en cuenta que estos marcos se harán públicos, salvo que se indique lo contrario.

SECCIÓN 5: Cobertura y diseño muestral



5.1. Cobertura

Esta sección requiere información sobre la cobertura geográfica y poblacional de la evaluación realizada.

5.1.1. Cobertura geográfica: Seleccione el área geográfica que cubre esta evaluación de la lista predefinida.

Tenga en cuenta que el objetivo de este estudio debería ser una cobertura nacional.

5.1.2. Población de interés: Describa y caracterice lo mejor posible la población objetivo de la evaluación con el fin de ofrecer una imagen completa de los participantes en la evaluación. Por ejemplo, la población objetivo podría describirse según su curso o nivel educativo, edad, lugar de residencia (rural, urbana), sexo, tipo de escolarización, titularidad del centro educativo (pública o no), necesidades especiales, etc.

5.1.3. Niños no escolarizados

i) Por favor, marque «Sí» si esta evaluación incluye a niños no escolarizados, y pase al punto 5.1.3 (ii). En caso contrario, pase al ítem 5.1.4, ya que las celdas correspondientes a 5.1.3.ii y 5.1.3.iii se completarán automáticamente con el código "a" para la categoría no aplicable.

ii) Por favor, marque «Sí» si los niños no escolarizados siguen programas de alfabetización equivalentes al curso que cubren el currículo en el que se centra esta evaluación (si procede). Por ejemplo, este podría ser el caso de los niños que sean tomados como muestra en programas de alfabetización.

iii) Por favor, marque «Sí» en el caso de que aprobar la evaluación permita a los niños no escolarizados reintegrarse en el sistema educativo en el siguiente curso escolar que corresponda.

Tenga en cuenta que algunos exámenes públicos están abiertos a todos los individuos, estén escolarizados o no. Es común que aprobar la evaluación sea un requisito para acceder a un programa educativo en concreto, independientemente de su trayectoria escolar previa.

5.1.4. Algunos grupos podrían quedar excluidos de la evaluación por motivos diversos, vinculados directamente al objetivo de la evaluación o no. Por ejemplo, los alumnos con deficiencias visuales podrían quedar excluidos si las pruebas no incluyen modalidades específicas que faciliten su participación. Otros niños podrían quedar excluidos por residir en zonas de difícil acceso. Por favor, describa los grupos oficialmente excluidos de esta evaluación, si procede.

5.1.5. Seleccione los lugares en los que los alumnos están invitados a participar en la evaluación, ya sea solo en las principales ciudades o centros locales diseminados por el país (centros educativos en pueblos, ciudades pequeñas y grandes, etc.). Para añadir otros lugares en los que los alumnos se someten a la evaluación recurra a la **celda amarilla**.



5.2. Participación

Cuando las evaluaciones nacionales no se centren en toda la población de alumnos de un cierto nivel, curso o edad, puede evaluarse una **muestra** representativa de la población total. Una muestra representativa se consigue sobre la base de diversas características o variables, como el número de centros educativos y su distribución en zonas urbanas/rurales y grandes ciudades, el número de aulas, el número de candidatos por centro educativo/aula, la edad y el sexo de los candidatos. En general, esta información figura en el informe técnico de la evaluación y se utiliza posteriormente para validar, imputar y ajustar los datos antes de efectuar el análisis.

Por favor, indique tanto la población objetivo total de alumnos (N) como el número de alumnos evaluados (n) por cada curso incluido en la evaluación y por su sexo.

- ✓ Las evaluaciones pueden abarcar uno o varios cursos. Seleccione el(los) curso(s) cubiertos (como en 3.2.2). Se puede elegir un curso por fila.
- ✓ En el caso de una evaluación basada en muestras, el total de alumnos evaluados (n) será el tamaño muestral, y el total (N) será el número de alumnos identificados en el marco muestral de la población objetivo de alumnos.
- ✓ En el caso de evaluaciones basadas en un censo, el total (N) será el número de alumnos inscritos para hacer las pruebas, mientras que los alumnos evaluados (n) será el número de alumnos que efectivamente se sometió a la prueba. En este caso, n y N deberían coincidir o ser prácticamente iguales.

En la última columna, debe incluir el rango de edad de los alumnos evaluados. Si es posible, distinga entre las edades objetivo y las edades reales de los alumnos evaluados. Por ejemplo, podría escribir «10-15 y 8-9, 16+» y, en el campo de metadatos del final de la sección, explicar que la prueba se dirige a alumnos de entre 10 y 15 años aunque, al estar abierta a cualquier participante, también incluye individuos de 8 a 9 años y mayores de 16.

Please, note that if the option "**Census (all students at the given grade(s) or age level(s)**" was chosen in 3.1.4, then all the cells in sub-sections 5.3 and 5.4 will turn automatically into code "a" for not applicable. In this case, please go directly to the section 6.

Tenga en cuenta que si la opción "**Censos (todos los estudiantes en el grado (s) o rango de edad)**" fue elegida en el punto 3.1.4, a continuación todas las celdas en las subsecciones 5.3 y 5.4 se completarán automáticamente con el código "a" para la categoría no aplicable. En este caso, por favor, vaya directamente a la sección 6.

5.3. Muestreo

5.3.1. Pueden utilizarse diversos métodos de muestreo para obtener una muestra representativa de la población objetivo. Sin embargo, esto depende, entre otras cosas, de la estructura de los marcos muestrales, el presupuesto, el nivel deseado de desglose de los datos y los informes requeridos para la elaboración de políticas. En el caso concreto de la educación se extraen muestras a distintos niveles, desde los niveles administrativos a los propios alumnos. Los planes de muestreo más comunes están incluidos en las opciones predefinidas de la pregunta 5.3.1. Para los objetivos del presente estudio, escoja el

plan de muestreo empleado para seleccionar las unidades administrativas (regiones/provincias/estados, divisiones, subdivisiones, distritos, etc.), las muestras de centros educativos, aulas y alumnos (si procede).

5.3.2. Incluya un breve resumen descriptivo de cada marco muestral empleado para obtener la(s) muestra(s).

5.3.3. Incluya un breve resumen de las omisiones del plan si procede. La omisión es una fuente de errores conocida como defecto de cobertura. Ocurre cuando la muestra se extrae de marcos muestrales inexactos, incompletos, insuficientes o desfasados.

5.4. Tasas de participación y ajustes

Consulte en el glosario la definición de las tasas de participación y los métodos de ajuste antes de cumplimentar esta tabla.



Consulte en el glosario la definición de **muestra**, **método de muestreo**, **tasas de participación** y **tasas de participación ponderadas**.

SECCIÓN 6: Procesamiento de datos



6.1. Procesamiento de datos

6.1.1. Edición de datos o verificación de datos: fase entre el levantamiento de datos y la entrada de datos. Es un paso necesario que garantiza que los datos están correctamente marcados en los cuestionarios o cuadernillos de las pruebas antes de ser capturados. Por favor, seleccione el método empleado para editar los datos levantados, antes de que se capturen. Cuando se utiliza la versión «de papel y lápiz» para recopilar datos, puede ocurrir que la **edición de datos** se efectúe directamente sobre las copias de papel antes de incluir los datos en una base de datos. También hay casos en los que los datos se introducen en una base de datos, pero la edición no la realiza un programa informático, sino que se efectúa una supervisión visual. En cualquiera de los dos casos, marque «control manual o visual». Por el contrario, si se utiliza un programa informático para llevar a cabo esta tarea, marque la opción **automático**, independientemente del método utilizado para introducir los datos en el ordenador. Si en el proceso de edición de datos se usan ambos métodos, seleccione ambos. En caso de que se utilicen otros métodos de edición de datos, indique de cuáles se trata en la **celda amarilla** de la última columna.

6.1.2. Entrada/captura de datos:

En general, los datos de la evaluación deben centralizarse en una base de datos para facilitar su gestión (limpieza, búsqueda, análisis, intercambio, etc.), independientemente de si la evaluación es en formato electrónico o con papel y lápiz. En el caso de las pruebas con papel y lápiz, los datos disponibles sobre el papel deberán introducirse meticulosamente en una base de datos, bien manualmente por un operador especializado, bien mediante un escáner. En este último caso, el escaneado convierte el documento de papel en una imagen o datos electrónicos que pueden utilizarse en aplicaciones informáticas y archivos. El uso del escáner se ha vuelto más frecuente en los últimos tiempos porque acelera el procesamiento de los datos de los censos de población y vivienda o de estudios a gran escala sobre los hogares, independientemente de que se recojan en persona o por correo. Indique si la captura o la entrada de datos es manual, por escáner u otro método, y explique su respuesta en la **celda amarilla**.

6.1.3. Lugar de la edición y entrada de datos:

Dependiendo de la cultura organizativa de cada evaluación, la edición y entrada de datos puede realizarse en centros locales de evaluación ubicados en todo el país. Por ejemplo, puede realizarse en el lugar en el que los alumnos se sometieron previamente a la prueba, en centros regionales o en la sede de la institución que administra la evaluación o ha sido designada para realizar estas tareas. Seleccione las opciones que correspondan a esta evaluación.



Consulte en el glosario la definición de **procesamiento de datos**, **edición de datos**, **captura de datos** y **codificación de datos**.



6.2. Valoración de datos (solo para evaluaciones basadas en una muestra)

Los datos solicitados en esta subsección suelen estar a disposición de los usuarios en los informes técnicos.

6.2.1. y 6.2.2. Por favor, compruebe si hay disponible un informe técnico de esta evaluación. Le resultará más fácil cumplimentar esta sección con los datos adecuados relativos a la estimación de errores en el muestreo y a las acciones emprendidas para reducir los errores no vinculados al muestreo. Si hay un informe técnico disponible en línea, por favor indique su URL en el campo para metadatos al final de la sección.

6.2.3. Los candidatos son clasificados en la categoría de «ausencia de respuesta» si:

- i. No se han levantado datos porque no se los ha podido contactar, o han preferido no participar en la evaluación. No hay datos en absoluto.
- ii. El número de preguntas/tareas/ítems completados no es suficiente (está por debajo del mínimo fijado, en porcentaje del total de preguntas/ítems/tareas) como para tener en cuenta sus datos en el subsiguiente análisis. Por ejemplo, podría darse el caso de que el evaluado solo respondiera a un 10% de la evaluación o menos y fuera por lo tanto considerado como no participante, o como si no se hubiera presentado a la prueba.

Si se hubiera fijado un porcentaje de cumplimentación como regla para identificar a los no participantes, indíquelo en la celda. Si se utiliza otro sistema para ello, indíquelo en la **celda amarilla** y explíquelo en los metadatos.

6.2.4. Ausencia de respuesta a los ítems

Un ítem es catalogado como de «no respuesta» en distintos casos. Por favor, elija los criterios utilizados en esta evaluación y sírvase de la **celda amarilla** para indicar si en la evaluación se utilizó un criterio que no aparece en la lista.



Incluya cualquier observación/aclaración que sea importante para entender e interpretar los datos en el campo para metadatos al final de la sección.

SECCIÓN 7: Mediciones y resultados

Le rogamos que consulte las definiciones y ejemplos incluidos en el manual de instrucciones antes de cumplimentar esta sección.



7.1. Mediciones

7.1.1. Categorías de modelos

i) Las definiciones de los modelos TCT, TRI y Rasch figuran en el glosario. El informe técnico de la evaluación suele incluir información sobre el uso de estos modelos y también puede incluirse en otros informes que presenten el desempeño de los alumnos. Por favor, marque todas las teorías de medición que correspondan a esta evaluación, e indique en la **celda amarilla** si se ha utilizado alguna que no aparezca en la lista para medir el desempeño de los alumnos.

ii) Si se escoge el modelo TRI, especifique de qué modelo se trata, ya sea logístico, de Rasch u otra categoría no citada. Si la opción IRT en 7.1.1.i es "No", las celdas en 7.1.1.ii se completarán automáticamente con el código "a" para la categoría no aplicable.

iii) Elija en la lista si los resultados se presentan por área evaluada, de manera global o ambas. Presentar los resultados por área significa que hay resultados disponibles para cada campo de conocimientos teórico o subárea, que al combinarse componen la asignatura evaluada. Por ejemplo, la evaluación de matemáticas puede constar de tres subáreas diferentes: álgebra, geometría y estadística.

7.1.2. Partiendo de los informes analíticos o técnicos es posible identificar el método de evaluación del curso empleado. Por favor, seleccione todos los que correspondan a esta evaluación y use la **celda amarilla** para indicar cualquier otro método de evaluación que no aparezca en la lista pero se utilice en la evaluación.



Consulte en el glosario la definición de **teoría clásica de pruebas (TCP)**, **teoría de respuesta a los ítems (TRI)**, **modelo de medición de Rasch**, **método normativo**, **método basado en criterios** y **método basado en el individuo**.



7.2. Resultados

7.2.1. Esta subsección debe cumplimentarse **únicamente si la evaluación o examen público descrito** en la sección 3.1 **no incluye divisiones del alumnado por desempeño o área de especialización**. En caso contrario, omita las subsecciones de 7.2.1 a 7.2.3 y pase directamente a 7.2.4.

Por favor, indique las métricas empleadas en la evaluación y señale el valor mínimo necesario para que el alumno alcance la norma nacional. Tenga en cuenta que cuando la respuesta es "No", la celda correspondiente al requisito mínimo para cumplir con la norma nacional se completará automáticamente con el código "a" para la categoría no aplicable.

Por ejemplo, si se usa el porcentaje de ítems correcto, un alumno alcanzará la norma nacional mínima si contesta correctamente al 50% de los ítems, siempre y cuando todos los ítems tengan el mismo peso. En este caso, la norma nacional mínima sería el 50%.

Por lo general, cuando se utiliza una escala (promedio) para indicar el desempeño de los alumnos, la norma nacional mínima se expresará mediante una cifra bruta. Suponiendo que la nota mínima exigida sea de 500, se considerará que los alumnos que consigan un resultado igual o superior a 500 habrán alcanzado la norma nacional.

Actualmente cada vez es más frecuente utilizar niveles de rendimiento para plasmar y categorizar las diferencias en el desempeño de los alumnos. Agrupar a los alumnos permite enfocarse en sus necesidades específicas. Facilita la creación de políticas adecuadas y el desarrollo de programas educativos para apoyar a grupos con necesidades especiales. También suelen utilizarse las letras del alfabeto (A, B, C, D, etc.) y rangos de números. Sea como fuere, la relación entre ambos formatos es unívoca. Cada letra o serie describe los tipos de competencias que los alumnos dominan o no. Si corresponde en esta evaluación, marque la casilla de niveles de rendimiento y facilite la explicación o descripción completa de cada nivel, tal y como aparece en la versión más reciente de los documentos oficiales de esta evaluación. Por favor, asegúrese de incluir las descripciones de los niveles de rendimiento en el campo para metadatos del final de la sección.

Tenga en cuenta que en las evaluaciones internacionales, las métricas y las normas mínimas deben ser las fijadas por la institución que efectúa las pruebas, y no las que elijan los países participantes.

7.2.2. En algunos países, las evaluaciones realizadas en las escuelas (cuyas pruebas se desarrollan y administran a nivel del centro educativo) suponen una parte o peso en el resultado final, sobre todo en exámenes públicos. Si este es el caso de esta evaluación, por favor indíquelo en la celda que corresponda y añada las aclaraciones adicionales en el campo para metadatos. En caso contrario (si este estudio no incluye evaluaciones a nivel de los centros educativos), escriba la letra «a» de «no aplicable».

7.2.3: Porcentaje de alumnos por encima de la norma nacional mínima exigida por sexo y edad.

- i. Algunas evaluaciones o exámenes a gran escala se administran en más de un curso. Por ello, algunas asignaturas podrían administrarse en un curso determinado, o en varios. La tabla facilita espacio para cuatro cursos. Si la evaluación cubre cuatro cursos o más, puede incluir ítems similares a los de 7.2.3.i-iv (copie las filas y péguelas debajo de 7.2.3.iv). Si la evaluación o examen público cubre sólo un grado, por favor seleccione la opción "a" para el grado evaluado en 7.2.3. ii a 7.2.3.iv. En ese caso, todas las celdas en estas tablas se completarán automáticamente con el código "a" para la categoría no aplicable.
- ii. En primer lugar escriba cada constructo/asignatura obligatoria o área del currículo objeto de la evaluación en la **celda amarilla** que le corresponda de la segunda fila. Por ejemplo, podría ser que las asignaturas obligatorias de un examen público nacional sean matemáticas, inglés, geografía y contabilidad.

Si están disponibles, escriba las distintas áreas o constructos evaluados. Mientras que las asignaturas se imparten dentro de un módulo concreto y en un entorno de aprendizaje estándar, los constructos están vinculados a ciertas competencias o aptitudes que emergen como resultado de combinar los conocimientos adquiridos en diversas disciplinas, la experiencia personal y el contexto. Por ejemplo, la aritmética o cálculo no es una disciplina impartida como tal, sino un rasgo que se mide en distintas evaluaciones. También podrían citarse como ejemplos de constructos los distintos tipos de alfabetización (en lectura, matemáticas, etc.) descritos en la literatura del siglo XXI relativa a las competencias.

Por ejemplo, en [PISA 2015](#) se evalúan cuatro constructos entre alumnos de 15 años: competencia matemática , *competencia lectora, competencia científica y aptitudes y competencias para la resolución colaborativa de problemas.*

Indique el porcentaje de alumnos cuyo desempeño es igual o superior a la norma mínima nacional en el conjunto de las pruebas y en función de las asignaturas/constructos evaluados, detallando el total para todos los estudiantes y luego desglosando entre hombres y mujeres. Incluya asimismo el porcentaje de alumnos cuyo desempeño es igual o superior a la norma mínima nacional para cada edad de los estudiantes evaluados. Por favor, indique las edades correspondientes entre paréntesis.

Ejemplo: En el país X se efectúa una evaluación nacional a gran escala entre alumnos del curso 6. Las edades correspondientes de los evaluados tal y como figuran en la base de datos de la evaluación son 10, 11, 12, 13, 14, 16, 17, 20 y 23 años.

Oficialmente, solo los alumnos menores de 14 años pueden inscribirse en el curso 6 en este país. Pero al tratarse de un examen público, incluso los candidatos no escolarizados pueden presentarse a las pruebas, siempre y cuando cumplan las condiciones y se hayan inscrito en ellas. En este caso concreto, es razonable presentar los datos de los alumnos de 10, 11, 12, 13 y 14 años por un lado, y los de 15 años en adelante por otro. En caso de que se trabaje con un rango de edades de este tipo, le rogamos que incluya explicaciones al respecto en el campo de metadatos para quienes se sirvan de estos datos.

En cada tabla también deberá indicar el idioma en el que se administran las pruebas de cada una de las asignaturas o constructos obligatorios, su duración oficial y el porcentaje que cada área representa en la nota total de la asignatura/constructo. Supongamos por ejemplo que las matemáticas incluyen cuatro áreas: álgebra, geometría, probabilidad y estadística. El álgebra podría suponer el 25% de la nota total; la geometría, el 30%; la probabilidad, el 15%; y la estadística el 30%.

Cuando sea apropiado, por favor también proporcione el peso de cada asignatura/constructo cubierto en forma de porcentaje sobre la nota total en la evaluación o examen. Este suele ser el caso en los exámenes públicos. Por ejemplo, en Mauritania, el CEPAS / CEF es un examen público organizado cada año por el Ministerio de Educación Nacional. Se administra en forma voluntaria a todos los estudiantes al final de la educación primaria (CINE 1), o de grado 6. Se compone de siete asignaturas obligatorias, que se indican a continuación, y el lenguaje de la prueba es árabe o francés. Los títulos de cada asignatura, el tiempo asignado en minutos, el idioma de las pruebas, y los pesos en porcentaje de la puntuación total obtenida en este examen para cada asignatura obligatoria se proporcionan a continuación:

Asignatura	Tiempo asignado (mn)	Idioma de la prueba	Peso (% de la nota total)
Árabe	90	Árabe	25
Matemáticas	90	Árabe	25
Francés	60	Francés	15
Historia y geografía	45	Árabe	10
Ciencias Naturales	45	Francés	10
Religión	45	Árabe	10
Formación cívica	30	Árabe	5

Los resultados de las evaluaciones de los estudiantes también pueden ser reportados solamente por asignatura o área evaluada, sin tener el resultado total. Ese es el caso en algunos exámenes públicos donde los estudiantes deben tomar una lista de materias obligatorias, así como algunas materias optativas para obtener una certificación. En ese caso, por favor, sólo informar el peso por área cubierta como porcentaje de la nota total de la asignatura / constructo, si está disponible, mientras que la fila anterior debe completarse con el código "a" para la categoría no aplicable.

7.2.4: Esta subsección debe cumplimentarse únicamente si la evaluación descrita en la sección 3.1 incluye divisiones o áreas de especialización (cumplimentar en las celdas).

- i. En primer lugar, indique cada división o área de especialización en las celdas correspondientes.
- ii. A continuación, elija la métrica utilizada para cada división, indique el mínimo exigido para alcanzar la norma nacional, el peso de las evaluaciones a nivel de los centros educativos y la lista de áreas de especialización disponible para esta división.

7.2.5: Las divisiones indicadas más arriba aparecen automáticamente en la primera fila de esta tabla.
Por favor, cumplimente esta tabla únicamente si las divisiones no tienen áreas de especialización.

En cada división (pero no para las asignaturas evaluadas en ella), indique los resultados conjuntos de hombres y mujeres, solo hombres, solo mujeres y por edad.

En la última fila de esta tabla, escriba la lista de asignaturas obligatorias examinadas en cada división.

7.2.6: Usos de los resultados

Los usos más comunes de las evaluaciones están enumerados en 7.2.6. Por favor, indique la intensidad con la que **se ha utilizado** la evaluación (solo una opción por fila) e incluya cualquier información adicional sobre el uso de los resultados de la evaluación en el campo para metadatos. Se trata de un punto importante, ya que una sección de los informes resúmenes nacionales está enfocada en los efectos de la evaluación en la formulación de políticas. Este uso es diferente del objetivo: el objetivo de una evaluación es la finalidad con que se concibió, y sus impulsores suelen declararlo abiertamente al principio del proceso; sin embargo, los usos de la evaluación podrían trascender su objetivo original, ya que otros actores tienen acceso a los datos de la evaluación y pueden servirse de ellos para sus propios fines.

Utilice el campo para metadatos para aportar cualquier otra información adicional que sirva para completar y entender mejor los datos solicitados en esta sección del cuestionario.

SECCIÓN 8: Divulgación, presentación y accesibilidad de los datos



8.1. Nivel de representación

Para que los datos sean relevantes en la formulación, seguimiento y evaluación de políticas, deben hacerse análisis a distintos niveles de unidades administrativas y a diferentes grupos poblacionales. Por ejemplo, un análisis basado en el género es necesario para supervisar las disparidades existentes entre estudiantes hombres y mujeres. Por favor, marque todas las opciones que correspondan a esta evaluación y utilice el campo para metadatos del final de la sección para incluir información adicional.



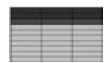
8.2. Divulgación de datos

Se pueden utilizar múltiples plataformas para informar al público y a la comunidad educativa de los resultados y la disponibilidad de los datos de la evaluación. Por favor, marque todas las plataformas empleadas para divulgar los datos de la evaluación en cuestión.



8.3. Accesibilidad de los datos

No es fácil acceder a los microdatos, sobre todo en evaluaciones de estudiantes. Esta subsección pretende recopilar toda la información práctica necesaria para los investigadores o para cualquier persona interesada en usar los microdatos y los documentos relacionados con ellos en futuras investigaciones. Esta información no siempre está documentada. Por favor, recurra a las personas responsables para recopilar esta información o para confirmar que la información inicial que pueda estar disponible sigue siendo válida, y cumplimente las celdas pertinentes del cuestionario.



8.4. Desafíos

La administración y gestión de evaluaciones a gran escala siempre representa un desafío, sobre todo en países en desarrollo. Estos desafíos se reportan generalmente para consideraciones futuras. Por favor, haga un breve resumen de estos desafíos en el espacio correspondiente con respecto a las siguientes etapas del proceso de evaluación de los estudiantes: planificación, diseño de la prueba, recopilación de datos, procesamiento de datos, análisis de datos, reforma de política educativa, divulgación de datos y otros.



Consulte en el glosario la definición de **divulgación de datos**.

5. GLOSARIO Y DEFINICIONES

Ausencia de respuesta del candidato: estatus de aquellas personas evaluadas que se da cuando no responden al cuestionario o cuando la cantidad de respuestas es insuficiente como para arrojar datos relevantes desde el punto de vista analítico.

Captura de datos: proceso en el que se introducen datos en un sistema de bases de datos, sobre todo cuando el levantamiento de datos no está informatizado.

Censo: estudio oficial que abarca toda la población de un sistema específico. Por ejemplo, un censo escolar comprende a todas las escuelas de un sistema educativo. Para el propósito del catálogo del UIS, la población objetivo incluye a los estudiantes de una cierta edad o curso escolar, o bien a los candidatos inscritos que cumplen las condiciones para aprobar un examen público o para participar en una evaluación.

CINE: Sistema de clasificación que define un marco para la descripción estadística completa de los sistemas educativos nacionales y una metodología que traduce los programas educativos nacionales a unos niveles educativos comparables a nivel internacional. La unidad básica de clasificación del CINE es el programa educativo. El CINE también clasifica los programas por ámbito de estudio, orientación y salidas de los programas. Por favor, haga clic en el siguiente vínculo para acceder a más información sobre la clasificación CINE:

<http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx>

Codificación de datos: proceso mediante el que se asignan valores numéricos a respuestas que originalmente tenían un formato específico, tal como numérico o de texto, audio o vídeo. Su principal objetivo es facilitar la gestión automática de los datos con fines analíticos.

Competencia: Capacidad de movilizar y utilizar recursos internos como conocimientos, destrezas y actitudes, así como recursos externos, como bases de datos, colegas, compañeros, bibliotecas, herramientas, etc., con el propósito de solucionar problemas concretos de una manera eficiente en situaciones de la vida real.

Curso: fase concreta de la instrucción en la educación inicial que suele abarcar un año académico. Los estudiantes que están en el mismo curso suelen tener la misma edad. También se utilizan los términos «clase», «cohorte» o «año» (glosario UIS).

Discapacidades: Término general que abarca las deficiencias, las limitaciones de la actividad y las restricciones de la participación. Las deficiencias son problemas que afectan a una estructura o función corporal; las limitaciones de la actividad son dificultades para ejecutar acciones o tareas, y las restricciones de la participación son problemas para implicarse en situaciones de la vida diaria. Es un fenómeno complejo que refleja una interacción entre las características del organismo humano y las de la sociedad en la que vive. Para superar las dificultades a las que se enfrentan las personas con discapacidad es preciso intervenir para eliminar las barreras sociales y del entorno (Organización Mundial de la Salud²).

Autismo: discapacidad del desarrollo que afecta significativamente la comunicación verbal, la no verbal y la interacción social; suele evidenciarse antes de los tres años y tiene efectos negativos en el desempeño educativo del niño. Otras características del autismo son la realización de actividades repetitivas y movimientos estereotipados, la resistencia a los cambios en su entorno o en sus rutinas diarias y respuestas insólitas a experiencias sensoriales. No se emplea el término autismo si los efectos negativos sobre el desempeño educativo del niño están causados por un trastorno emocional. Un niño que muestra rasgos de autismo después de los tres años podría ser diagnosticado de autismo si cumple dichos criterios (Congreso de los Estados Unidos, IDEA 1997).

²<http://www.who.int/topics/disabilities/es/>

Deficiencia auditiva: trastorno de la audición, ya sea permanente o no, que tiene un efecto negativo en el desempeño educativo del niño pero no está incluido en la definición de «sordera» (Congreso de los Estados Unidos, IDEA 1997).

Deficiencia auditiva (sordera incluida): trastorno de la audición, ya sea permanente o no, que tiene un efecto negativo en el desempeño educativo del niño.

Deficiencia del lenguaje o habla: trastorno en la comunicación como el tartamudeo o los problemas de articulación o en la voz que afecta negativamente al desempeño educativo del niño (Congreso de los Estados Unidos, IDEA 1997).

Deficiencia visual: trastorno en la vista que afecta negativamente al desempeño educativo del niño, incluso con corrección. El término incluye la hipovisión y la ceguera (Congreso de los Estados Unidos, IDEA 1997).

Factores ambientales: Elementos del entorno que impiden que ciertos alumnos puedan realizar la evaluación con normalidad y a los que la prueba por tanto se adapta para posibilitar su participación normal. Por ejemplo, podría entregárseles una prueba alternativa o dárseles la oportunidad de realizarla en otro momento.

Sordera: deficiencia auditiva tan grave que incapacita al niño para procesar información lingüística a través del oído, con o sin amplificador, y afecta negativamente a su desempeño educativo (Congreso de los Estados Unidos, IDEA 1997).

Sordoceguera: deficiencias visual y auditiva concomitantes [simultáneas] cuya combinación implica graves problemas comunicativos, educativos y del desarrollo para los niños que las padecen, de tal modo que no pueden integrarse en programas especiales concebidos únicamente para niños sordos o ciegos (Congreso de los Estados Unidos, IDEA 1997).

Trastorno emocional: estado que presenta una o varias de las siguientes características durante un período prolongado de tiempo y en tal medida que afecta negativamente el desempeño educativo del niño: incapacidad de aprender que no puede explicarse mediante factores intelectuales, sensoriales o de salud; incapacidad de crear o mantener relaciones interpersonales satisfactorias con sus compañeros y maestros; comportamientos o sentimientos inadecuados en circunstancias normales; estado anímico dominante de tristeza o depresión; tendencia a desarrollar síntomas físicos o miedos relacionados con sus problemas personales o escolares; el término incluye la esquizofrenia. El término no se aplica a niños socialmente inadaptados, salvo que se determine que tienen un trastorno emocional (Congreso de los Estados Unidos, IDEA 1997).

Diseño muestral

Muestra: subconjunto de individuos de una población concreta escogido en función de un proceso de selección (con o sin sustitución y de manera aleatoria o no). Una muestra escogida mediante una selección aleatoria de individuos basada en probabilidades conocidas se conoce como muestra aleatoria o probabilística. Una muestra no aleatoria se forma sobre la base de un método de selección subjetivo.

Muestreo: método para la creación de una muestra basado en un marco (de muestreo) que contiene a todos los individuos de la población objetivo. Los métodos de muestreo de la población dependen de múltiples parámetros, como el proceso de selección de la muestra (con o sin sustitución y de manera aleatoria o no), la estructura del marco muestral, el nivel requerido de desglose en los datos y el análisis necesario, y el presupuesto disponible para el estudio. Algunos métodos comunes de **muestreo probabilístico** y **no probabilístico** se presentan más adelante en el glosario (Statistics Canadá)⁴.

³IDEA: Ley para la Educación de Individuos con Discapacidades (Congreso de los Estados Unidos, IDEA 1997).

⁴<http://www.statcan.gc.ca/edu/power-pouvoir/ch13/prob/5214899-eng.htm#a1>
<http://www.statcan.gc.ca/edu/power-pouvoir/ch13/nonprob/5214898-eng.htm>

Diseño muestral matricial: proceso mediante el que se administran distintos subconjuntos de ítems a distintos subconjuntos de personas evaluadas, de tal modo que cada ítem se somete al menos a un subconjunto de la muestra de personas evaluadas. En general, hay ítems importantes llamados «ítems principales» que se administran a todas las personas evaluadas, al contrario que los ítems «fraccionados», que solo se administran a un subconjunto de encuestados. Por lo general, los ítems principales tienen alta capacidad predictiva respecto de los ítems fraccionados de la prueba.

Divulgación de datos: difusión de datos a través de diversos medios (modernos o tradicionales) como Internet o medios en línea, conferencias o comunicados de prensa, artículos en un periódico impreso, entrevistas en televisión o radio, etc.

Edición de datos: controles destinados a identificar los datos que falten o no sean coherentes o válidos con el fin de corregirlos o reemplazarlos mediante consultas con el encuestado, revisiones manuales o imputaciones.

Estudio piloto: estudio realizado con algunos individuos de la población objetivo o de la muestra de un estudio para poner a prueba y perfeccionar los instrumentos del estudio (cuestionario y manual de instrucciones, manual y programas para el procesamiento de datos) antes del levantamiento de datos principal con toda la población objetivo o toda la muestra.

Evaluación de los resultados de aprendizaje se refiere a la medición del éxito de los individuos en la consecución de los objetivos de aprendizaje. Pueden concentrarse en áreas concretas del currículo, servirse de diversos métodos de evaluación (escritos, orales, pruebas/exámenes prácticos, proyectos y expedientes) y ser administradas durante el transcurso de un programa educativo o a su finalización.

Evaluación obligatoria: evaluación de alto impacto que se administra a todos los alumnos de un curso concreto. Para el propósito de este estudio, sirven como ejemplo de evaluaciones obligatorias el General Certificate of Education Ordinary level (GCE-OL) o Advanced level (GCE-AL) del sistema británico, el Baccalaureate (Bacc) del sistema francés o el Diplôme d'étude collégiale (DEC) de Quebec-Canadá. Solo los alumnos que tengan el GCE-AL, Bacc, DEC o equivalente podrán inscribirse en educación superior o presentarse a las pruebas de selección para la enseñanza secundaria alta.

Evaluación a nivel de los centros educativos: evaluaciones organizadas y administradas regularmente por cada centro educativo de un país. Las herramientas de la evaluación suelen estar diseñadas por el personal docente del centro. Los resultados se utilizan para ofrecer la información directamente a estudiantes y padres, para monitorear el funcionamiento del curso y mejorar el proceso de enseñanza y aprendizaje. En algunos países, los resultados de estas evaluaciones cuentan (peso en la nota final) para la graduación y selección de los estudiantes.

Formato del estímulo - Texto (fuente: OECD 2015 PISA reading framework⁵, páginas 17-18)

Texto continuos: los textos continuos están compuestos por frases organizadas en párrafos. Entre otros, las noticias de los periódicos, los ensayos, las novelas, los relatos cortos, las reseñas y las cartas son textos corridos.

Texto discontinuo: Los **textos discontinuos**, también llamados documentos, están organizados de un modo distinto a los textos continuos, por lo que requieren un enfoque diferente en su lectura. Entre otros, las listas, las tablas, los gráficos, los diagramas, los anuncios, los horarios, los catálogos, los índices y los formularios son textos discontinuos. Estos textos pueden ser fijos o dinámicos.

Textos mixtos: los textos mixtos constan de una serie de elementos con formato corrido y discontinuo. Los componentes de un texto mixto bien construido (por ejemplo, una explicación en prosa que incluya un gráfico o una tabla) se refuerzan mutuamente mediante vínculos de coherencia y cohesión a nivel local y global. Es común encontrarlos en revistas, informes o libros de referencia en los que los autores se sirven de distintos métodos para comunicar la información.

Textos múltiples: Textos generados independientemente y que tienen sentido por sí solos; se juxtaponen para una ocasión concreta y podrían estar ligeramente vinculados, según el objetivo de la evaluación. La relación entre los textos no tiene por qué ser evidente; podrían ser complementarios o contradictorios. Por ejemplo, las páginas web de distintas empresas que ofrezcan asesoramiento sobre viajes podrían ofrecer indicaciones distintas o iguales a los

⁵<http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Reading%20Framework%20.pdf>

turistas. Los textos múltiples pueden tener un único formato «puro» (por ejemplo, continuo) o incluir tanto textos corridos como discontinuos.

Ítem: una única pregunta o tarea dentro de una prueba de una evaluación.

Nota: Un ítem tiene tres partes: el estímulo, que presenta el contexto de la tarea en formato de audio, vídeo o texto; una clara descripción de la tarea que debe realizarse a partir de los estímulos; y las posibles opciones de respuesta en un formato concreto (dicotómico, respuesta múltiple, crédito parcial, abierta, ensayo, etc.).

Mediciones:

Teoría clásica de los tests (TCT): Teoría de medición que consiste en una serie de supuestos sobre la relación entre los resultados reales u observados de las pruebas y los factores que afectan esos resultados. Se utiliza para medir y gestionar los datos del desempeño en las pruebas y los ítems. Al contrario que en la teoría de respuesta a los ítems, comprende un conjunto de métodos psicométricos más tradicionales (DFID⁶). Según la TCT, el nivel global de competencias de la persona evaluada depende de la dificultad general de la prueba y viceversa. No hay manera de obtener los parámetros de cada ítem ni las aptitudes de la persona evaluada en cada ítem en una escala continua de desempeño, al contrario que en la TRI o los modelos de Rasch.

Modelo de medición de Rasch: Conjunto de modelos matemáticos que permiten relacionar y predecir el resultado de un individuo para un ítem de una prueba, su nivel de desempeño según una escala de la capacidad o rasgo evaluados, y el parámetro de la dificultad del ítem. El modelo de Rasch considera que la probabilidad de que un individuo responda correctamente a un ítem o realice la tarea asignada de manera eficiente solo depende de la diferencia entre su nivel de capacidad y el nivel de dificultad del ítem. La probabilidad aumenta cuando la capacidad es superior a la dificultad del ítem, pero es del 50% si ambos parámetros son iguales.

Teoría de respuesta a los ítems (TRI): Conjunto de modelos matemáticos que permiten predecir el desempeño de un individuo en un ítem de una prueba basándose en su desempeño según una escala de la capacidad o rasgo evaluados y en los parámetros característicos del ítem: discriminación, dificultad y pseudo-azar o adivinación.

Nota: A diferencia de la teoría clásica de pruebas (TCP), tanto la TRI como el modelo de Rasch tienen potencial para estimar la capacidad de la persona evaluada independientemente de los ítems, así como las características del ítem, también independientemente del grupo de la prueba concreto. A esto se le conoce como propiedad de invariancia.

Método de ajuste a la ausencia de respuesta o no respuesta: método estadístico empleado para reemplazar los datos que falten o no sean válidos. La ponderación y la imputación son ejemplos de métodos de ajuste a la ausencia de respuesta que suelen usarse para contabilizar las no respuestas antes de proceder con los análisis.

Imputación: procedimiento estadístico empleado para determinar los valores de remplazo para los datos que falten o no sean válidos o coherentes. Algunos ejemplos de métodos de imputación frecuentes son: «hot deck», «cold deck», exclusión por lista y por par, imputación por la media, imputación por regresión, proyección de la última observación, imputación estocástica e imputaciones múltiples.

Método de evaluación del curso

Métrica de aprobado/reprobado: Las decisiones de aprobar o reprobar pueden determinarse con escalas de notas o de porcentajes. El formato de presentación de aprobado/reprobado es una variante de la presentación en función del nivel de rendimiento que solo tiene dos niveles: aprobado o reprobado. En ocasiones, el grupo de los aprobados puede ser subdividido en «aprobado con méritos», «aprobado primera división», etc. **Para el propósito de este catálogo, todos quedarán englobados en la categoría «aprobado».** Las decisiones de aprobar o reprobar pueden determinarse con escalas de notas o de porcentajes.

⁶<http://www.heart-resources.org/wp-content/uploads/2012/04/04-Learning-Outcomes-How-To-Note.pdf>

*Presentación típica de resultados individuales: El alumno ha aprobado la prueba. El alumno ha reprobado la prueba.
Presentación típica de resultados grupales: El 22% de los alumnos ha aprobado la prueba.*

Nota: Normalmente, la decisión de aprobar o reprobado está basada en una nota de corte o porcentaje establecido para la prueba. La nota de corte puede fijarse como el desempeño mínimo aceptado, por ejemplo, el 50% o el 80%. No hay una norma general sobre dónde situar el porcentaje de corte, ya que depende de la dificultad o contenidos de los ítems de las pruebas y del objetivo de la decisión de aprobar/reprobar. Desde el punto de vista operativo, el número de aprobados también puede depender de un número fijo de plazas disponibles. Por ejemplo, si solo hay 30 plazas y 1000 personas evaluadas, se considerará que el 3% que obtenga los mejores resultados habrá aprobado el examen, independientemente de sus notas, si el objetivo es cubrir esas 30 plazas. Una combinación también es posible: una nota mínima y un número máximo de plazas, es decir, dentro de quienes obtengan la nota mínima Y, solo 30 aprobarán; si la nota mínima es alcanzada por menos de 30 personas, la agencia que organice la evaluación podría bajar la nota mínima o dejar plazas sin cubrir (por ejemplo, en escuelas secundarias superiores).

Escala de porcentajes: El número de ítems de la prueba al que el alumno ha respondido correctamente, dividido entre el número total de ítems y multiplicado por 100 sería la nota en porcentaje obtenida para el alumno en esa prueba. En otras palabras, la nota es un porcentaje, no el número de respuestas correctas. Si una prueba incluye ítems valorados con varios puntos, el porcentaje obtenido por el alumno sería el total de puntos conseguido dividido entre el total de puntos posibles y multiplicado por 100.

Presentaciones típicas de resultados individuales: El alumno ha contestado correctamente al X% de los ítems de la prueba. El alumno ha logrado el X% del total de puntos posibles de la prueba.

Presentaciones típicas de resultados grupales: El porcentaje medio de respuestas correctas en el grupo fue el X%, con una desviación estándar del Y%. El porcentaje de puntos conseguidos osciló entre el X% y el Y% con una media de Z%.

Evaluación con referencia a criterios (Criterion-referenced assessment, CRA): los estudiantes son evaluados en función de un nivel predefinido de desempeño, o un objetivo, un desempeño deseado, un referente o un criterio. En las CRA se evalúa a todos los alumnos en función de ciertos criterios o referentes, que podrían ser un conjunto concreto de conocimientos y destrezas. En el ámbito educativo, los CRA suelen realizarse para comprobar que un alumno domina el material impartido en un curso o clase concreta.

Evaluación individual/ipsativa: consiste en evaluar el desempeño actual de una persona en función de su desempeño anterior con el objetivo de descubrir si ha habido alguna evolución. Esta práctica requiere datos longitudinales.

Evaluación con referencia a normas (Norm-referenced assessment, NRA): consiste en comparar a la persona evaluada con sus pares, por ejemplo, cuando las notas de los alumnos se clasifican de más alta a más baja, y el resultado del alumno se compara con el del resto. No se pretende interpretar los resultados ni saber lo que los alumnos saben o pueden hacer, salvo en el sentido limitado de que el desempeño de un alumno es similar al de otros alumnos con desempeños bajos, medios o altos dentro del grupo.

Presentación típica de resultados individuales: El alumno ha obtenido una nota de 450 (sobre una escala con media en 500 y desviación estándar de 100).

Presentación típica de resultados grupales: La nota media del curso X en la evaluación Y es S y la desviación estándar es D.

Notas: 1-En una evaluación normativa, los resultados suelen representarse como percentiles: a un alumno que haya obtenido la nota X se le asignará el percentil P, lo que implica que el P% de los alumnos habrán obtenido la nota X o inferior. Así, la NRA está concebida para clasificar a los alumnos "según la curva", y no para comprobar si cumplen unos estándares o criterios. Por lo tanto, no deben usarse NRA para determinar si los alumnos cumplen los estándares deseados (o mínimos). 2- En teoría, una prueba puede ser a la vez normativa y basada en criterios. Una CRA puede ser útil para medir el aprendizaje de un alumno en comparación con la norma si se incluyen notas de corte en la escala para distinguir los niveles de desempeño. En ese caso puede elaborarse una clasificación con los resultados. Sin embargo, esta podría no ser lo suficientemente precisa para los objetivos de la NRA. Cabría considerar usar una NRA como CRA, pero podría carecer de los contenidos y precisión suficientes para determinar el

nivel de conocimientos y aptitudes. Solo debería usarse una prueba para ambos fines cuando estos hayan sido claramente definidos durante el diseño de la prueba.

Métricas de presentación de resultados: hace referencia a las distintas formas de presentar los resultados de las evaluaciones de aprendizaje, o los logros de los estudiantes, para comunicárselos a las partes interesadas en el ámbito de la educación, incluyendo los mismos estudiantes. Los resultados pueden presentarse para individuos o agregados para grupos específicos. Algunas formas de presentar los resultados son los porcentajes de aprobados o que reprobaron, las escalas y los niveles de rendimiento.

Métrica por niveles de rendimiento: clasificación de alumnos por franjas de desempeño identificadas mediante una serie de notas de corte situadas en la escala total de desempeño. Los niveles de rendimiento se utilizan con frecuencia en las pruebas basadas en criterios. Debería haber descripciones de lo que significa pertenecer a cada nivel respecto de los conocimientos, destrezas, aptitudes, etc. Cada nivel (o franja) representa el grado de dominio de lo que la prueba pretende medir. Los mismos niveles de rendimiento pueden expresarse con palabras o letras. Por ejemplo, «inferior al básico», «básico», «por encima del básico»; o alto, medio, bajo; o A, B, C.

Presentación típica de resultados individuales: El alumno está clasificado en nivel 2, o básico. El alumno está clasificado en nivel 4, o avanzado.

Presentación típica de resultados grupales: En los centros educativos de este tipo, X% pertenece al nivel 1 (inferior al básico), Y% pertenece al nivel 2 (básico), Z% pertenece al nivel 3 (competente) y W% pertenece al nivel 4 (avanzado). En esta escuela, el X% de los alumnos se encuentra en el nivel 2 (competente) o superior.

Nota: Los niveles de rendimiento suelen asociarse a los puntajes de las escalas de rendimiento. La determinación de las notas de corte de una escala concreta comienza con la definición y descripción de lo que significan los niveles de rendimiento en materia de conocimientos y destrezas, evalúa los ítems de la escala en función de su contenido y complejidad y, por último, sopesa sistemáticamente los puntos en los que colocar las notas de corte.

Muestreo no probabilístico:

Muestreo de conveniencia: solo se escogen unidades muestrales de acceso fácil y cómodo.

Muestreo discrecional: La muestra se escoge en función de ciertos juicios sobre la población general. La hipótesis subyacente es que el investigador seleccionará unidades características de la población.

Muestreo por cuotas: Se efectúa el muestreo hasta que se haya seleccionado un número concreto de unidades (cuotas) para diversas subpoblaciones. Al no haber reglas que dicten el modo en que deben ser cubiertas las cuotas, el muestreo por cuotas es en realidad una manera de cumplir los objetivos de tamaño de las muestras en ciertas subpoblaciones.

Muestreo voluntario: Este tipo de muestreo se da cuando la gente se ofrece a participar en el estudio.

Muestreo probabilístico:

Muestreo aleatorio simple: cada miembro de una población tiene las mismas probabilidades de ser incluido en la muestra. Para que eso ocurra todas las unidades de la población del estudio deben aparecer en la lista.

Muestreo con probabilidad proporcional al tamaño: El muestreo probabilístico requiere que cada miembro de la población estudiada tenga posibilidades de ser incluido en la muestra, pero no exige que esta probabilidad sea la misma para todos. Si hay información disponible sobre el tamaño de cada unidad (ej. número de estudiantes por escuela o aula) y si esas unidades son de tamaño variable, esta información puede emplearse en la selección del muestreo para aumentar su eficiencia. Esto se conoce como muestreo con probabilidad proporcional al tamaño (PPT). Según este método, cuanto mayor sea el tamaño de la unidad, mayores serán sus posibilidades de ser incluida en la muestra. Para que este método sea más eficiente, la medición de las unidades debe ser precisa.

Muestreo estratificado: la población se divide en grupos homogéneos, grupos excluyentes entre sí llamados estratos, tras lo que se seleccionan muestras independientes de cada estrato. Cualquiera de los métodos de muestreo puede usarse dentro de cada estrato, y el método de muestreo puede variar de un estrato a otro. Si la muestra dentro de cada estrato se elige mediante un muestreo aleatorio simple, el plan se denominará muestreo aleatorio simple estratificado. Una población puede estratificarse antes del muestreo según cualquier variable que esté disponible en todas las unidades del marco muestral (ej. notas de los alumnos, sexo, edad, provincia/región/área de residencia, titularidad del centro educativo).

Muestreo multietápico: es similar al método por conglomerado, salvo que se toma una muestra dentro de cada grupo seleccionado, en lugar de incluir todas las unidades del grupo. Este tipo de muestreo requiere al menos dos etapas. En la primera etapa se identifican y seleccionan grandes grupos o agregados. Estos agregados contienen más unidades poblacionales de las necesarias para la muestra final. En la segunda etapa, las unidades poblacionales se escogen a partir de los grupos seleccionados (mediante cualquier método de muestreo probabilístico) para configurar la muestra definitiva. Si hay más de dos etapas, el proceso de escoger unidades poblacionales dentro de los grupos continúa hasta obtener la muestra definitiva.

Muestreo multifásico: Una muestra multifásica recopila información básica de una amplia muestra de unidades y a continuación recaba información más detallada de una submuestra de esas unidades. La forma más común de muestreo multifásico es el muestreo en dos fases (o muestreo doble), pero también es posible realizar tres o más fases.

El muestreo multifásico es bastante distinto al multietápico, a pesar de lo similar de sus nombres. Aunque el muestreo multifásico también conlleva la toma de dos o más muestras, todas ellas se extraen del mismo marco, y la estructura de las unidades es igual en cada fase. Sin embargo, al igual que en el muestreo multietápico, cuantas más fases se utilicen, más complejo será el plan de muestreo y la estimación.

Muestreo por conglomerado: En primer lugar se divide la población en grupos o agregados. Se elige aleatoriamente una serie de grupos para que represente al total de la población, y todas las unidades de los grupos seleccionados se incluyen en la muestra. La muestra no incluye ninguna unidad de grupos no seleccionados (quedan representadas por las incluidas en los grupos seleccionados). Es un procedimiento distinto al muestreo estratificado, en el que se escogen algunas unidades de cada grupo.

Muestreo sistemático o por intervalos: aquel en el que existe un salto o intervalo entre cada unidad seleccionada de la muestra. Para crear una muestra sistemática es preciso seguir los siguientes pasos:

- i. Numerar de 1 a N las unidades del marco (donde N es el tamaño total de la población).
- ii. Determinar el intervalo de muestreo (K) dividiendo el número de unidades de la población entre el tamaño deseado de la muestra.
- iii. Seleccionar un número al azar entre 1 y K. Este número se conoce como inicio aleatorio y será el primer número incluido en la muestra.
- iv. Elegir cada K^a unidad después de ese número.

Nivel de impacto de la evaluación

Evaluación de alto impacto (high-stake assessment): evaluación con importantes consecuencias para la **persona evaluada** basada en su desempeño. Aprobar conlleva importantes beneficios, como pasar al curso siguiente, obtener el diploma de estudios secundarios o una beca, acceder al mercado laboral o recibir la licencia para ejercer una profesión. La reprobación también tiene consecuencias, como la obligación de asistir a clases de recuperación o la incapacidad de ejercer una profesión. Algunos ejemplos de evaluaciones de alto impacto son los exámenes de acceso a la universidad, los exámenes de fin de estudios secundarios y los exámenes de acceso a una profesión.

Evaluación de bajo impacto (low-stake assessment): aquella que tiene el objetivo de comprobar el funcionamiento del sistema educativo (o escolar).

Las evaluaciones de bajo impacto no tienen consecuencias directas sobre las personas evaluadas. Por ejemplo, independientemente de sus resultados, los estudiantes no tendrán que repetir curso ni se les denegará el acceso al siguiente nivel educativo, ni se les derivará a programas más o menos académicos.

Sin embargo, esta neutralidad podría no aplicarse a todas las partes implicadas. Las repercusiones podrían ser elevadas si las autoridades competentes toman medidas (despidos, repetir la formación, etc.) respecto de los profesores y centros educativos cuyos alumnos obtienen peores resultados. El gobierno podría ser objeto de críticas por parte de los partidos de la oposición si los resultados globales son bajos o empeoran. Por lo tanto, mientras que la evaluación puede tener un bajo impacto en para las personas evaluadas o alumnos, las consecuencias podrían ser importantes para otros.

Nota de corte: punto concreto de la escala de puntuación de una prueba que determina una interpretación específica de los resultados, dependiendo de si estos quedan situados en ese punto, por encima o por debajo de él.

Procesamiento de datos: serie de operaciones manuales, automáticas o electrónicas como la validación, la clasificación, la creación de resúmenes, o la integración de datos. Estas operaciones suelen ir seguidas de la recuperación, la transformación, la clasificación, el análisis y la presentación de los datos.

Prueba adaptativa informatizada (CAT): método de administración de las pruebas que adapta la lista de ítems administrados en la prueba a los logros del alumno, determinados a partir de sus respuestas anteriores. El modelo CAT es muy distinto a las pruebas fijas informatizadas, ya que selecciona los ítems en función de las respuestas anteriores de cada alumno evaluado y es capaz de memorizar la manera en que éste busca, almacena y recupera la información.

Prueba fija informatizada (CFT): la versión electrónica de las pruebas tipo «papel y lápiz» orientada a facilitar el procesamiento de datos y reducir la probabilidad de errores humanos a la hora de confeccionar una base de datos.

Prueba estandarizada: Aquella en la que los ítems/tareas o preguntas, las condiciones administrativas, la edición, la puntuación y la interpretación de los resultados se implementan de manera consistente y predeterminada para todas las personas evaluadas.

Puntajes escalares (scale scores): Los puntajes o valores escalares son estimaciones de la habilidad de los estudiantes derivados de modelos de Respuesta al Ítem (TRI). Estas estimaciones constituyen transformaciones generadas a partir de los vectores de respuesta de los estudiantes. Los puntajes escalares están diseñados para proveer una métrica común para los estudiantes que tomaron versiones diferentes (o paralelas) de una prueba en el mismo año y/o el mismo test en distintos años.

Rasgo/constructo: aspecto o rasgo hipotético de índole mental no observable que se utiliza para explicar el desempeño de un individuo en una evaluación. Solo se mide a través de observaciones o del desempeño en tareas a partir de las que se infiere el nivel de la persona evaluada. Los constructos no siempre pueden observarse o medirse directamente.

Tasa de participación: número total de individuos de la muestra escogida que ha participado en el estudio (i.e. realizado el cuestionario del estudio), expresado como porcentaje del número inicial de individuos seleccionados (tamaño muestral) o del total de la población objetivo.

Tasa de participación ponderada: tasa de participación de una muestra específica que recoge el peso de cada individuo que ha participado en el estudio.

Tipos de evaluaciones:

Evaluación internacional: evaluación de los resultados de aprendizaje que ofrece información similar a la evaluación nacional, pero que abarca varios sistemas educativos nacionales. Su principal objetivo es establecer una comparación entre los resultados de los países participantes.

Evaluación nacional: En el contexto del estudio del catálogo del UIS, se refiere a una evaluación de los resultados de aprendizaje que pretende describir los logros de los alumnos de cierta edad o curso y arrojar datos sobre un número limitado de medidas de resultados consideradas importantes para el sistema educativo por parte de los políticos, quienes elaboran las políticas y la comunidad educativa en general. Por lo general se administra a una

muestra de estudiantes y recoge asimismo información contextual (de estudiantes, profesores y padres) importante para vincular el análisis con los planteamientos de políticas a nivel nacional, regional y local.

Exámenes públicos: En el contexto del catálogo del UIS, designa exámenes estandarizados de salida o de fin de estudios, normalmente fijados por un consejo evaluador estatal o regional para promover, seleccionar u otorgar un certificado a **todos** los candidatos que hayan aprobado o se presuponga que han adquirido de manera formal o informal los contenidos del currículo de un programa educativo formal, como parte de los requisitos para su graduación. Los exámenes públicos suelen administrarse anualmente a todo el que se inscriba en ellos, independientemente de su edad y profesión (en algunos países). Al contrario que en las evaluaciones nacionales, en los exámenes públicos no suele recogerse información contextual de los alumnos.

Valores plausible: son estimaciones de la habilidad de los estudiantes producidos por medio de una combinación de técnicas de modelos de respuesta al ítem e imputaciones sobre la base de modelos de regresión latente. Aunque los **valores plausibles** no son resultados realmente observados de las pruebas, ellos tienen la propiedad del resultado de una prueba o test y permite a los analistas usar programas informáticos y técnicas estándares para analizar datos que han sido recopilados mediante una evaluación con un diseño muestral matricial complejo.