

# Guía para la sistematización de resultados de las evaluaciones

---



---



## **Guía para la sistematización de resultados de las evaluaciones**

Primera edición, 2019

### **D. R. © Instituto Nacional para la Evaluación de la Educación**

Barranca del Muerto 341, Col. San José Insurgentes,  
Del. Benito Juárez, C.P. 03900, Ciudad de México.

La elaboración de este documento estuvo a cargo de la Unidad de Normatividad y Política Educativa. El contenido, la presentación, así como la disposición en conjunto y de cada página de esta obra son propiedad del INEE. Se autoriza su reproducción por cualquier sistema mecánico o electrónico, para fines no comerciales. Cítese de la siguiente manera:

INEE (2019). *Guía para la sistematización de resultados de las evaluaciones*. México: autor.

# Índice

Presentación.....	4
<b>1. Principios básicos de la evaluación educativa.....</b>	<b>6</b>
Objetivo de la guía.....	6
Fases y pasos para el desarrollo, el uso y el mantenimiento de instrumentos de evaluación.....	7
Componentes de la guía.....	7
Recomendaciones para la lectura, aprovechamiento y uso de la guía.....	8
<b>2. Fase 4. Análisis de resultados del instrumento de evaluación.....</b>	<b>10</b>
Paso 8. Evaluación de la métrica del instrumento.....	11
Paso 9. Resultados de las puntuaciones obtenidas a partir del instrumento.....	27
<b>3. Análisis de resultados en el enfoque cualitativo.....</b>	<b>32</b>
Fase 7. Análisis de datos.....	33
Fase 8. Interpretación de resultados.....	35
Portafolios de evidencias del análisis de resultados en el enfoque cualitativo.....	36
Lista de cotejo para el análisis de resultados en el enfoque cualitativo.....	36
Glosario.....	37
Bibliografía.....	43
Siglas y acrónimos.....	46
<b>Anexo</b>	
Principales características y técnicas del enfoque cualitativo.....	47
Listas de cotejo general. Fase 4. Análisis de resultados del instrumento de Evaluación. Paso 8. Evaluación de la métrica del instrumento.....	58

## Presentación

El propósito de esta guía es dar continuidad al trabajo realizado en las guías previas y orientar a las autoridades educativas locales (AEL) en el análisis de los resultados de los Instrumentos de Evaluación de los Proyectos de Evaluación y Mejora Educativa (PROEME) y de los subproyectos que se derivan del Proyecto Nacional de Evaluación y Mejora Educativa de las Escuelas Multigrado (PRONAEME).

Los PROEME se clasifican, según su punto de partida, en tres grandes categorías (INEE, 2017, p. 17):

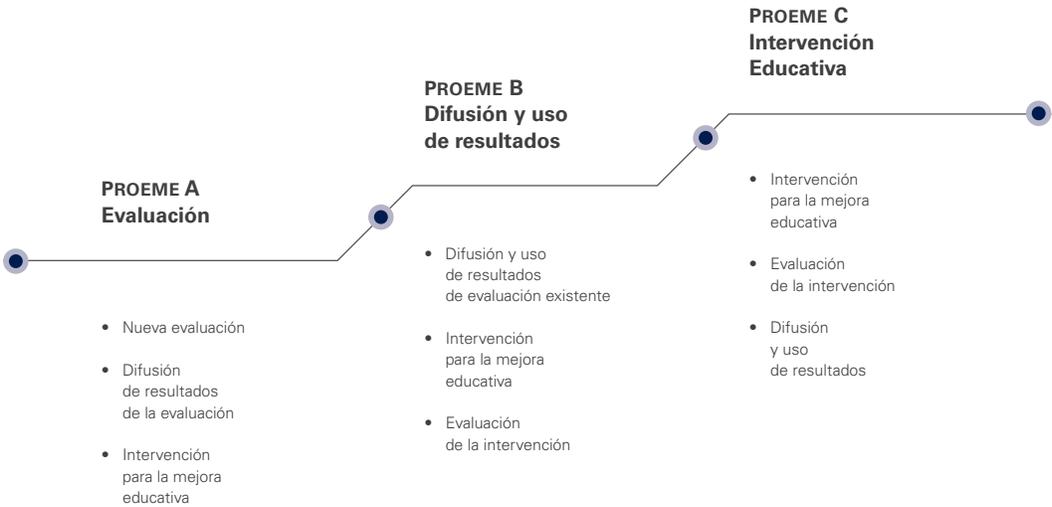
- *Proyectos de evaluación* (PROEME A): son aquellos que, no habiendo identificado evaluaciones pertinentes sobre las causas asociadas a determinado problema educativo, plantean el desarrollo de nuevas evaluaciones cuyos resultados eventualmente serán difundidos a distintas audiencias, se promoverá su uso entre distintos actores educativos y darán sustento a la definición de intervenciones orientadas a la mejora.
- *Proyectos de difusión y uso* (PROEME B): reconociendo la existencia de evaluaciones, éstos plantean difundirlas de manera suficiente y adecuada, usar sus resultados para definir intervenciones sustentadas para la mejora y promover la cultura de la evaluación en las distintas audiencias identificadas.
- *Proyectos de intervención* (PROEME C): éstos plantean que, dada la existencia de evaluaciones suficientemente difundidas y la identificación de usos exitosos probados, pueden definirse intervenciones sustentadas en evidencia y orientadas a la mejora.

Los PROEME deberán cumplir un ciclo. En general, todos los proyectos deberán desarrollar alguna evaluación:

- PROEME A: evaluación-difusión y uso-intervención.
- PROEME B: difusión y uso-intervención-evaluación de la intervención.
- PROEME C: intervención-evaluación de la intervención-difusión y uso.

En la figura 1 se muestran el punto de inicio y el ciclo de acuerdo al tipo de PROEME A.

Figura 1. Tipo de PROEME



Fuente: elaboración propia a partir de INEE, 2017, p. 17.

# 1. Principios básicos de la evaluación educativa

Esta guía ha sido elaborada para los Proyectos de Evaluación y Mejora Educativa (PROEME) A, que inician con el diseño de nuevos instrumentos de evaluación en los que aplicarán los principios de validez, confiabilidad y equidad, y, asimismo, los de la evaluación educativa: la mejora educativa, la equidad, la justicia, la diversidad y la participación.

La concepción de evaluación que orienta la guía es la establecida en el artículo 6 de la Ley del Instituto Nacional para la Evaluación de la Educación (LINEE), que consiste en “la acción de emitir juicios de valor que resultan de comparar los resultados de una medición u observación de componentes, procesos o resultados del Sistema Educativo Nacional (SEN) con un referente previamente establecido” (INEE, 2013).

La guía se ha diseñado a partir de los Criterios Técnicos para el Desarrollo, Uso y Mantenimiento de Instrumentos de Evaluación (CTDUMIE), que establecen que “Los principios básicos para todo instrumento de evaluación que tiene como propósito la medición son la validez, confiabilidad y equidad” (DOF, 2017, 28 de abril, p. 4).

La guía presenta la fase 4 para el análisis de los resultados del instrumento de evaluación, necesario para la recolección de datos de los Proyectos de Evaluación en la Educación Obligatoria: Educación Básica y Educación Media Superior; de acuerdo con las categorías establecidas en el Programa de Mediano Plazo del Sistema Nacional de Evaluación Educativa (PMP-SNEE) 2016-2020:

- Logro educativo.
- Docentes, directivos, supervisores y asesores técnicos pedagógicos.
- Currículo, materiales y métodos educativos.
- Organización escolar y gestión del aprendizaje.
- Condiciones de la oferta educativa.
- Políticas, programas y sistemas de información.

Debido a los enfoques adoptados para la investigación de cada PROEME, la guía se ha dividido en dos grandes bloques para orientar el análisis de los resultados de proyectos que aplican un enfoque cuantitativo (fase 4) y para el análisis de los resultados de proyectos que aplican un enfoque cualitativo.

## Objetivo de la guía

El objetivo general de esta guía es proporcionar elementos conceptuales y metodológicos para el análisis de los resultados de los instrumentos de evaluación de los Proyectos de Evaluación, de Difusión y Uso, y de Intervención de los Programas Estatales de Evaluación

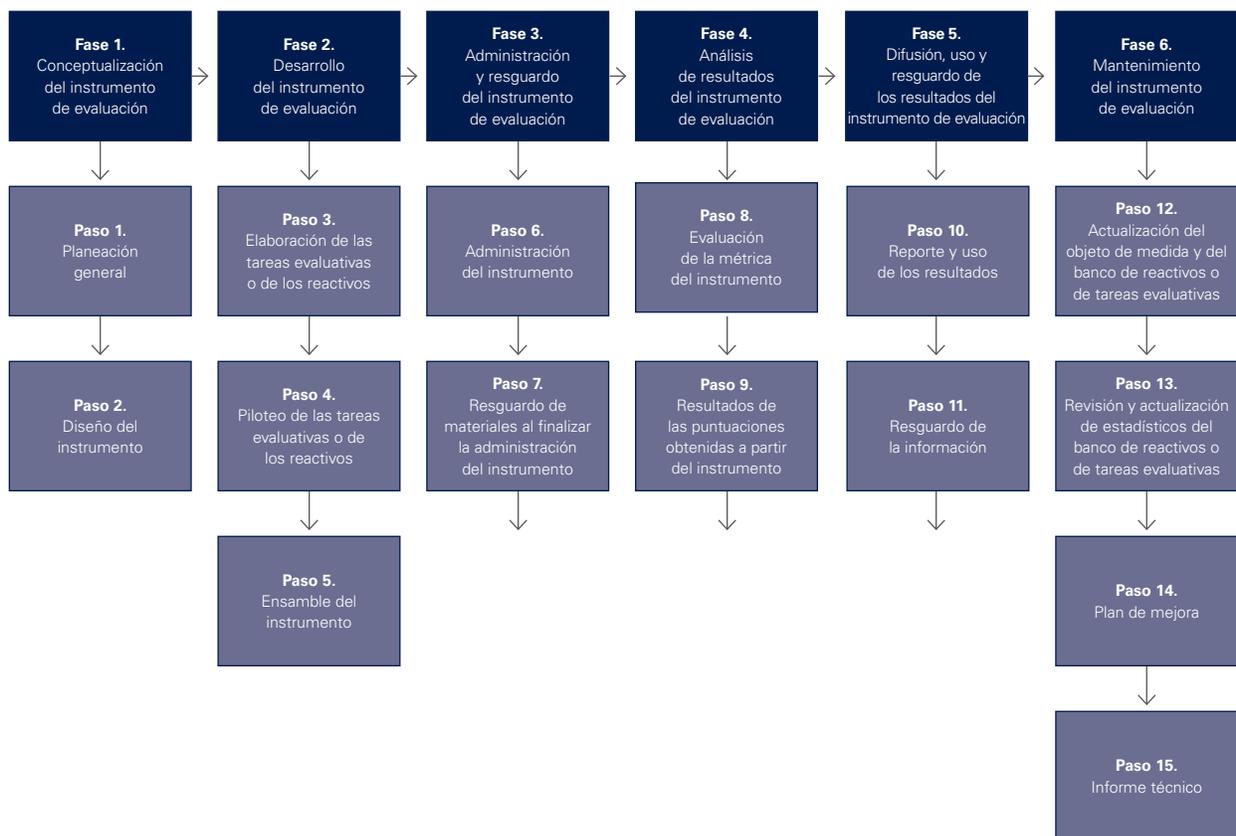
y Mejora Educativa (PEEME) con apego a los CTDUMIE emitidos por el Instituto Nacional para la Evaluación de la Educación (INEE).

## Fases y pasos para el desarrollo, el uso y el mantenimiento de instrumentos de evaluación

La figura 1.1 muestra las seis fases para el desarrollo, el uso y el mantenimiento de instrumentos de evaluación, de acuerdo con los CTDUMIE.

El alcance de esta guía sólo considera la Fase 4, que corresponde al Análisis de resultados del instrumento de evaluación. Las fases 5 y 6 serán abordadas en las guías posteriores.

**Figura 1.1. Fases y pasos para el desarrollo de instrumentos de evaluación**



Fuente: elaboración propia a partir de los CTDUMIE (DOF, 2017, 28 de abril).

## Componentes de la guía

Para una orientación más didáctica, dentro de la Fase 4 y el enfoque cualitativo se incluyen los siguientes componentes:

- *Diagrama*. Al inicio se muestra un diagrama con los pasos que se trabajarán en la fase y se especifican las evidencias a desarrollar.
- *Diagrama de pasos*. Por cada paso se muestra el diagrama con las actividades a realizar y las evidencias a generar.
- *Apoyo conceptual y metodológico*. Se incluyen los conceptos de los términos esenciales de los pasos o actividades, y, en caso necesario, se hace una breve explicación metodológica para realizar las actividades.
- *Portafolio de evidencias<sup>1</sup> (PE)*. Es un documento en formato DOCX que contiene la estructura de las fichas técnicas y evidencias de formato libre (EFL). Al ir desarrollando las fichas técnicas y las EFL integrará el portafolio de evidencias de cada paso.<sup>2</sup>
- *Lista de cotejo de cada paso*. Al final de cada paso se incluye una lista de cotejo para confirmar que los productos obtenidos cubren los criterios de calidad.
- *Lista de cotejo de fase*. Al concluir esta fase, se incluye una lista de cotejo para confirmar que se desarrollaron todas las evidencias de la fase.

Al final del documento se incluye:

- *Bibliografía consultada y sugerida*. Se presentan las referencias de los textos citados o consultados y los textos sugeridos para complementar los temas de la guía.

## Recomendaciones para la lectura, aprovechamiento y uso de la guía

1. Para la revisión de la guía, se sugiere identificar el enfoque cualitativo o cuantitativo que orienta el PROEME.
2. Para el enfoque cuantitativo (Fase 4):
  - a. Identificar los pasos y evidencias a obtener en la fase.
  - b. Identificar las evidencias a obtener en cada paso.
  - c. Realizar las actividades en orden y considerar los productos que ya se han generado desde la Guía para el desarrollo del Programa Estatal de Evaluación y Mejora Educativa (PEEME); la Guía 1, para la elaboración de Marcos de Referencia de Evaluaciones y difusión y uso de resultados de las Evaluaciones; la Guía 2, para la elaboración de Estrategias de Difusión y uso de resultados de las evaluaciones, y la Guía 3, para la elaboración del instrumento de evaluación.

<sup>1</sup> Es importante aclarar que, de existir análisis psicométricos personales, o si alguna de las pruebas es personalizada y no anónima, debe considerarse esta información como sensible y apegada al tratamiento establecido sobre información y datos personales en la Ley Federal de Transparencia y Acceso a la Información Pública. En este caso es necesario aplicar medidas de seguridad en el manejo de los archivos, que pueden ser desde una encriptación en documento PDF hasta el desarrollo de sitios seguros (https) para compartir la información.

<sup>2</sup> En total se generarán dos portafolios de evidencias, uno por cada paso. El nombre del archivo se integrará de la siguiente forma: PE\_P8\_P0\_000.docx, que corresponde al paso 8, y PE\_P9\_P0\_000.docx, que corresponde al paso 9. Para identificar los PE que entreguen las entidades, se requerirá que en el nombre original del portafolio de evidencias, PE\_P8\_P0\_000.docx, se sustituya P0 por el número de PROEME que le corresponda en la entidad, y 000 con las tres letras que identifiquen el nombre de la entidad. Por ejemplo: PE\_P8\_P3\_SON.docx (portafolio de evidencias del paso 8 del Proeme 3 de Sonora).

- d. Elaborar las fichas técnicas y evidencias de formato libre incluidas en el portafolio de evidencias de cada paso.
- e. Aplicar la lista de cotejo para cada paso realizado y verificar que se da cumplimiento a los criterios de calidad de cada documento.
- f. Al finalizar la fase, aplicar la lista de cotejo para confirmar que se han desarrollado todas las evidencias especificadas.

### 3. Para el enfoque cualitativo.

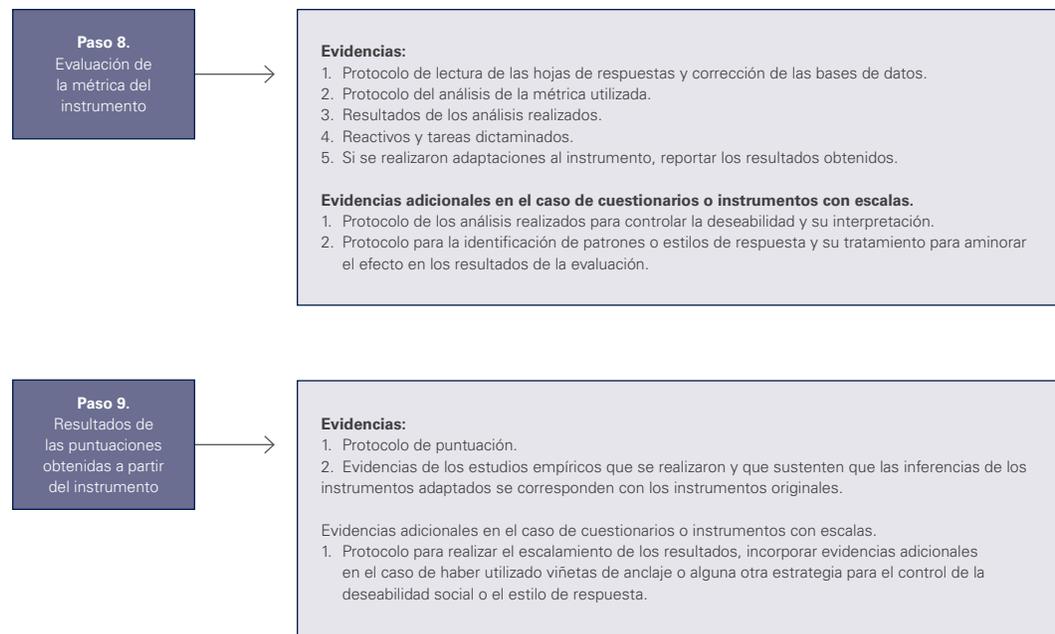
- a. Recordar las principales características y técnicas del enfoque cualitativo (anexo 1).
- b. Identificar las evidencias a obtener en cada paso.
  - Paso A. Estructurar los datos recopilados.
    - Organización de datos.
    - Transcripción de materiales.
  - Paso B. Análisis del material recopilado.
  - Paso C. Interpretación de resultados.
- c. Realizar las actividades en orden y considerar los productos que ya se han generado desde la Guía para el desarrollo del Programa Estatal de Evaluación y Mejora Educativa (PEEME); la Guía 1, para la elaboración de Marcos de Referencia de Evaluaciones y difusión y uso de resultados de las Evaluaciones; la Guía 2, para la elaboración de Estrategias de Difusión y uso de resultados de las evaluaciones, y la Guía 3, para la elaboración del instrumento de evaluación.
- d. Elaborar las fichas técnicas y evidencias de formato libre incluidas en el portafolio de evidencias de cada paso.
- e. Aplicar la lista de cotejo para cada paso realizado y verificar que se da cumplimiento a los criterios de calidad de cada documento.
- f. Al finalizar, aplicar la lista de cotejo para confirmar que se han desarrollado todas las evidencias especificadas.

## 2. Fase 4. Análisis de resultados del instrumento de evaluación

En esta fase se mencionan los elementos que deben considerarse para el análisis estadístico del instrumento y de los reactivos o las tareas evaluativas que lo conforman, así como el modelo de puntuación utilizado. Conocer las propiedades métricas de los reactivos permite identificar los que son aptos para ser considerados en la puntuación de los sujetos, así como aquellos que no aportan información relevante o son defectuosos, por lo que deben ser excluidos (Cook y Beckman, 2006; Downing, 2004; Stemler y Tsai, 2008) (DOF, 2017, 28 de abril, p. 21).

En la figura 2.1 se muestran los pasos de la Fase 4 asociados a las evidencias a desarrollar.

Figura 2.1. Fase 4. Análisis de resultados del instrumento de evaluación. Pasos y Evidencias



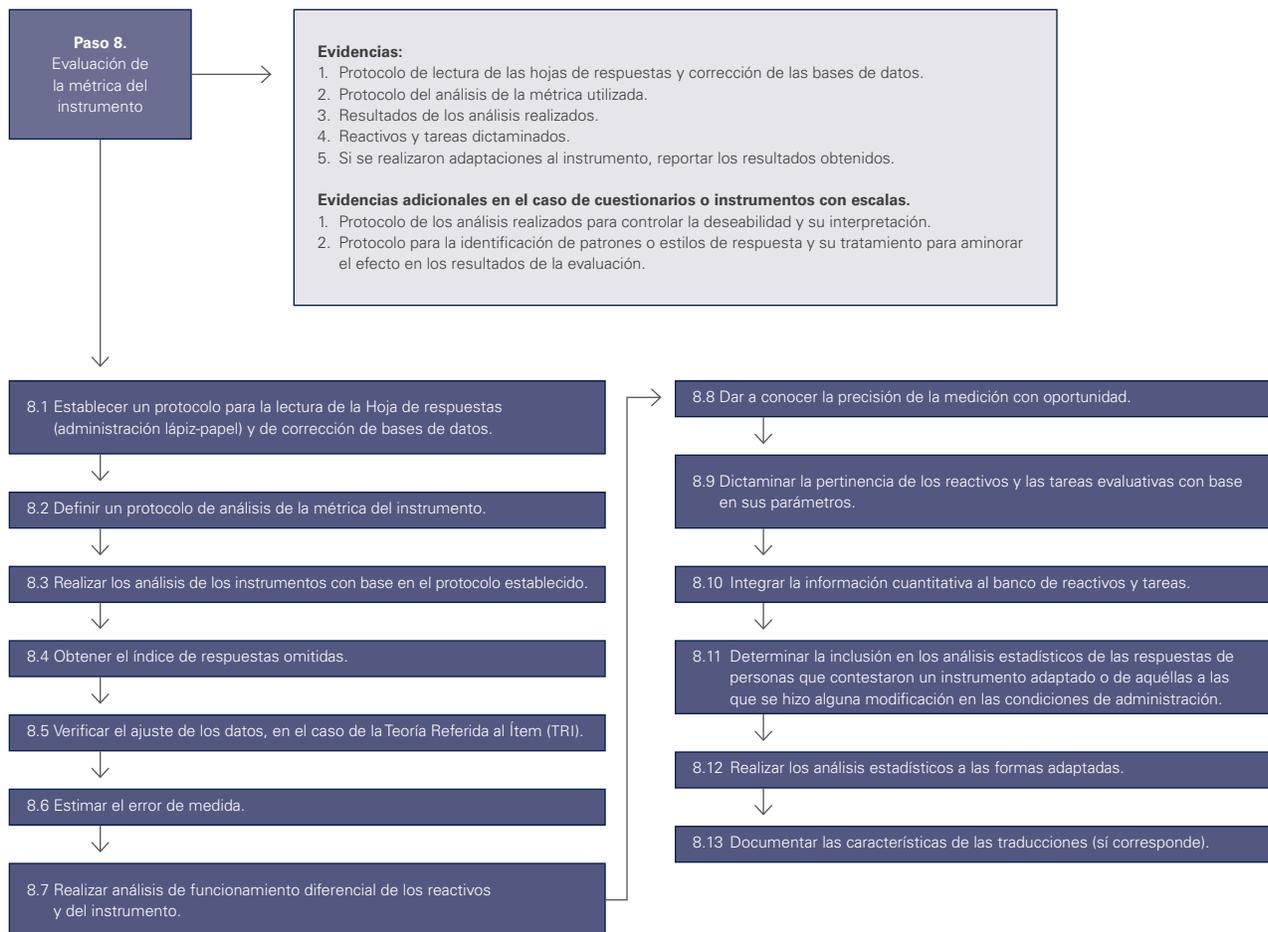
Fuente: elaboración propia a partir de los Criterios Técnicos para el Desarrollo, Uso y Mantenimiento de Instrumentos de Evaluación (CTDUMIE) (DOF, 2017, 28 de abril).

## Paso 8. Evaluación de la métrica del instrumento

En este paso se llevan a cabo los análisis estadísticos considerando las respuestas a los reactivos o las tareas evaluativas, con el objetivo de aportar evidencia para la validez de las inferencias que se desean obtener a partir de los datos recopilados. Esta información es de vital importancia, debido a que el propósito de este paso es fundamentar empíricamente las características del instrumento de evaluación (DOF, 2017, 28 de abril, p. 21).

En figura 2.2 se muestran las actividades y evidencias a obtener en el paso 8.

**Figura 2.2. Paso 8. Evaluación de la métrica del instrumento**



Fuente: elaboración propia a partir de los CTDUMIE (DOF, 2017, 28 de abril).

## 8.1. Establecer un protocolo para la lectura de la Hoja de respuestas (administración lápiz-papel) y de corrección de bases de datos

Antes de iniciar el análisis estadístico, es conveniente contar con los procedimientos documentados para la digitalización y la sistematización de las respuestas de las personas, además de los mecanismos que se emplean para depurar las bases de datos con el propósito de minimizar los errores de codificación (DOF, 2017, 28 de abril, p. 21).

### Protocolo para la lectura de respuestas de las personas

El protocolo tendrá que responder a la pregunta: ¿cómo se van a codificar las respuestas del instrumento de evaluación?

La codificación de las respuestas tendrá que considerar desde los datos de los participantes en el instrumento aplicado hasta las respuestas dadas a cada reactivo o tarea evaluativa.

La estructura propuesta para la elaboración del protocolo es la siguiente:<sup>3</sup>

- Título del instrumento al que se refiere el protocolo.
- Área solicitante de la lectura del instrumento.
- Fecha de solicitud.
- Breve descripción del instrumento de evaluación.
- Productos que recibirá el área solicitante de la lectura del instrumento.
- Referentes del levantamiento de la información.
  - Cantidad de instrumentos aplicados.
  - Estructura del instrumento.
  - Tipo de preguntas que incluye el instrumento (selección de respuesta especificando la cantidad de opciones, respuesta abierta).
  - En caso de la aplicación de varios instrumentos de evaluación (para alumnos, profesores, padres, etc.), especificar si se requiere que la información esté ligada entre ellos.
- Describir detalladamente el instrumento aplicado del cual se requiere la lectura.
  - Forma en que se deberán leer las respuestas.<sup>4</sup>
  - Codificación de los reactivos especificando cuando hay variedad en la cantidad de opciones a seleccionar (con cuatro, cinco o más opciones de respuesta), se deberá incluir una descripción de los reactivos.

<sup>3</sup> Si en su Proyecto de Evaluación y Mejora Educativa (PROEME) se aplicaran diferentes instrumentos, deberá elaborar un protocolo para cada uno o una sección que corresponda a cada instrumento.

<sup>4</sup> Es recomendable incluir en el protocolo la imagen de la hoja de respuestas con sus códigos a utilizar en la lectura y verificar que corresponde a la última versión.

- Especificar los códigos a utilizar para los datos de los participantes (sexo, edad, etcétera).

Aspecto	Nombre del campo en tabla	Código
Sexo	sexo	1 Hombre 2 Mujer 8 Respuesta múltiple* 9 Respuesta omitida**

\* Corresponde a que se seleccionó más de una opción.

\*\* Corresponde a que no se seleccionó ninguna opción.

- Especificar los códigos a utilizar para las preguntas del instrumento de evaluación según la organización del mismo, asignando letras y números que permitan su fácil identificación.
- Tabla de datos en donde se integrarán las respuestas de los instrumentos.
  - Estructura de la tabla de datos (campo y descripción de cada campo).

Un ejemplo de Protocolo de lectura de hojas de respuesta (EXCALE-INEE) lo puede consultar en la siguiente liga:

<http://www.inee.edu.mx/index.php/normateca-descentralizada/documentos-del-inee/comite-de-validez/documentacion-sobre-los-excale/excale-cd/especificaciones-y-resultados-de-piloteo/especificaciones/excale09-piloteo-2008-especificaciones-de-lectura/download>

A continuación se mencionan varios aspectos básicos para la asignación de puntuaciones (Navas, Fidalgo, Gabriel, Suárez, Briosó, Gil, Sarria, 2010):

- La asignación de puntuaciones al test permite representar numéricamente los indicadores empíricos —las respuestas a los ítems del test— utilizados, así como representar empíricamente la característica de interés que evalúa el test.
- La asignación de puntuaciones a un test consiste en asignar previamente puntuaciones a los ítems que lo componen para posteriormente combinarlas en una única —la puntuación del test—, que expresa el valor de escala del sujeto en el rasgo o la característica medida por la prueba.
- En pruebas cognitivas con ítems de elección la asignación de puntuaciones a los ítems supone el uso de una plantilla de corrección: si la respuesta dada por el sujeto coincide con la indicada como correcta en la plantilla, se le asigna 1 a la respuesta, y si no coincide, 0.
- En pruebas no cognitivas con ítems de respuesta cerrada se asigna como puntuación al ítem el código numérico correspondiente a la alternativa o categoría elegida por el sujeto.
- En pruebas no cognitivas con ítems de respuesta cerrada con enunciados positivos y negativos es preciso recodificar apropiadamente las categorías de respuesta con el fin de que un número signifique siempre lo mismo (una actitud/sentimiento/opinión favorable o desfavorable), con independencia de que el enunciado del ítem denote una actitud/sentimiento/opinión favorable o desfavorable hacia la cuestión examinada.

- En pruebas cognitivas con ítems de respuesta corta se confecciona un listado de posibles respuestas aceptables y se asigna 1 si la respuesta dada está en la lista y 0 en caso contrario.
- En pruebas no cognitivas con ítems de respuesta abierta, para poder asignar puntuaciones a los ítems es preciso categorizar las respuestas: establecer un sistema de categorías con las posibles respuestas y asignar a cada respuesta el código correspondiente a la categoría de la respuesta en cuestión.
- En pruebas cognitivas con ítems de construcción se pueden utilizar el método holístico o el analítico para asignar puntuaciones a los ítems (mediante una rúbrica).
- La puntuación total del test se define como la suma ponderada de las puntuaciones obtenidas por el sujeto en los ítems del test.
- Para dotar de significado a las puntuaciones de los test se puede optar por una interpretación normativa o criterial. En la normativa se compara la actuación del sujeto en el test con la actuación en esa prueba de un grupo de referencia del sujeto (grupo normativo), utilizando para ello puntuaciones normativas o tipificadas. En la criterial se compara la actuación del sujeto en el test con un criterio que puede ser un punto de corte en la escala —que sirve para adscribir al sujeto a una categoría u otra previamente establecida— o un dominio de conducta o de contenido perfectamente definido —que sirve para describir la actuación del sujeto en términos de lo que éste sabe o no sabe hacer—.

### ¿Qué debe realizar?

Se deberá elaborar el protocolo para la lectura de las respuestas de las personas que contestaron el instrumento de evaluación del PROEME en el Portafolio de evidencias 8, en la Ficha técnica 21. Protocolo de lectura (documento PE\_P8\_P0\_00.docx).

Para mantener organizados los documentos será necesario guardarlos considerando lo siguiente:

Sustituya P0\_000 del nombre original del Portafolio de Evidencias, PE\_P8\_P0\_000.docx, de la siguiente forma:

- P0, por el número de PROEME que le corresponda en la entidad.
- 000, con las tres letras que identifiquen el nombre de la entidad.  
Ejemplo: PE\_P8\_P3\_SON.docx (Portafolios de Evidencias del paso 8, del PROEME 3 de Sonora).

## 8.2. Definir un protocolo de análisis de la métrica del instrumento

Se debe formalizar el procedimiento que se realizará para valorar la pertinencia estadística de los reactivos, tareas evaluativas y del instrumento integrado (opción múltiple o respuesta construida). Esta información debe ser congruente con el modelo de medición establecido en la planeación general del instrumento al inicio de su construcción. Además, se deben incorporar los valores deseados de cada parámetro e indicador, así como su justificación (DOF, 2017, 28 de abril, p. 21).

## Protocolo para el análisis de la métrica del instrumento

El trabajo realizado para el PROEME ha pasado por diferentes planos, como lo mencionan Navas Ara y sus colaboradores (2010); se ha pasado del plano teórico, en el que se encuentran las características o constructos no directamente observables que se desea medir, al plano empírico, en el que se encuentran los indicadores —directamente observables— que representan empíricamente esas características que se busca medir, y de éste, al plano matemático, en el que los indicadores son representados numéricamente a través de las puntuaciones de los test.<sup>5</sup>

Las técnicas empleadas para el análisis de un instrumento dependen de su naturaleza, de los objetivos específicos para los cuales se diseñó, así como del tamaño de la población evaluada. Sin embargo, en todos los casos debe aportarse información sobre la dificultad y la discriminación de sus reactivos o tareas evaluativas, así como de la precisión del instrumento y los indicadores de consistencia interna o estabilidad del instrumento, los cuales, además de los elementos asociados a la conceptualización del objeto de medida, forman parte de las evidencias que servirán para valorar la validez de la interpretación de sus resultados. Estos elementos deberán reportarse en el informe o manual técnico del instrumento (DOF, 2017, 12 de abril).

Las principales teorías para el análisis de los instrumentos son la Teoría Clásica de los Test (TCT) o modelo lineal de Spearman y la Teoría Referida al Ítem (TRI). En la siguiente tabla se destacan las principales características de ambas teorías:

**Tabla 2.1. Comparativo entre TCT y TRI**

Aspectos	Teoría Clásica de los Test (TCT)	Teoría de Respuesta a los Ítems (TRI)
Modelo	Lineal	No lineal
Asunciones	Débiles (fáciles de cumplir por los datos)	Fuertes (difíciles de cumplir por los datos)
Invarianza de las mediciones	No	Sí
Invarianza de las propiedades del test	No	Sí
Escala de las puntuaciones	Entre 0 y la puntuación máxima en el test	Entre $-\infty$ y $+\infty$
Énfasis	Test	Ítem
Relación Ítem-Test	Sin especificar	Curva Característica del Ítem
Descripción de los ítems	Índice de Dificultad y de Discriminación	Parámetros a, b, c
Errores de medida	Error típico de medida común para toda la muestra	Función de Información (varía según el nivel de aptitud)
Tamaño muestral	Puede funcionar bien con muestras de entre 200 y 500 sujetos aproximadamente	Se recomienda más de 500 sujetos, aunque depende del modelo

Fuente: Muñiz, 2010.

<sup>5</sup> Para realizar este trabajo, se debe contar con personal que tenga experiencia en el análisis psicométrico de instrumentos de evaluación.

Podrá consultar una introducción sencilla y clara de las teorías de los test TCT y TRI en la siguiente liga: <http://www.papelesdelpsicologo.es/pdf/1796.pdf>

El protocolo deberá responder a las preguntas:

- ¿Qué teoría para el análisis de los reactivos utilizará?
- ¿De qué forma se valorará la pertinencia estadística de los reactivos?
- ¿Cuáles son los valores deseados de cada parámetro e indicador?

### ¿Qué debe realizar?

Se deberá elaborar el protocolo de análisis del instrumento de evaluación del PROEME en el Portafolio de evidencias 8, en la Ficha técnica 22. Protocolo para el análisis de la métrica (documento PE\_P8\_P0\_00.docx).

## 8.3. Realizar los análisis de los instrumentos con base en el protocolo establecido

- a) Estimación de estadísticos de reactivos o tareas;
  - b) confiabilidad/función de información<sup>6</sup>/índice de generalizabilidad;
  - c) dimensionalidad, y
  - d) funcionamiento diferencial (reactivos e instrumento).
- Para el caso de cuestionarios, verificar que la medición no esté influenciada por la deseabilidad social. En este sentido, una de las técnicas o procedimientos que se emplean con más frecuencia es utilizar una escala explícita de deseabilidad social en conjunto con la escala de interés y verificar si existe una asociación significativa y moderada entre ellas; de ser así, se asume la presencia de deseabilidad social (Havercamp y Reiss, 2003).<sup>7</sup>
  - En cuanto a los cuestionarios, elegir y justificar el método para la corrección de los puntajes, en caso de detectar la presencia de algún estilo de respuesta, con el propósito de reducir el valor de la invarianza en la medida.

Como se precisa en la tabla comparativa de la TCT y la TRI, la descripción de los ítems se realiza con procedimientos diferentes; a continuación se presentan ambos caminos.

<sup>6</sup> Concepto en la TRI que indica la precisión con la que el reactivo estima el grado de constructo latente que posee el objeto evaluado a lo largo de la escala. En el caso de pruebas, un reactivo o tarea evaluativa proporcionará mayor información sobre los sustentantes cuyo nivel de habilidad esté cerca de su dificultad que de aquellos con una habilidad alejada de este punto.

<sup>7</sup> Estadísticamente, existen otros métodos para detectar la deseabilidad social y controlarla, como el modelo de observaciones espurias, el modelo de supresión o modelo moderador (Ganster, Hennessey y Luthans, 1983), el análisis de factores confirmatorio, el análisis confirmatorio de clases latentes, el método del punto medio (*midpoint responding*), las respuestas extremas (*extreme responding*), entre otros (Van Vaerenbergh y Thomas, 2012).

## Cálculo del Índice de Dificultad y de Discriminación<sup>8</sup> (TCT)

Se divide a los examinados que contestaron la prueba en tres grupos según sus calificaciones obtenidas en la prueba como un todo:<sup>9</sup>

- Grupo Superior, con las calificaciones más altas, formado por 27%.
- Grupo Inferior, con las calificaciones más bajas, compuesto por 27%.
- Grupo Intermedio, el restante, compuesto por 46%.

Los índices estadísticos de Dificultad y Discriminación se calculan a partir de los resultados de los grupos inferior y superior aplicando las siguientes fórmulas:

$$p = \frac{U_p + L_p}{U + L}$$

$$D = \frac{U_p - L_p}{U}$$

$U_p$  y  $L_p$  constituyen la cantidad de individuos que hay en los grupos superior e inferior, respectivamente, y que aciertan en el reactivo;  $U$  y  $L$  son el número total de personas en los grupos superior e inferior (se puede observar que  $U = L$ ), respectivamente. Al valor  $p$  se le conoce como índice de dificultad del reactivo, y a  $D$  como índice de discriminación del reactivo.

El índice de dificultad del reactivo tiene un rango de 0.00 a 1.00. Un reactivo con  $p = 0.00$  es uno que nadie contestó correctamente, y un reactivo con  $p = 1.00$  es el que todos respondieron en forma acertada. El valor  $p$  óptimo para un reactivo depende de varios factores, incluyendo los objetivos de la prueba y la cantidad de opciones de respuesta. En una prueba elaborada para medir un rango amplio de habilidad, el valor de  $p$  óptimo es más cercano a 0.50. El valor óptimo de  $p$  también varía de acuerdo con el número de opciones de respuesta.

**Tabla 2.2. Índices medios óptimos de la dificultad de reactivos para pruebas con reactivos de opción múltiple**

Número de opciones ( $k$ )	Índice medio óptimo de dificultad ( $p$ )
2	0.85
3	0.77
4	0.74
5	0.69
Abierta (ensayo, respuesta breve)	0.50

Fuente: elaborado por Navas Ara y sus colaboradores (2010, p. 66) a partir de datos proporcionados por F. M. Lord, *Psychometrika*, 17(1952), pp. 181-194.

<sup>8</sup> Estos índices aplican a reactivos tanto con referencia a normas como con referencia a criterios (Navas Ara *et al.*, 2010, p. 82).

<sup>9</sup> Cuando el grupo que presentó el instrumento de evaluación es pequeño, 50%, correspondiente a los grupos inferior y superior, a menudo se utiliza para propósitos de análisis de reactivos.

El índice de discriminación del reactivo ( $D$ ) es una medida de eficacia de un reactivo para discriminar entre quienes obtienen altas y bajas calificaciones en una prueba. Mientras más elevado sea el valor de  $D$ , resulta más eficaz para establecer dicha distinción. Cuando  $D$  es igual a 1.00, todos los examinados del grupo superior y ninguno del grupo inferior en las calificaciones totales de la prueba respondieron el reactivo en forma adecuada. Sin embargo, casi nunca resulta  $D$  igual a 1.00 y, por lo regular, se considera aceptable un reactivo si tiene un índice  $D$  de 0.30 o mayor.

Pero  $D$  y  $p$  no son índices independientes, y el valor de  $D$  mínimo aceptable varía de acuerdo con el valor de  $p$ . Un valor de  $D$  en cierta medida inferior a 0.30 es aceptable mientras  $p$  cada vez aumenta o disminuye más que el valor óptimo, sobre todo cuando los grupos de comparación superior e inferior son numerosos. Asimismo, un reactivo con el índice  $D$  bajo no se descarta automáticamente: es posible salvarlo modificándolo.

### Cálculo de los parámetros $a$ , $b$ y $c$ (TRI)

La TRI se basa en la relación funcional teórica entre un continuo de capacidad latente supuesto y las respuestas a reactivos individuales en una prueba. Los cálculos que conlleva son muy complicados y en general deben realizarse con la ayuda de un programa de cómputo como LOGIST, BILOG, ASCAL o BIGSTEPS (Navas *et al.*, 2010, p. 71).

El modelo usual de la TRI es una función logística que tiene uno, dos o tres parámetros: La fórmula para el modelo de tres parámetros es:

$$P(\theta) = c + (1 - c) \frac{1}{1 + e^{-a(\theta - b)}}$$

En esta fórmula:

- $e$  es la base del logaritmo natural (2.718282),
- $a$  es un parámetro de la pendiente de un reactivo,
- $b$  es un parámetro de ubicación de un reactivo,
- $c$  es un parámetro pseudoadivinatorio,
- $\theta$  es el nivel de capacidad del examinado en una escala de calificación estándar, y
- $P(\theta)$  es la probabilidad de que una persona con nivel de capacidad conteste el reactivo correctamente.

La fórmula para el modelo de dos parámetros es:

$$P(\theta) = \frac{1}{1 + e^{-a(\theta - b)}}$$

Esta fórmula supone que  $c = 0$  (pseudoadivinatorio es igual a 0), donde:

- $e$  es la base del logaritmo natural (2.718282),
- $a$  es un parámetro de la pendiente de un reactivo,
- $b$  es un parámetro de ubicación de un reactivo,

- $\theta$  es el nivel de capacidad del examinado en una escala de calificación estándar, y
- $P(\theta)$  es la probabilidad de que una persona con nivel de capacidad conteste el reactivo correctamente.

La fórmula para el modelo de un parámetro, o Modelo de Rasch, es:

$$P(\theta) = \frac{1}{1 + e^{-1(\theta-b)}}$$

Esta fórmula supone que todos los reactivos son igualmente discriminantes:  $a = 0$  (discriminación es igual a 0).

- $e$  es la base del logaritmo natural (2.718282),
- $b$  es un parámetro de ubicación de un reactivo,
- $\theta$  es el nivel de capacidad del examinado en una escala de calificación estándar, y
- $P(\theta)$  es la probabilidad de que una persona con nivel de capacidad conteste el reactivo correctamente.

#### ¿Qué debe realizar?

Se deberá aplicar el protocolo de análisis del instrumento de evaluación del PROEME especificado en la Ficha técnica 22 y registrar los resultados en el Portafolio de evidencias 8, en la Ficha técnica 23. Resultados del análisis (documento PE\_P8\_P0\_00.docx).

## 8.4. Obtener el índice de respuestas omitidas

Independientemente del marco de la métrica, es recomendable calcular el índice de no respuesta para el reactivo o la tarea evaluativa, es decir, la proporción de las personas que no lo contestaron. Un análisis exhaustivo de esta información ayuda a decidir qué tratamiento dar a los reactivos sin respuesta en la puntuación del instrumento y reconsiderar el tiempo de administración.

#### ¿Qué debe realizar?

- Establecer un código de respuestas omitidas para la lectura de las respuestas de los participantes en los instrumentos de evaluación y especificarlo en la Ficha técnica 21. Protocolo de lectura.
- Se deberá registrar en el Portafolio de Evidencias 8 la forma en que se obtendrá el índice de respuestas omitidas en el protocolo de análisis del instrumento de evaluación del PROEME, especificado en la Ficha técnica 22, y también los resultados, en la Ficha técnica 23. Resultados del análisis (documento PE\_P8\_P0\_00.docx).

## 8.5. Verificar el ajuste de los datos, en el caso de la TRI

Es importante incluir la valoración del ajuste del comportamiento de los reactivos al modelo de medición establecido. Se recomienda llevar a cabo contrastes para evaluar la bondad de ajuste de cada reactivo.<sup>10</sup> Es decir, es recomendable no sólo considerar la estimación puntual de cada índice, sino también su precisión, particularmente en caso de que el tamaño de la muestra sea pequeño.

### ¿Qué debe realizar?

- Determinar la forma en que se verificará el ajuste de los datos (en el caso de TRI).
- Se deberá registrar, en el Portafolio de Evidencias 8, la forma en que se verificará el ajuste de datos, en el protocolo de análisis del instrumento de evaluación del PROEME, especificado en la Ficha técnica 22, y registrar los resultados, en la Ficha técnica 23. Resultados del análisis (documento PE\_P8\_P0\_00.docx).

## 8.6. Estimar el error de medida

Cualquier instrumento es vulnerable a todas las fuentes de error que a veces causan fluctuaciones en los resultados de una fuente de información a otra o de un área de medición a otra. Debido a que hay diversas fuentes de error, hay diferentes procedimientos para estimar la confiabilidad, por ejemplo: la consistencia interna, la estabilidad temporal y la concordancia entre los evaluadores, por lo que se debe elegir el más apropiado al tipo de instrumento que se construyó. La forma para obtener las evidencias de confiabilidad —índice de generalizabilidad, función de información, error estándar condicional, índice de consistencia, entre otros— debe ser apropiada para los usos previstos de los resultados, la población involucrada y los modelos de medición utilizados para obtener las puntuaciones. Con el fin de conocer qué tan precisa es la medición y cuáles son sus principales fuentes de varianza se emplean la confiabilidad en el caso de la TCT, la función de información en el caso de la TRI, o el índice de generalizabilidad para el caso de la Teoría de la Generalizabilidad.<sup>11</sup>

Un mayor grado de precisión es requerido para evaluaciones de alto impacto. A la inversa, un grado más bajo puede ser aceptable cuando una decisión con base en la puntuación de un instrumento es reversible o depende de su corroboración de otras fuentes de información.

<sup>10</sup> Al interpretar los índices de los reactivos se debe tomar en cuenta que son estimaciones de parámetros poblacionales propensas a errores por fluctuaciones muestrales.

<sup>11</sup> La confiabilidad informa cuánto influyen en la estimación los errores aleatorios, no sistemáticos; una confiabilidad baja refleja mucho error de medición y, por lo tanto, las inferencias que se hacen a partir de la estimación son menos precisas y, por ende, tendrán menos validez. Del mismo modo, la función de información indica la precisión de la estimación a lo largo de la escala de habilidad, es decir, para niveles de habilidad distintos, la precisión con la que se midan también será diferente.

### ¿Qué debe realizar?

- Determinar la forma en que calculará el error de medida, según la teoría seleccionada.
- Se deberá registrar, en el Portafolio de Evidencias 8, la forma en que se calculará el error de medida en el protocolo de análisis del instrumento de evaluación del PROEME, especificado en la Ficha técnica 22, y registrar los resultados en la Ficha técnica 23. Resultados del análisis (documento PE\_P8\_P0\_00.docx).

## 8.7. Realizar análisis de funcionamiento diferencial de los reactivos y del instrumento

Los análisis DIF (Funcionamiento Diferencial del Reactivo) y DFT (Funcionamiento Diferencial del Instrumento) deben hacerse considerando grupos de interés donde se sospeche que los reactivos pueden funcionar de manera diferente; si esto llegara a ocurrir, se recomienda que el reactivo no sea considerado para emitir los resultados; algunas variables que pueden considerarse para estos análisis son: género, creencia religiosa, lengua materna, estrato socioeconómico y grupo étnico. El aspecto más relevante de este tipo de análisis es fundamentar que la prueba, en su conjunto, no afecta diferencialmente a los sujetos debido a su grupo de pertenencia (Muraki, 1999; Fidalgo y Madeira, 2008).

- Para los cuestionarios se debe identificar si existen diferencias transculturales en el uso de las escalas Likert o en reactivos categóricos individuales extraídos de esas escalas. Si se encuentra que alguna subpoblación muestra un estilo de respuesta específico para responder debido a sus características culturales, se debe reportar y considerar en el momento de comunicar las inferencias de los resultados. Los posibles estilos o patrones de respuesta son: sesgo de positividad —tendencia a estar de acuerdo con los reactivos independientemente de la actitud real—, estilo discrepante —tendencia a no estar de acuerdo con los elementos, independientemente de su contenido—, estilo de respuesta extrema —elección de los puntos más lejanos de la escala— y, por último, estilo de respuesta no contenciosa —selección aleatoria de las respuestas de los reactivos— (Buckley, 2009).

### ¿Qué debe realizar?

- Si se sospecha que los reactivos pueden funcionar de forma diferente para ciertos grupos de interés, se debe especificar la forma en que se realizarán los análisis DIF y DFT (en caso de cuestionarios, describir la forma en que se identificará el estilo de respuesta), y registrarlo en el Portafolios 8, en la Ficha técnica 22.
- Registrar en la Ficha técnica 23 los resultados de los análisis realizados.

## 8.8. Dar a conocer la precisión de la medición con oportunidad

Los resultados de los análisis para determinar el error de medida o la consistencia de la medición se deben reportar para que se les considere en el momento de la interpretación de los resultados, y ésta sea adecuada (JCSEE, 2011 y AERA, APA y NCME, 2014).

## 8.9. Dictaminar la pertinencia de los reactivos y las tareas evaluativas con base en sus parámetros

Con base en el modelo psicométrico seleccionado y los estándares establecidos en el protocolo, se deben revisar los resultados del análisis cuantitativo para determinar cuáles reactivos y tareas evaluativas son métricamente adecuados; este ejercicio consiste en comparar los índices obtenidos con los valores deseados, lo que debe dar como resultado una etiqueta que resuma su funcionamiento.

### ¿Qué debe realizar?

- Establecer una tabla con etiquetas y su significado para la clasificación de los reactivos o tareas evaluativas.
- Evaluar los reactivos o tareas evaluativas aplicando las etiquetas correspondientes.

## 8.10. Integrar la información cuantitativa al banco de reactivos y tareas

Los parámetros o indicadores obtenidos en este paso deben incorporarse de manera periódica en el repositorio destinado para el resguardo de reactivos, de modo que esta información pueda consultarse en futuros ensambles y en el mantenimiento del instrumento.

### ¿Qué debe realizar?

Incorporar en el repositorio las etiquetas correspondientes a cada reactivo o tarea evaluativa.

## 8.11. Determinar la inclusión en los análisis estadísticos de las respuestas de personas que contestaron un instrumento adaptado o de aquellas a las que se hizo alguna modificación en las condiciones de administración

Según el impacto de la evaluación y sus características, se debe hacer un ejercicio de reflexión para determinar, a la luz del principio de equidad (UNICEF, 2012), si las respuestas de estas personas son incluidas en los análisis con el resto de los resultados. El producto de esta reflexión debe hacerse explícito en el protocolo correspondiente. Con el propósito de facilitar esta decisión se propone apoyarse en la siguiente clasificación (Lewis, Patz, Sheinker y Barton, 2002):

- *Categoría 1.* Se definen como adaptaciones y modificaciones que no se espera que influyan en el desempeño de la persona que contesta el instrumento de una manera que altere la interpretación de los resultados de la evaluación. Se caracterizan por ser ajustes que no alteran la medición ni sus condiciones de manera directa. Por ejemplo: magnificar el tamaño de letra del instrumento o el sonido en los estímulos

auditivos; destinar a las personas un espacio exclusivo para la presentación del instrumento o tomar descansos adicionales sin que esto signifique aumentar el tiempo de administración.

- **Categoría 2.** Se definen como adaptaciones y modificaciones que pueden tener un efecto en el desempeño de las personas y que deben ser consideradas al interpretar su puntuación. Es posible que mediante estudios cuantitativos se muestre que este tipo de ajustes no se comporta estadísticamente diferente, en comparación con el resto. En caso contrario, los resultados del análisis deberán interpretarse con esa salvedad. Un ejemplo de esta categoría es aumentar el tiempo de administración, leer en voz alta, resaltar con negritas las instrucciones, entre otras.
- **Categoría 3.** Se definen como adaptaciones y modificaciones que de alguna manera cambian el objeto de medida y pueden alterar la interpretación de los puntajes obtenidos. Esto ocurre cuando el ajuste está fuertemente relacionado con el constructo o la variable latente que se está midiendo (por ejemplo: tener una prueba de comprensión de lectura leída en voz alta). A falta de una investigación que muestre lo contrario, los resultados de los instrumentos y las consecuencias o decisiones asociadas con ellos deben interpretarse a la luz no sólo de los ajustes realizados, sino también de cómo éstos pueden alterar lo que se mide. Ejemplos de esto son la traducción de un instrumento a otro idioma o a Braille, el uso exclusivo de un diccionario o el empleo de un lector o un escriba.

## 8.12. Realizar los análisis estadísticos a las formas adaptadas

Se debe determinar el modelo de medición a partir del cual se realizarán los análisis de este tipo de formas. Este criterio incluye la equiparación de puntajes para que sean comparables los resultados de todas las formas, e identificables los aspectos o componentes del instrumento que pudieran evaluarse equivocadamente para alguna población (Kolen y Brennan, 2014). Si las puntuaciones totales se presentan por separado para los grupos evaluados, se debe valorar la comparabilidad de las puntuaciones (adaptaciones contra población total). Si esta evidencia indica que existen diferencias entre los grupos, es necesario examinar la validez de las interpretaciones de los resultados y proporcionar declaraciones preventivas.

*Proceso para la equiparación de instrumentos de evaluación* (DOF, 2017, 12 de abril, p. 13). El procedimiento que permite hacer equivalentes los resultados obtenidos en diferentes formas o versiones de un mismo instrumento es una equiparación. La que aquí se plantea considera dos estrategias:

- a) si el número de sustentantes es de al menos 100 en ambas formas, se utilizará el método de equiparación lineal de Levine para puntajes observados;
- b) si el número de sustentantes es menor de 100 en alguna de las formas, se utilizará el método de equiparación de identidad (*identity equating*).

## Método de equiparación lineal de Levine

La equiparación de las formas de un instrumento deberá realizarse utilizando el método de equiparación lineal de Levine (Kolen y Brennan, 2014), para puntajes observados bajo un diseño de grupos no equivalentes con reactivos comunes. Dicho diseño es uno de los más utilizados en la práctica. A cada muestra de sujetos se le suministra solamente una forma de la prueba, con la peculiaridad de que ambas muestras comparten un conjunto de reactivos en común llamado *ancla* que permite establecer la equivalencia entre las formas a equiparar.

Cualquiera de los métodos de equiparación de puntajes que se construya involucra dos poblaciones diferentes. Sin embargo, una función de equiparación de puntajes se define sobre una población única. Por lo tanto, las poblaciones 1 y 2, que corresponden a aquellas donde se aplicaron la forma nueva y la antigua, deben combinarse para obtener una población única a fin de definir una relación de equiparación.

Esta única población se conoce como población sintética, y en ella se les asignan pesos  $w_1$  y  $w_2$  a las poblaciones 1 y 2, respectivamente, esto es,  $w_1 + w_2 = 1$  y  $w_1, w_2 \geq 0$ . Para este proceso se utilizará:

$$w_1 = \frac{N_1}{N_1 + N_2}$$

$$w_2 = \frac{N_2}{N_1 + N_2}$$

Donde corresponde al tamaño de la población 1, y corresponde al tamaño de la población 2.

Los puntajes de la forma nueva, aplicada a la población 1, serán detonados por  $X$ ; los puntajes de la forma antigua, aplicada a la población 2, serán detonados por  $Y$ .

Los puntajes comunes están identificados por  $V$ , y se dice que los reactivos comunes corresponden a un anclaje interno cuando  $V$  se utiliza para calcular los puntajes totales de ambas poblaciones.

Usando el concepto de población sintética, la relación lineal de equiparación de puntajes para el diseño de grupos no equivalentes con reactivos comunes se escribe de la siguiente forma:

$$I_{y_s}(x) = \frac{\sigma_s(Y)}{\sigma_s(X)} [x - \mu_s(X)] + \mu_s(Y)$$

Donde  $s$  denota la población sintética y

$$\mu_s(X) = \mu_1(X) - w_2\gamma_1[\mu_1(V) - \mu_2(V)]$$

$$\mu_s(Y) = \mu_2(Y) - w_1\gamma_2[\mu_1(V) - \mu_2(V)]$$

$$\sigma_s^2(X) = \sigma_1^2(X) - w_2\gamma_1^2[\sigma_1^2(V) - \sigma_2^2(V)] + w_1w_2\gamma_1^2[\mu_1(V) - \mu_2(V)]^2$$

$$\sigma_s^2(Y) = \sigma_2^2(Y) - w_2\gamma_2^2[\sigma_1^2(V) - \sigma_2^2(V)] + w_1w_2\gamma_2^2[\mu_1(V) - \mu_2(V)]^2$$

Donde los subíndices 1 y 2 se refieren a las poblaciones 1 y 2, respectivamente.

$$\gamma_1 = \frac{\sigma_1(X,V)}{\sigma_1^2(V)} \quad \text{y} \quad \gamma_2 = \frac{\sigma_2(X,V)}{\sigma_2^2(V)}$$

Específicamente, para el método de Levine para puntajes observados bajo un diseño de grupos no equivalentes con reactivos comunes, las  $\gamma$  se representan de la siguiente manera:

$$\gamma_1 = \frac{\sigma_1^2(X)}{\sigma_1(X,V)}$$

$$\gamma_2 = \frac{\sigma_2^2(Y)}{\sigma_2(Y,V)}$$

Para aplicar este método basta con remplazar estos coeficientes en las ecuaciones lineales antes descritas. Por su parte, Kolen y Brennan proveen justificaciones para usar esta aproximación.

Es importante señalar que los puntajes a los que se les aplique la equiparación como pendiente y como ordenada al origen, el procedimiento es análogo al descrito en la sección "Procedimiento para el error estándar condicional. Método delta"; y el error estándar condicional de medición para la transformación, que ya incorpora la confiabilidad, está dado por:

$$\sigma(P_{ie}) = \frac{A}{2} \left[ \frac{b_1}{2(k+1) \left( \sqrt{\frac{x_e}{k+1}} \right) \left( \sqrt{1 - \frac{x_e}{k+1}} \right)} + \frac{b_1}{2(k+1) \left( \sqrt{\frac{x_e+1}{k+1}} \right) \left( \sqrt{1 - \frac{x_e+1}{k+1}} \right)} \right] \sigma(x_e)$$

Donde  $X_e$  son las puntuaciones equiparadas, son una transformación de las puntuaciones crudas, por lo que el error estándar de medida de dicha transformación se define como:

$$\sigma(x_e) = b_1 * \sigma(x)$$

### Método de equiparación de identidad (*identity equating*)

La equiparación de identidad es la más simple, toda vez que no hace ningún ajuste a la puntuación  $x$  en la escala de la forma  $X$  en el momento de convertirla en la puntuación equiparada  $y$  en la escala de la forma  $Y$ .

Es decir, dichas puntuaciones se consideran equiparadas cuando tienen el mismo valor, por lo que las coordenadas de la línea de equiparación de identidad están definidas simplemente como  $x = y$  (Holland y Stawderman, 2011).

## ¿Qué debe realizar?

- Elegir el método de equiparación más pertinente al tipo de evaluación realizada.
- Realizar el cálculo correspondiente
- Describir la forma en que se realizará la equiparación (en caso de haber realizado adaptaciones del instrumento de evaluación) y registrarla en el Portafolio 8, en la Ficha técnica 22.
- Registrar los resultados de equiparación en el Portafolio 8, en la Ficha técnica 23.

### 8.13. Documentar las características de las traducciones (si corresponde)

Si se hace una traducción del instrumento a otro idioma o lengua, se debe describir el proceso y evaluar los resultados y su comparabilidad con la versión original.

#### Evidencias del paso 8. Evaluación de la métrica del instrumento<sup>12</sup>

Para la elaboración de las evidencias se ha preparado el Portafolio de Evidencias 8, correspondiente al paso 8, Evaluación de la métrica del instrumento, en el que se incluyen las fichas técnicas y los espacios para la incorporación de las evidencias con formato libre. Las evidencias a realizar en este paso se listan a continuación:

- *Protocolo de lectura de las hojas de respuestas y corrección de las bases de datos.* Describe los mecanismos para garantizar una base de datos libre de errores de codificación. Utilice la Ficha técnica 21. Protocolo de lectura, en el Portafolio de evidencias 8 (documento PE\_P8\_PO\_000.docx).
- *Protocolo del análisis de la métrica utilizada.* Establece los métodos para estimar las propiedades cuantitativas de los reactivos, de las tareas y del instrumento como una unidad; debe ser congruente con el alcance de la evaluación y con el modelo estadístico seleccionado. Utilice la Ficha técnica 22. Protocolo para el análisis de la métrica, en el Portafolio de evidencias 8 (documento PE\_P8\_PO\_000.docx)
- Resultados de los análisis realizados (estimación de parámetros; confiabilidad/función de información/índice de generalizabilidad; dimensionalidad; funcionamiento diferencial-reactivos e instrumento). Utilice la Ficha técnica 23. Resultados del análisis, en el Portafolio de evidencias 8 (documento PE\_P8\_PO\_000.docx).
- *Reactivos y tareas dictaminados.* El banco donde se resguardan los reactivos y las tareas del instrumento debe almacenar los estadísticos, indicadores y etiquetas que dan cuenta de su funcionamiento. Utilice la Ficha técnica 22. Protocolo para el análisis de la métrica, y la Ficha técnica 23. Resultados del análisis, en el Portafolio de evidencias 8 (documento PE\_P8\_PO\_000.docx).

<sup>12</sup> Es importante mencionar que, de existir análisis psicométricos personales, o si alguna de las pruebas es personalizada y no anónima, debe considerarse esta información como sensible y apegada al tratamiento establecido sobre la información y datos personales en la Ley Federal de Transparencia y Acceso a la Información Pública. En este caso, es necesario aplicar medidas de seguridad en el manejo de los archivos, que pueden ser desde una encriptación en documento PDF hasta el desarrollo de sitios seguros (https) para compartir la información.

- Si se realizaron adaptaciones al instrumento, reportar los resultados obtenidos. Utilice la Ficha técnica 22. Protocolo para el análisis de la métrica, y la Ficha técnica 23. Resultados del análisis, en el Portafolio de evidencias (documento PE\_P8\_PO\_000.docx).
- Evidencias adicionales en el caso de cuestionarios o instrumentos con escalas. Utilice la Ficha técnica 22. Protocolo para el análisis de la métrica, y la Ficha técnica 23. Resultados del análisis, en el Portafolio de evidencias 8 (documento PE\_P8\_PO\_000.docx).
- Protocolo de los análisis realizados para controlar la deseabilidad y su interpretación.
- Protocolo para la identificación de patrones o estilos de respuesta y su tratamiento para aminorar el efecto en los resultados de la evaluación.

### Lista de cotejo del paso 8. Evaluación de la métrica del instrumento

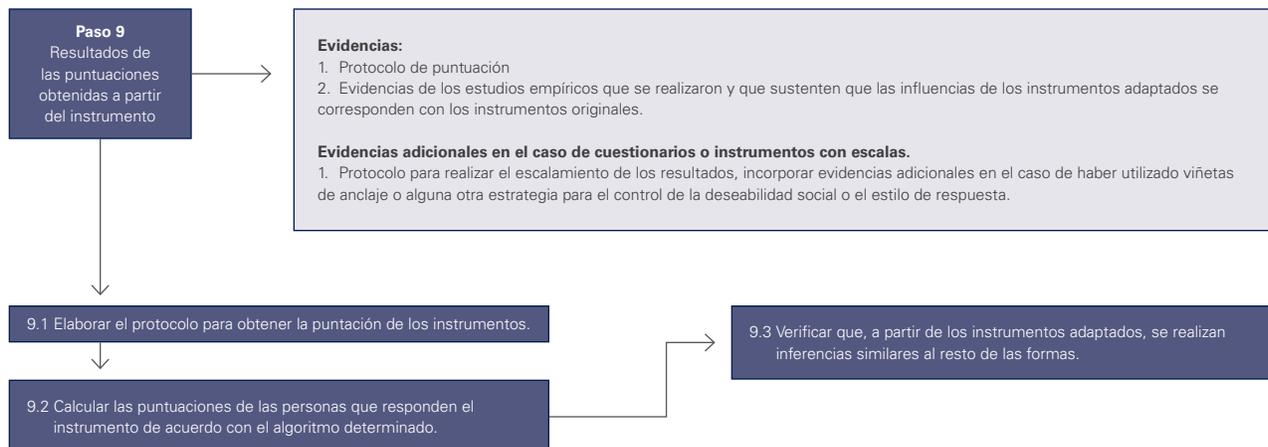
Una vez que han concluido las 13 actividades del paso 8, aplique la lista de cotejo 8, Evaluación de la métrica del instrumento, que se especifica en el anexo 1. Si cumple con todos los criterios de calidad, puede realizar el paso 9; si no cumple con todos los aspectos, justifique la razón en la lista de cotejo.

## Paso 9. Resultados de las puntuaciones obtenidas a partir del instrumento

Este paso se refiere a la asignación de un valor a las respuestas de las personas en cada reactivo o tarea evaluativa del instrumento, a fin de que al combinarlos se obtenga una o más puntuaciones, con base en el modelo establecido en la fase de planeación (AERA, APA y NCME, 2014).

En la figura 2.3 se muestran las actividades y evidencias a obtener en el paso 9.

**Figura 2.3. Paso 9. Resultados de las Puntuaciones obtenidas a partir del instrumento**



Fuente: elaboración propia a partir de los CTDUMIE (DOF, 2017, 28 de abril).

## 9.1. Elaborar el protocolo para obtener la puntuación de los instrumentos

La manera para derivar las puntuaciones de las personas que responden el instrumento debe ser descrita con suficiente detalle en un protocolo para garantizar su exactitud y pertinencia técnica. Se debe especificar cómo se trabajará el modelo de puntuación, por ejemplo, si se trata de una escala que se puntúa de manera dicotómica o politómica. Cuando se incluyan procesos de jueceo se deben documentar con claridad los procedimientos y criterios que se utilizarán para asignar la puntuación a las personas que contestaron el instrumento.

En el caso de pruebas, el protocolo para obtener la puntuación debe considerar:

- *Validar las claves de respuesta.* El primer paso para calificar un instrumento es verificar las claves de respuesta correcta en los reactivos de opción múltiple. En el caso de instrumentos de respuesta construida, se debe revisar que el protocolo para la asignación de las puntuaciones favorezca a que la rúbrica de calificación se emplee de manera adecuada, además de verificar que no haya errores en el registro. Este criterio es obligatorio y debe quedar constancia de que se implementó.
- Obtener, en el caso de los instrumentos de respuesta construida en los que el desempeño es evaluado a través de una rúbrica, el porcentaje de acuerdos interjueces e intrajueces antes de dar la calificación definitiva a las personas que contestaron la prueba. Se deben tener previstas las acciones que se implementarán en caso de haber una discordancia mayor a la tolerada (considerando un criterio establecido previamente), por ejemplo: reforzar la capacitación de los jueces o prescindir de los servicios de otros cuando existan decisiones consistentemente discordantes.
- Establecer criterios para seleccionar a los evaluadores de los instrumentos que requieran la calificación de jueces. Para calificar las tareas evaluativas por medio de una rúbrica se deben considerar, al menos, dos evaluadores, y es indispensable establecer criterios de selección: el perfil académico y profesional, el dominio de la tarea y, en caso de que haya participado en el proyecto anteriormente, los porcentajes según inter o intrajuez (Jonsson y Svingby, 2007; Rezaei y Lovorn, 2010; Stemler y Tsai, 2008; Stellmack *et al.*, 2009).
- Se deben establecer los mecanismos de seguridad para que el personal que realice la calificación sólo tenga acceso a las respuestas de los sustentantes sin conocer ninguno de los datos de identificación; por ejemplo: nombre, Registro Federal de Contribuyentes (RFC) y Clave Única de Registro de Población (CURP). Dichos mecanismos asegurarán el anonimato de las respuestas de los sustentantes.
- Capacitar y monitorear a los evaluadores del instrumento de respuesta construida. Se debe capacitar a los evaluadores para que se familiaricen con la rúbrica y para unificar sus decisiones (Council of Europe, 2011). También se proporcionará evidencia del grado de acuerdo entre las puntuaciones dadas por los jueces, a fin de que la calificación obtenida sea un resultado imparcial y objetivo de la medición.
- En cuanto a los cuestionarios, el protocolo para obtener la puntuación debe considerar el tratamiento que se le dará, en caso de detectar algún estilo de respuesta (extrema, respuestas intermedias o sobrevaloración), con el fin de controlar la deseabilidad social y aumentar la validez de las inferencias que se realizan a partir de los resultados de la evaluación.

## ¿Qué debe realizar?

Elaborar el protocolo para obtener la puntuación de los instrumentos y registrarlo en la Ficha técnica 24. Protocolo de puntuación, del Portafolios de evidencias 9.

En el contenido deberá incluir:

- Forma que se utilizará para derivar puntuaciones.
- Describir el proceso de jueceo (en caso de requerirlo para asignar puntuaciones).

Para pruebas:

- La forma para validar las claves de respuesta (instrumentos de opción múltiple) o para que la rúbrica de calificaciones se use de manera adecuada y que no haya errores en el registro (para instrumentos de respuesta construida).
- Para evidencias de desempeño, determinar el porcentaje según inter e intrajueces antes de dar la calificación definitiva a los evaluados. Y considerar la estrategia en caso de discordancias.
- Establecer criterios para la selección de los evaluadores de los instrumentos que requieren la calificación de jueces.
- Establecer mecanismos de seguridad que consideren el acceso exclusivo a las respuestas de los sustentantes que evalúa y sin acceso a los datos personales.
- Establecer el mecanismo para capacitar y monitorear a los evaluadores del instrumento de respuesta construida.
- Para cuestionarios, describir las estrategias para aumentar la validez de las inferencias que se realizan (detección de estilo de respuesta y control de la deseabilidad social).

## 9.2. Calcular las puntuaciones de las personas que responden el instrumento de acuerdo con el algoritmo determinado

El cálculo de las puntuaciones debe hacerse de acuerdo con lo establecido en el diseño del instrumento. Los análisis estadísticos pueden motivar cambios en el esquema de puntuación utilizado; sin embargo, cualquier desvío del plan original debe documentarse y justificarse.

- En el caso de pruebas, escoger el método de equiparación más adecuado, si las condiciones del instrumento lo permiten.<sup>13</sup> Si debido a la exigencia de las reglas para equiparar no se pudo llevar a cabo dicho proceso, entonces se deben escalar los puntajes de las diferentes formas para garantizar equidad en la evaluación.<sup>14</sup>

<sup>13</sup> La equiparación se utiliza para ajustar las puntuaciones de las formas de un mismo instrumento, y permite que las puntuaciones de una forma a otra sean utilizadas de manera intercambiable. Este método ajusta, por dificultad, las distintas formas que fueron ensambladas bajo las mismas reglas estadísticas y de contenido (Holland y Strawderman, 2011, y Kolen y Brennan, 2014).

<sup>14</sup> El escalamiento se lleva a cabo a partir de las puntuaciones crudas (cantidad de aciertos) de los sustentantes, y se obtiene una métrica común para todos los instrumentos de evaluación (Shun-Wen, 2006, y Wilson, 2005).

- En el caso de pruebas, si el reporte de calificaciones incluye la clasificación de los evaluados en niveles de desempeño, se debe documentar el procedimiento empleado para el establecimiento de puntos de corte, así como el método seleccionado —Angoff, Beuk, Bookmark, Nedelski, Hofstee, grupos de contraste, etc. (INEE, 2016, y Downing 2004)— y su justificación. Los descriptores de los niveles de desempeño y los puntajes asociados deben estar fundamentados y conocer el error estándar de medida asociado a las puntuaciones, e incorporarlo en el reporte técnico del instrumento.
- En el caso de instrumentos de autoevaluación, si se incluyeron viñetas de anclaje o alguna otra estrategia para el control de la deseabilidad social o el estilo de respuesta, se debe realizar el escalamiento de los resultados a través del método seleccionado (paramétrico o no paramétrico) (King y Wang, 2006).

### ¿Qué debe realizar?

- Aplicar el protocolo para obtener la puntuación de los instrumentos y registrar los resultados en la Ficha técnica 25. Resultados de la evaluación, del Portafolios de evidencias 9.
- En caso de pruebas, utilice la Ficha técnica 26. Protocolo de escalamiento, y la Ficha técnica 27. Protocolo para establecer niveles de desempeño, del Portafolio de evidencias 9 (documento PE\_P9\_P0\_000.docx).
- En caso de autoevaluaciones, utilice la Ficha técnica 28. Protocolo de escalamiento de resultados, del Portafolios de evidencias 9 (documento PE\_P9\_P0\_000.docx).

### 9.3. Verificar que, a partir de los instrumentos adaptados, se realizan inferencias similares al resto de las formas

Las adaptaciones de un instrumento deben producir conclusiones equivalentes a las generadas con las formas no adaptadas, por lo que las inferencias que se pueden derivar de ellas se deben justificar con evidencias o estudios empíricos (ETS, 2015).

*Ejemplo de la Fase 4. Análisis de resultados del instrumento de evaluación.* En la siguiente liga podrá consultar los Criterios Técnicos y de procedimiento para el análisis de los instrumentos de evaluación, el proceso de calificación y la definición de las listas de prelación de los concursos de oposición para el ingreso al Servicio Profesional Docente en educación básica (EB) y educación media superior (EMS) para el ciclo escolar 2016-2017.

[http://www.dof.gob.mx/nota\\_detalle.php?codigo=5436465&fecha=10/05/2016](http://www.dof.gob.mx/nota_detalle.php?codigo=5436465&fecha=10/05/2016)

## Evidencias del paso 9. Resultados de las puntuaciones obtenidas a partir del instrumento<sup>15</sup>

Para la elaboración de las evidencias se ha preparado el Portafolio de Evidencias 9 correspondiente al paso 9, Resultados de las puntuaciones obtenidas a partir del instrumento, en el que se incluyen las fichas técnicas y los espacios para la incorporación de las evidencias con formato libre. Las evidencias a realizar en este paso se listan a continuación:

- *Protocolo de puntuación.* Contiene los algoritmos y las reglas de puntuación establecidos en la Ficha técnica del instrumento y la manera correcta de interpretar los resultados. Utilice la Ficha técnica 24. Protocolo de puntuación, y la Ficha técnica 25. Resultados de la evaluación, del Portafolio de evidencias 9 (documento PE\_P9\_P0\_000.docx).
- Evidencias de los estudios empíricos que se realizaron y que sustenten que las inferencias de los instrumentos adaptados se corresponden con los instrumentos originales.
- *Evidencias adicionales en el caso de cuestionarios o instrumentos con escalas.* Utilice la Ficha técnica 26. Protocolo de escalamiento, y la Ficha técnica 27. Protocolo para establecer niveles de desempeño, del Portafolio de evidencias 9 (documento PE\_P9\_P0\_000.docx).
- Protocolo para realizar el escalamiento de los resultados, incorporar evidencias adicionales en el caso de haber utilizado viñetas de anclaje o alguna otra estrategia para el control de la deseabilidad social o el estilo de respuesta. En caso de autoevaluaciones, utilice la Ficha técnica 28. Protocolo de escalamiento de resultados, del Portafolios de evidencias 9 (documento PE\_P9\_P0\_000.docx).

## Lista de cotejo del paso 9. Resultados de las puntuaciones obtenidas a partir del instrumento

Una vez que han concluido las tres actividades del paso 9, aplique la lista de cotejo 9, Resultados de las puntuaciones obtenidas a partir del instrumento que se especifica en el anexo 1. Si cumple con todos los criterios de calidad, ha concluido la Fase 4.

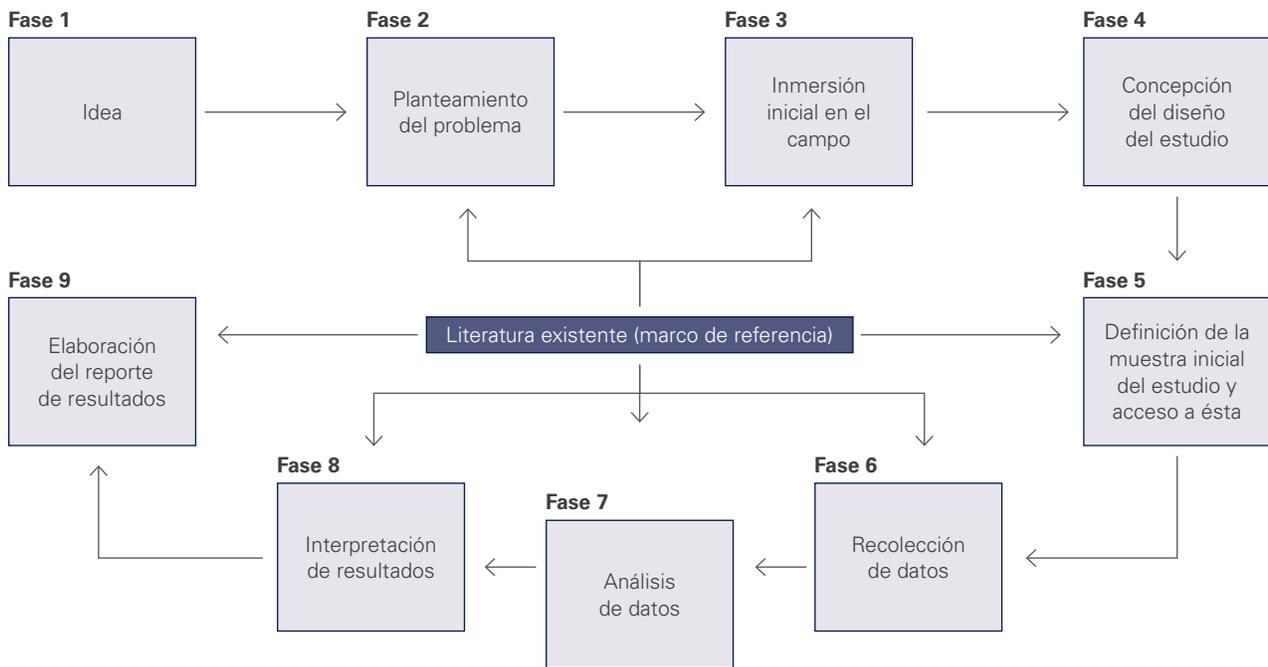
<sup>15</sup> Es importante mencionar que, de existir análisis psicométricos personales, o si alguna de las pruebas es personalizada y no anónima, debe considerarse esta información como sensible y apegada al tratamiento sobre la información y los datos personales en la Ley Federal de Transparencia y Acceso a la Información Pública. En este caso, es necesario aplicar medidas de seguridad en el manejo de los archivos, que pueden ser desde una encriptación en archivo PDF hasta el desarrollo de sitios seguros (https) para compartir la información.

### 3. Análisis de resultados en el enfoque cualitativo

De acuerdo con Hernández Sampieri, Fernández Collado y Baptista (2010, p. 8), el enfoque cualitativo es un proceso integrado por 9 fases, que se muestran en la figura 3.1.

Figura 3.1. Proceso cualitativo

32



Fuente: Hernández Sampieri, Fernández Collado y Baptista Lucio, 2010, p. 8.

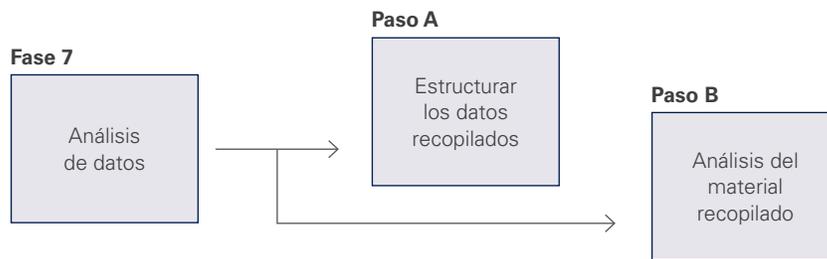
En esta guía se trabajará con las fases 7 y 8, que corresponden al análisis de los datos y a su interpretación.

En el enfoque cualitativo las fases de recolección, análisis e interpretación de datos se realizan de forma simultánea en algunos casos, de tal manera que, al tener una inmersión profunda dentro del campo de la evaluación, se podrá tener información que posiblemente requiera mayor precisión, por lo que el evaluador podrá regresar al campo

y recopilar más datos; incluso pueden surgir nuevas preguntas que van reorientando la evaluación. Las observaciones se van enfocando en responder al planteamiento del problema, que en algunas ocasiones se replantea y complementa con los nuevos datos que van surgiendo.

## Fase 7. Análisis de datos

Figura 3.2. Fase 7. Análisis de datos



Fuente: Hernández Sampieri, Fernández Collado y Baptista Lucio, 2010.

## Paso A. Estructurar los datos recopilados

### Organización de datos

En el enfoque cualitativo se ha obtenido la información mediante diversas fuentes: observación del ambiente, bitácoras y recolección enfocada, que pudo realizarse por medio de revisión de documentos, entrevistas, historias de vida, descripciones, conversatorios, etc., dependiendo de la evaluación que se realiza y de las características de los sujetos que participan. Por supuesto, aunque en el enfoque se tiene flexibilidad para incrementar o disminuir el número de entrevistados o descripciones, o la cantidad de grupos de enfoque a realizar, el trabajo debe estar orientado por el objeto de la evaluación.

Generalmente se obtiene una gran cantidad de datos que han de ser organizados, por lo que se debe elegir la forma en que se organizará la información. Para empezar, se deberá revisar nuevamente el material y crear una bitácora de análisis<sup>16</sup> cuya función es documentar paso a paso el proceso analítico. Durante esta revisión se debe verificar la calidad de la información (pertinencia, relevancia y objetividad).

<sup>16</sup> Tiene la función de documentar el procesamiento de análisis y las propias reacciones del investigador al proceso, y contiene fundamentalmente: anotaciones sobre el método utilizado; las ideas, conceptos, significados, categorías e hipótesis que van surgiendo del análisis; anotaciones en relación con la credibilidad, y verificación del estudio para que otro investigador pueda evaluar su trabajo.

## Transcripción de materiales

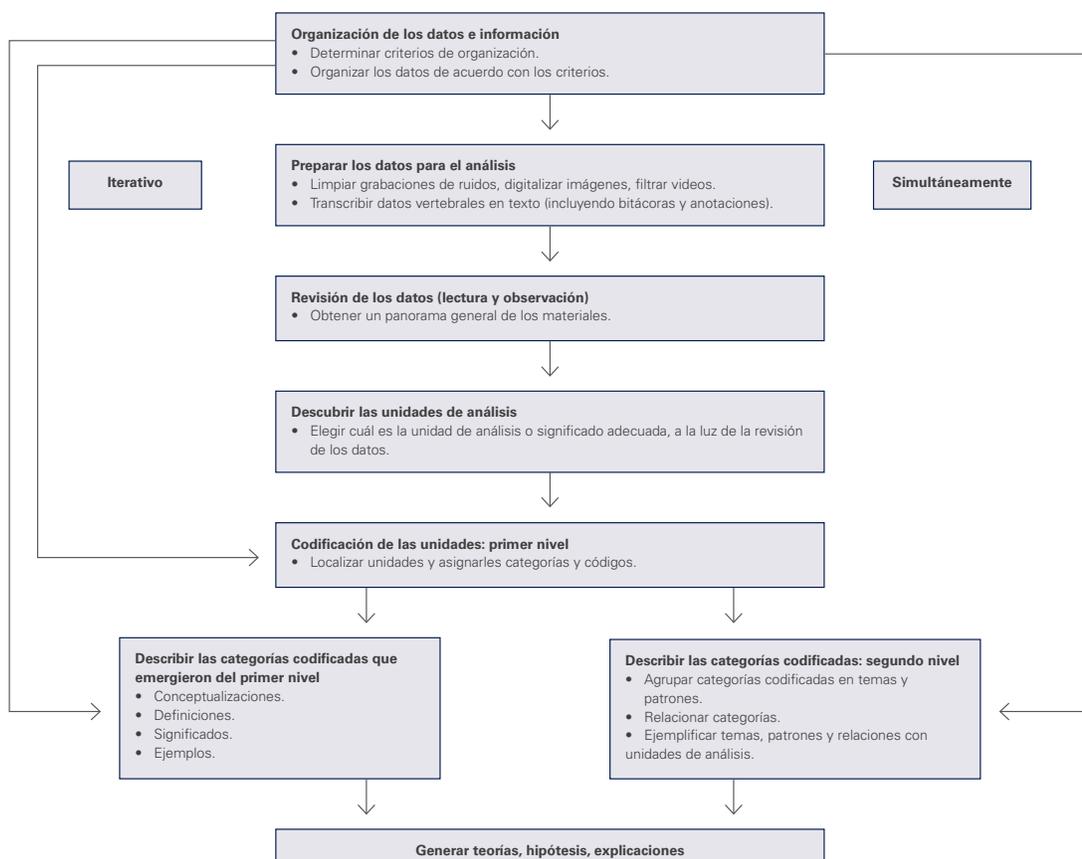
En la organización de los datos se identificaron algunos materiales que requieren su transcripción a medios electrónicos (procesador de datos). En este paso se deberán transformar todos los archivos verbales, videos o imágenes en texto, por lo que se requerirá la captura de toda la información generada en las entrevistas, grupos de enfoque, bitácoras, etc. Este trabajo necesitará la organización del personal que realizará las transcripciones, y el tiempo que tarden en esta actividad será variable; dependerá de la cantidad de información recopilada, lo complejo de la misma y del personal que participe en el proyecto.

## Paso B. Análisis del material recopilado

Una vez que se tiene todo el material organizado en formato de texto, se puede revisar nuevamente la información y reorganizarla mediante los criterios que se crean más convenientes (cronológicos, por sucesión de eventos, por tipo de datos, por grupo o participantes, por ubicación, por tema, por importancia del participante o el criterio que el equipo de evaluadores acuerde).

En la figura 3.3 se presenta el proceso completo para el análisis del material recopilado.

Figura 3.3. Análisis del material recopilado



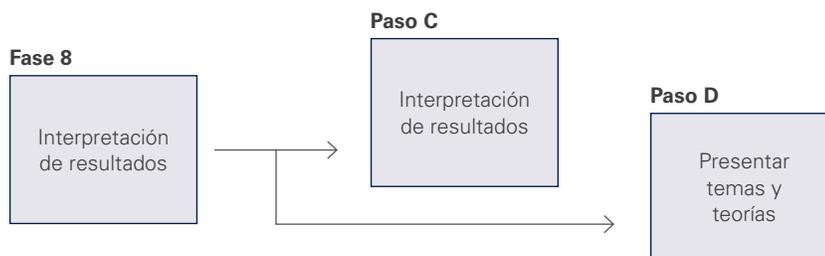
Fuente: Hernández Sampieri, Fernández Collado y Baptista Lucio, 2010.

## ¿Qué deberá realizar?

- Realizar el análisis de los datos recopilados mediante la organización de datos y la transcripción de materiales.
- Describa el procedimiento utilizado e incorpore sus resultados en la Evidencia de Formato Libre Cualitativo A (EFLCA. Estructurar los datos recopilados), en el Portafolios de Evidencias de Análisis Cualitativo.
- Revise los datos, describa las unidades de análisis, codifique las unidades, describa las categorías de primer nivel y codifique las categorías de segundo nivel. Registre sus resultados en la Evidencia de Formato Libre Cualitativo B (EFLB. Análisis del material recopilado).

## Fase 8. Interpretación de resultados

Figura 3.4. Fase 8. Interpretación de resultados.



Fuente: Hernández Sampieri, Fernández Collado y Baptista Lucio, 2010.

Para realizar la interpretación de los resultados, resulta fundamental darle sentido a:

- Las descripciones de cada categoría, es decir, dar una descripción completa de la categoría y ubicarla en el fenómeno que se evalúa.
- Los significados que cada categoría tiene para los participantes en la evaluación.
- La presencia de cada categoría, es decir, la frecuencia con la cual aparece en los materiales analizados en cierto sentido cuantitativo sin llegar a realizar un análisis estadístico.
- La relación entre categorías para encontrar vinculaciones, nexos y asociaciones entre ellas (temporales, causales, de conjunto-subconjunto).

Con base en la selección de temas y el establecimiento de relaciones entre categorías, se comienza a interpretar los resultados y a entender el fenómeno de evaluación, así como a generar hipótesis, explicaciones y teorías.

Para completar idealmente el ciclo del análisis cualitativo, es necesario:

- Producir un sistema de clasificación (tipologías).
- Presentar temas y teorías.

Con la finalidad de identificar las relaciones entre temas, se deben desarrollar interpretaciones de los mismos, los cuales emergen de manera consistente con respecto a los esquemas iniciales de categorizaciones y las unidades. Es una labor encontrar sentido y significado a las relaciones entre temas, y para lograrlo se pueden utilizar diversas herramientas a fin de visualizar con mayor facilidad dichas relaciones. Algunas herramientas para lograr la presentación gráfica y ordenada de las relaciones son:

- Mapas conceptuales.
- Matrices.
- Metáforas.
- Establecimiento de jerarquías.
- Calendarios.
- Fotografías.
- Videos.

#### ¿Qué deberá realizar?

- Interprete y registre los resultados a partir de la producción de sistemas de clasificación; asimismo, presente los temas y teorías; utilice la Evidencia de Formato Libre Cualitativo C (EFLCC. Interpretación de resultados).

## Portafolios de evidencias del análisis de resultados en el enfoque cualitativo

Las evidencias para el análisis de resultados en el enfoque cualitativo son:

- Procedimiento utilizado para estructurar los datos recopilados: utilice la evidencia de Formato Libre Cualitativo A (EFLCA. Estructurar los datos recopilados) en el Portafolios de Evidencias de Análisis Cualitativo.
- Análisis del material recopilado: utilice la Evidencia de Formato Libre Cualitativo B (EFLB. Análisis del material recopilado).
- Interpretación de resultados: utilice la Evidencia de Formato Libre Cualitativo C (EFLCC. Interpretación de resultados).

## Lista de cotejo para el análisis de resultados en el enfoque cualitativo

Una vez que ha concluido los cuatro pasos de las fases 7 y 8 del análisis cualitativo, aplique la lista de cotejo A, Análisis e interpretación de resultados que se especifica en el anexo 1. Si cumple con todos los criterios de calidad, ha concluido las fases 7 y 8.

## Glosario

- *Accesibilidad*: Grado en el que se integra a las personas con discapacidad al proceso de evaluación en igualdad de condiciones con las demás.
- *Adaptaciones al instrumento*: En el contexto de la evaluación, representan los ajustes necesarios al instrumento para garantizar que las personas con alguna discapacidad participen en igualdad de condiciones con el resto de los evaluados.
- *Análisis de ítems*: Procedimiento cuantitativo mediante el cual se determina el funcionamiento de los ítems de una prueba o encuesta. Permite establecer el número y la proporción de quienes responden la clave y cada uno de los distractores, así como la correlación biserial de las opciones con el puntaje total de la prueba o con un criterio de validación externo.
- *Administración del instrumento*: Proceso en el que una o más personas contestan el instrumento de evaluación.
- *Administrador del instrumento*: Persona responsable de llevar a cabo la aplicación de los instrumentos de evaluación conforme a los protocolos establecidos.
- *Alto impacto*: Se entiende que una evaluación es de alto impacto cuando sus resultados tienen consecuencias importantes para las personas o las instituciones; por ejemplo, los procesos de admisión o certificación.
- *Alfa de Cronbach*: Indicador que determina la consistencia interna del instrumento analizando el promedio de las correlaciones entre los reactivos.
- *Autoevaluación*: Ejercicio de valoración de las características (conocimientos, actitudes, valores, conducta, etcétera) que la persona que contesta el instrumento hace de sí misma.
- *Banco de reactivos*: Repositorio donde se resguardan y clasifican los reactivos que integran los instrumentos de evaluación; en él se administran los datos de identificación del reactivo, sus características métricas, las formas en las que se incorporó y las fechas en las que se utilizó.
- *Calificación*: Proceso de asignación de una puntuación o nivel de desempeño logrado a partir de los resultados de una medición.
- *Clave de corrección*: Conjunto de respuestas correctas a las preguntas de una prueba psicométrica. La clave de corrección también debe indicar la ponderación dada a cada una de las respuestas para el cálculo del puntaje bruto total.
- *Cuestionario*: Tipo de instrumento de evaluación que sirve para recolectar información sobre actitudes, conductas, opiniones, contextos demográficos o socioculturales, entre otros.
- *Coefficiente de correlación de Pearson*: Valor estadístico que mide la relación entre dos variables. Los resultados del cálculo se dan en valores de -1 a 1, lo que significa la relación entre las variables (positiva o negativa). Para realizar el cálculo se pueden utilizar programas como SPSS.
- *Confiabilidad*: Calidad de las mediciones obtenidas con un instrumento que se caracterizan por ser consistentes y estables cuando éste se aplica en distintas ocasiones.

Determina el grado en que un test es consistente al medir una variable. Algunas palabras asociadas con la confiabilidad son *estabilidad*, *exactitud*, *consistencia* y *precisión*, así como el Error Estándar de Medida. La confiabilidad se expresa usualmente mediante coeficientes de correlación obtenidos a partir de los puntajes en dos aplicaciones de la misma prueba al mismo grupo de sujetos, o haciendo uso de pruebas paralelas aplicadas al mismo grupo de sujetos. También se define como *Consistencia Interna*, que se obtiene mediante la determinación de las intercorrelaciones o covarianzas de los ítems (véanse *KR21* y *Alfa de Cronbach*).

- *Constructo*: Elaboración teórica formulada para explicar un proceso social, psicológico o educativo. Es un concepto que se ha formulado para ser usado en la ciencia. Es decir, es un concepto con valor científico. Se usa en esquemas teóricos y se define de tal manera que sea susceptible de ser observado y medido. Por ende, se convierte en algo cuantificable.
- *Correlación punto biserial*: Medida de consistencia que se utiliza en el análisis de reactivos; indica si hay una correlación entre el resultado de un reactivo y el resultado global del examen.
- *Corrección (puntuación) de pruebas*: Asignación de puntajes a las respuestas dadas por un sujeto a las preguntas de una prueba. La mayoría de las veces se realiza mediante la asignación de un punto por cada respuesta correcta (ítems binarios), es decir, que coincida con la clave de respuestas (véase *Clave de corrección*), aunque también son posibles otras formas de corrección, como restar a las respuestas correctas una proporción de las incorrectas.
- *Corrección de puntaje por azar*: Los ítems de opción múltiple permiten responder al azar y alcanzar un cierto número de aciertos, es decir, permiten la adivinación de las respuestas. La fórmula más común para el cálculo de puntaje por azar es: en donde  $X_c$  es el puntaje corregido,  $C$  es el número de respuestas correctas,  $I$  es el número de incorrectas y  $k$  es el número de opciones.
- *Criterio de evaluación*: Indicador de un valor aceptable sobre el cual se puede establecer o fundamentar un juicio de valor sobre el desempeño de una persona.
- *Criterial*: Tipo de medición en donde se compra a la persona que responde la prueba con un estándar previamente establecido.
- *Dificultad de un reactivo*: Indicador de la proporción de personas que responden correctamente el reactivo de un examen.
- *Distractores*: Opciones de respuesta incorrectas del reactivo de opción múltiple que probablemente serán elegidas por los sujetos con menor dominio en lo que se evalúa.
- *Deseabilidad social*: Tendencia de las personas a dar una imagen más favorable de sí mismas en el momento de responder un instrumento de evaluación, lo que ocasiona una distorsión en la medición.
- *Dicotómico*: Cualidad de una respuesta que adopta dos valores, como 0 o 1.
- *Dominio*: Conjunto de conocimientos, habilidades, destrezas, actitudes u otros atributos que tienen las siguientes propiedades: límites, extensión y definición. También se puede aplicar a contenidos, procedimientos u objetos.
- *Educación básica*: Tipo de educación que comprende los niveles de preescolar, primaria y secundaria en todas sus modalidades, incluyendo la educación indígena, la especial y la que se imparte en los centros de educación básica para adultos.
- *Educación media superior*: Tipo de educación que comprende el nivel de bachillerato, los demás niveles equivalentes a éste, así como la educación profesional que no requiere bachillerato o sus equivalentes.

- *Equiparación*: Método estadístico que se utiliza para ajustar las puntuaciones de las formas o versiones de un mismo instrumento, de manera tal que al sustentante le sea indistinto, en términos de la puntuación que se le asigne, responder una forma u otra.
- *Error de medida*: Diferencia entre el valor medido y el “valor verdadero”. Cuando la medida es más precisa, el error es más pequeño y viceversa.
- *Error estándar de medida*: Estimación de mediciones repetidas de una misma persona en un mismo instrumento que tienden a distribuirse alrededor de un puntaje verdadero. El puntaje verdadero siempre es desconocido porque ninguna medida puede ser una representación perfecta de un puntaje verdadero.
- *Escala*: Conjunto de números, puntuaciones o medidas que pueden ser asignados a objetos o sucesos con propiedades específicas a partir de reglas definidas.
- *Escala de actitud*: Conjunto de reactivos que tiene como propósito recolectar información del grado de aceptación o preferencia sobre algún aspecto de interés.
- *Escala Likert*: Conjunto de reactivos presentados en forma de afirmaciones o juicios, y ante los cuales se pide la reacción de los participantes. Se presenta la afirmación y se solicita al sujeto que externé su reacción eligiendo uno de los cinco puntos o categorías de la escala.
- *Escalamiento*: Proceso a través del cual se construye una escala que facilita la interpretación de los resultados que se obtienen en uno o varios instrumentos de evaluación, colocando las puntuaciones de los distintos instrumentos o formas a una escala común.
- *Especificaciones de tareas evaluativas o de reactivos*: Descripción detallada de las tareas específicas susceptibles de medición que deben realizar las personas que contestan el instrumento de evaluación. Deben estar alineadas al constructo definido en el marco conceptual.
- *Estándar*: Principio de valor o calidad en la conducción y el uso de los procedimientos de evaluación. Constituye el referente para emitir un juicio de valor sobre el mérito del objeto evaluado.
- *Estructura*: Está compuesta por los contenidos o aspectos disciplinares que mide un instrumento, así como por el número y el porcentaje relativo de reactivos o tareas evaluativas con que se integran el instrumento y su distribución.
- *Evaluación*: Proceso sistemático mediante el cual se recopila y analiza información, cuantitativa o cualitativa, sobre un objeto, sujeto o evento, con el fin de emitir juicios de valor al comparar los resultados con un referente previamente establecido. La información resultante puede ser empleada como insumo para orientar la toma de decisiones.
- *Formas de un instrumento*: Dos o más versiones de un instrumento que se consideran equivalentes, pues se construyen con los mismos contenidos y especificaciones estadísticas.
- *Funcionamiento diferencial del instrumento (DFT)*: Tendencia del instrumento a funcionar de manera diferente en distintas subpoblaciones, a pesar de que los individuos que las componen obtengan puntuaciones similares en el instrumento. Las subpoblaciones son definidas por algo no contenido entre los aspectos relacionados con el constructo evaluado, y suelen considerar características de los individuos que las componen, tales como el sexo, la edad, el grupo étnico o el estatus socioeconómico.

- *Funcionamiento diferencial del reactivo (DIF)*: Tendencia del reactivo a funcionar de manera diferente en diversas subpoblaciones, a pesar de que los individuos que las componen obtengan puntuaciones similares en el reactivo.
- *Grupo de enfoque*: Algunos autores los consideran como una especie de entrevistas que consisten en reuniones de grupos pequeños o medianos (tres a diez personas) en las cuales los participantes conversan en torno a uno o varios temas en un ambiente relajado e informal, bajo la conducción de un especialista en dinámicas grupales.
- *Índice de generalizabilidad*: Indicador que se calcula con el propósito de identificar y estimar la magnitud de las distintas fuentes de variación que pueden intervenir en las diferencias entre puntuaciones, o variación debida a las puntuaciones del universo y a las múltiples fuentes de error.
- *Instrumento de evaluación*: Herramienta de recolección de datos que suele tener distintos formatos atendiendo a la naturaleza de la evaluación, por ejemplo, instrumentos de selección de respuesta, instrumentos de respuesta construida, cuestionarios, observaciones, portafolios, entre otros.
- *Ipsativa*: Tipo de medición de la evaluación en donde se compara al sujeto contra sí mismo; este tipo de puntuación es utilizada en mayor medida en cuestionarios de autoevaluación.
- *Jueceo*: Método en el cual se utiliza la opinión de expertos (denominados jueces) para valorar y calificar distintos aspectos, tales como las respuestas y ejecuciones de las personas que participan en una evaluación, o la calidad de los reactivos, las tareas evaluativas y los estándares de un instrumento.
- *KR21*: Estadístico que permite determinar la confiabilidad de consistencia interna de una prueba. Deriva su nombre de las fórmulas propuestas por Kuder y Richardson en *Psychometrika* en los años treinta, de las cuales la número 21 fue particularmente exitosa. Tiene estrecha relación con el Alfa de Cronbach. Se calcula con la fórmula siguiente:  $KR_{21} = \frac{m}{n} \frac{S^2 - \sum s_i^2}{S^2}$ , en donde  $KR_{21}$  es el coeficiente de consistencia interna,  $m$  es el número de ítems de la prueba,  $S^2$  es la varianza del ítem y  $S^2$  es la varianza total del test.
- *Medición*: Proceso de asignación de valores numéricos a atributos de las personas; características de objetos o eventos de acuerdo con reglas específicas que permitan que sus propiedades puedan ser representadas cuantitativamente.
- *Modificaciones a las condiciones de aplicación*: En el contexto de la evaluación, representan los ajustes necesarios para garantizar que la administración del instrumento a las personas con alguna discapacidad se lleva a cabo de manera correcta.
- *Muestra*: Subconjunto de la población de interés que refleja las variables medidas en una distribución semejante a la de la población.
- *Nivel de desempeño*: Criterio conceptual que delimita el marco interpretativo de las puntuaciones obtenidas en una prueba y que refiere a lo que el sustentante es capaz de hacer en términos de conocimientos, destrezas o habilidades en el contexto del instrumento.
- *Normativa*: Tipo de medición en donde se compara al sujeto con el resto de las personas evaluadas.
- *Objeto de medida*: Conjunto de características o atributos que se miden en el instrumento de evaluación.
- *Parámetro estadístico*: Número que resume un conjunto de datos derivados del análisis de una cualidad o característica del objeto de estudio.

- *Parámetro de referencia:* Indicador estadístico del reactivo o de la tarea evaluativa que se obtiene al considerar un número grande de observaciones que provienen de las administraciones del instrumento ocurridas durante un lapso determinado.
- *Persona que responde el instrumento:* Sujeto que se enfrenta a algún instrumento de evaluación; este concepto incluye a un sustentante que responde una prueba de logro y a un informante en el caso de cuestionarios o encuestas.
- *Piloteo de las tareas evaluativas o los reactivos:* Recolección preliminar de datos mediante la administración de un nuevo instrumento de evaluación para valorar su funcionamiento en una muestra de la población objetivo o en una población con características similares a las de la población objetivo, y realizar ajustes orientados a su mejora y a su posterior administración.
- *Población objetivo:* Grupo de individuos sobre los cuales se desea que las inferencias elaboradas a partir de los resultados obtenidos con un instrumento de evaluación sean válidas.
- *Politómico:* Se refiere a que la calificación puede adoptar más de dos valores, como calificación parcial o gradual.
- *Porcentaje de acuerdos inter-jueces:* Medida del grado en que dos jueces coinciden en la puntuación asignada a un sujeto cuyo desempeño es evaluado a través de una rúbrica.
- *Porcentaje de acuerdo intrajueces:* Medida del grado en que el mismo juez, a través de dos o más mediciones repetidas a los mismos sujetos que evalúa, coincide en la puntuación asignada al desempeño de los sujetos, evaluados a través de una rúbrica.
- *Protocolo:* Conjunto de reglas o normas que establecen cómo se deben realizar ciertas actividades o pasos, además de incorporar de manera detallada los procedimientos y los estándares que se deben cumplir. Para su realización es fundamental considerar las características especiales de cada evaluación.
- *Prueba:* Instrumento de evaluación que tiene como propósito medir el grado de dominio, conocimiento o aptitud para valorar el mérito de personas, instituciones, programas, sistemas, entre otros.
- *Punto de corte:* En instrumentos de evaluación con referencia a un estándar de desempeño, es la puntuación mínima o el criterio a alcanzar o a superar para considerar que el nivel de desempeño de una persona cumple con lo esperado y distinguirlo de otro que no.
- *Puntuación:* Valor numérico obtenido durante el proceso de medición.
- *Referente:* Es aquello a lo que uno se refiere cuando se formula un juicio; toma forma de puntuación, valor, o de descripción de realizaciones.
- *Reactivo:* Unidad básica de medida de un instrumento de evaluación que consiste en una pregunta o instrucción que requiere una respuesta del sujeto.
- *Rúbrica:* Es una herramienta de valoración usada para medir el grado de cumplimiento de una actividad o trabajo. Se presenta como una pauta o tabla de doble entrada que permite unir y relacionar criterios de evaluación, niveles de logro y descriptores que se presentan en cuadro de doble entrada.
- *Sesgo:* Error en la medición de un atributo debido a una variable no controlada, como las diferencias culturales o lingüísticas de las personas evaluadas.
- *Tabla de especificaciones:* Matriz que permite identificar con toda precisión el objeto de medida o evaluación. Concentra la estructura del instrumento y las definiciones operacionales de lo que se desea medir (especificaciones de reactivos o de tareas evaluativas).

- *Tarea evaluativa*: Unidad básica de medida de un instrumento de evaluación que consiste en la respuesta que construye una persona o en la ejecución de una actividad susceptible de ser observada y graduada en su nivel de cumplimiento.
- *Teoría Clásica de los Tests (TCT)*: Teoría psicométrica que parte del supuesto de que el puntaje observado de una persona que responde un instrumento es la suma de su puntaje verdadero y un error aleatorio independiente del puntaje.
- *Teoría de la Generalizabilidad (TG)*: Teoría psicométrica que reconoce que existen diferentes fuentes de error de medida y enfatiza la estimación de cada uno por separado; proporciona un mecanismo para la optimización de la confiabilidad denominado coeficiente de generalización, esto es, se centra en los componentes de varianza que indican la magnitud de cada fuente de error que afecta la medición.
- *Teoría de Respuesta al Ítem (TRI)*: Teoría psicométrica que consiste en una familia de modelos estadísticos que modelan la probabilidad de cierta(s) respuesta(s) (por ejemplo, la respuesta correcta en un test de rendimiento óptimo) como función de las características tanto de la persona evaluada (por ejemplo, su nivel de habilidad en el constructo latente) como del reactivo (por ejemplo, su grado de dificultad).
- *Validez general*: Juicio que se hace respecto del grado en que el instrumento de medición mide lo que debe medir. Este juicio consiste en tener una idea clara de la variable que desea medirse y evaluar si las preguntas o los artículos del instrumento en realidad lo miden.
- *Validez del contenido*: Juicio sobre el grado en que el instrumento representa la variable objeto de medición, es decir, el grado en que representa el universo de la variable objeto de estudio.
- *Validez relacionada con el criterio*: Juicio que se hace al instrumento en cuanto a su capacidad para predecir la variable objeto de la medición.
- *Validez relacionada con el constructo*: Juicio sobre el instrumento respecto del grado en que una medición se relaciona consistentemente con otras mediciones sobre conceptos que están midiéndose.
- *Variable latente*: Variable "oculta", es decir, que no es susceptible de medirse directamente, sino a través de otras variables manifiestas (observables).
- *Varianza*: La varianza de una variable aleatoria es una medida de dispersión definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media.
- *Varianza irrelevante para el constructo*: Efecto de variables ajenas al constructo que mide el instrumento de evaluación y que afecta sus resultados; por consiguiente, es información que compromete la validez de las inferencias que se realizan.

## Bibliografía

- AERA, APA y NCM. American Educational Research Association, American Psychological Association y National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Amezcu Viedma, C., Jiménez Lara, A., Bermejo Ascorbe, R., García Alonso, R., y Ramalle Gómara, E. (1996). *Evaluación de Programas Sociales*. Madrid: Díaz de Santos.
- Aragón, L. E., y Silva, A. (2004). *Fundamentos teóricos de la evaluación psicológica*. México: Pax México.
- Best, J. (1982). *Cómo investigar en educación*. Novena edición. Madrid: Morata.
- Centro de Investigación en pólizas de salud de UCLA (2017, 17 de julio). Sección 2: Grupos de Enfoque (Focus Groups). Recuperado de: [http://healthpolicy.ucla.edu/programs/health-data-espanol/Documents/seccion\\_2\\_apendice\\_A.pdf](http://healthpolicy.ucla.edu/programs/health-data-espanol/Documents/seccion_2_apendice_A.pdf)
- Cook, D. A. y Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American Journal of Medicine*, 119(2), 166.e7-166.e16.
- Council of Europe. (2011). *Manual for language test development and examining*. United Kingdom: Association of Language Testers in Europe and Council of Europe Language Policy Division. Recuperado de: [http://www.coe.int/t/dg4/linguistic/ManualLanguage-Test-Alte2011\\_EN.pdf](http://www.coe.int/t/dg4/linguistic/ManualLanguage-Test-Alte2011_EN.pdf)DOF.
- DOF. Diario Oficial de la Federación (2017, 12 de abril). *Criterios técnicos y de procedimiento para el análisis de los instrumentos de evaluación, el proceso de calificación y la definición de las listas de prelación de los concursos de oposición para el ingreso al SPD en EB y EMS para el ciclo esc. 2017-2018*. México: INEE.
- DOF (2017, 28 de abril). *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación*. México: Unidad de Normatividad y Política Educativa-INEE.
- Downing, S.M. (2004). Reliability: On the reproducibility of assessment data. *Medical Education*, 38(9), 1006-1012.
- ETS. Educational Testing Service (2015). *ETS Standards for quality and fairness*. Recuperado de: <https://www.ets.org/s/about/pdf/standards.pdf>
- Fidalgo, A. M. y Madeira, J. M. (2008). Generalized Mantel-Haenszel methods for differential item functioning detection. *Educational and Psychological Measurement*, 68(6), 940-958.
- Ganster, D. C., Hennesey, H. W. y Luthans, F. (1983). Social desirability response effects: Three alternative models. *Academy of Management Journal*, 26(2), 321-331.
- Havercamp, S.M. y Reiss, S. (2003). A comprehensive assessment of human strivings: Test-retest reliability and validity of the Reiss profile. *Journal of Personality Assessment*, 81(1), 123-132.
- Hernández Sampieri, R., Fernández Collado, C., y Baptista Lucio, P. (2010). *Metodología de la Investigación*. México: McGraw-Hill Interamericana.

- Holland, P. W. y Strawderman, W. E. (2011). How to average equating functions, if you must. En A. A. von Davier (ed.), *Statistical models for test equating, scaling, and linking* (pp. 89–107). Nueva York: Springer.
- INAE. Instituto Nacional de Asesoría Especializada (2017). *Guía 1. Marcos de Referencia*. México: autor.
- INEE. Instituto Nacional para la Evaluación de la Educación (2013). Ley del Instituto Nacional para la Evaluación de la Educación. Recuperado el 29 de julio de 2017, de: <http://www.inee.edu.mx/index.php/517-reforma.../marco.../1604-ley-del-inee>.
- INEE (2017). Programa de Mediano Plazo del Sistema Nacional de Evaluación Educativa 2016-2020. México: autor.
- JCSEE. Joint Committee on Standards for Educational Evaluation (2011). *The program evaluation standards* (3a ed.). Thousand Oaks, CA: Corwin Press.
- Jonsson, A. y Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review* 2, 130–44. doi:10.1016/j.edurev.2007.05.002
- King, W. y Wand, J. (2006). Comparing incomparable survey responses: Evaluating and selecting anchoring vignettes. *Political Analysis*, 15(1), 46-66.
- Kolen, M. J., y Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3a ed.). Nueva York: Springer
- Lewis, D. M., Patz R. J., Sheinker, A. y Barton, K. (2002). Reconciling standardization and accommodation: Inclusive norms and inclusive reporting using a taxonomy for testing accommodations. En *Supporting inclusion in large-scale assessment: Reconciling standard school testing practices and standardized tests*. Simposio llevado a cabo en la reunión anual 2002 de la American Educational Research Association. Nueva Orleans, LA.
- Martínez-Miguélez, M. (2017, 19 de noviembre). Los grupos focales de discusión como método de investigación. Recuperado de: <http://miguelmartinezm.atSPACE.com/gruposfocales.html>.
- Muñiz, J. (2010). Las teorías de los test: Teoría Clásica y Teoría de Respuesta a los Ítems. *Papeles del Psicólogo*, 31(1), pp. 57-66. Recuperado el 15 de septiembre de 2017, de: <http://www.papelesdelpsicologo.es/pdf/1796.pdf>.
- Muraki, E. (1999). Stepwise analysis of differential item functioning based on multiple-group partial credit model. *Journal of Educational Measurement*. 36(3), 217-232. Recuperado de: <http://www.jstor.org/stable/1435155>
- Navas Ara, M. J., Fidalgo Aliste, Á. M., Gabriel Esteban, C., Suárez Falcón, J. C., Brioso Díez, Á., Gil Escudero, G., Sarria Sánchez, E. (2010). *Métodos, diseños y técnicas de investigación psicológica*. Madrid: UNED.
- Rezaei, A. R. y Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15(1), 18–39.
- Shun-Wen, Ch. (2006). Methods in Scaling the Basic Competence Test. *Educational and Psychological Measurement*, 66(6), 907-927.
- Stemler, S.E. y Tsai, J. (2008). 3 Best practices in interrater reliability three common approaches. En J. Osborne (ed.), *Best practices in quantitative methods* (pp. 29–49). SAGE Publications.

- 
- Stellmack, M. A., Konheim-Kalkstein, Y. L., Manor, J. E., Massey, A. R., y Schmitz, J. A. P. (2009). An assessment of reliability and validity of a rubric for grading APA-style introductions. *Teaching of Psychology, 36*(2), 102-107
- Tenbrink, T. D. (1999). *Evaluación. Guía práctica para profesores*. Quinta edición. C. Fernández Aguinaco (trad.). Madrid: Narcea.
- UNICEF [por sus siglas en inglés]. Fondo de las Naciones Unidas para la Infancia. (2012). *Cómo diseñar y gestionar evaluaciones centradas en la equidad*. Nueva York: autor. Recuperado de: <http://siare.clad.org/fulltext/2241800.pdf>

## Siglas y acrónimos

AEL	Autoridad Educativa Local
ATP	Asesores Técnicos Pedagógicos
CTDUMIE	Criterios Técnicos para el Desarrollo, Uso y Mantenimiento de Instrumentos de Evaluación
DOF	Diario Oficial de la Federación
EXCALE	Exámenes de la Calidad y el Logro Educativos
INEE	Instituto Nacional para la Evaluación de la Educación
PEEME	Programas Estatales de Evaluación y Mejora Educativa
PMP-SNEE	Programa de Mediano Plazo del Sistema Nacional de Evaluación Educativa
PNEE	Política Nacional de Evaluación de la Educación
PRONAEME	Proyecto Nacional de Evaluación y Mejora Educativa de Escuelas Multigrado
SEN	Sistema Educativo Nacional
SNEE	Sistema Nacional de Evaluación Educativa
SPD	Servicio Profesional Docente

## Anexo

### Principales características y técnicas del enfoque cualitativo

#### Características del enfoque cualitativo

En el enfoque cuantitativo, como en el cualitativo, se deben plantear preguntas claras y precisas; la diferencia entre lo *cuanti* y lo *cuali* estriba en que en los cuestionarios cuanti o la observación cuali las preguntas están más formalizadas/estructuradas que en el enfoque cualitativo. En el enfoque cuantitativo lo que se busca es estimar magnitudes, mientras que en el cualitativo lo que se busca es observar a profundidad cualidades significativas, por eso se dice que una explicación fuerte debería usar coherentemente los dos enfoques.

Ambos pueden desarrollar hipótesis *antes o después* (es el caso de algunos cuestionarios cuanti, que parten de *hipótesis nula* y que permiten, al analizar e interpretar los datos, postular después una hipótesis *a posteriori*; lo mismo suceden en el enfoque cuali, que puede partir de preguntas sin hipótesis (es el caso de los grupos de discusión) o de una hipótesis previa (es el caso del grupo focal). Ambos, sin embargo, tienen que elaborar protocolos rigurosos y sistemáticos para recopilar la información, analizarla e interpretarla.

Mientras que un enfoque cuanti se basa en el criterio de representatividad (lógico-matemático), uno cuali se fundamenta en el criterio de significatividad (lógico-determinista). Pero ambos son útiles, y más si se usan para explicar problemas complejos.

Algunas de las características más importantes del enfoque cualitativo son (INAE, 2017, pp. 71-73):

- Tiene una visión de los eventos, acciones, valores, normas, etc., desde la posición y la percepción de los agentes que están siendo evaluados.
- Se opta por estudiarlos en el contexto en el que ocurren.
- Pone el énfasis en la interpretación, es decir, en la comprensión del fenómeno educativo como un todo y el significado que ello tiene para sus participantes en el marco de realidades concretas.
- Tiene un carácter procesal.
- Se inclina por una estrategia de trabajo relativamente abierta y no estructurada donde la teoría y el levantamiento de información empírica se entremezclan, por lo que las propuestas teóricas y conceptuales puede ocurrir durante o al finalizar el trabajo de campo, más que como un punto de partida del proyecto de evaluación.

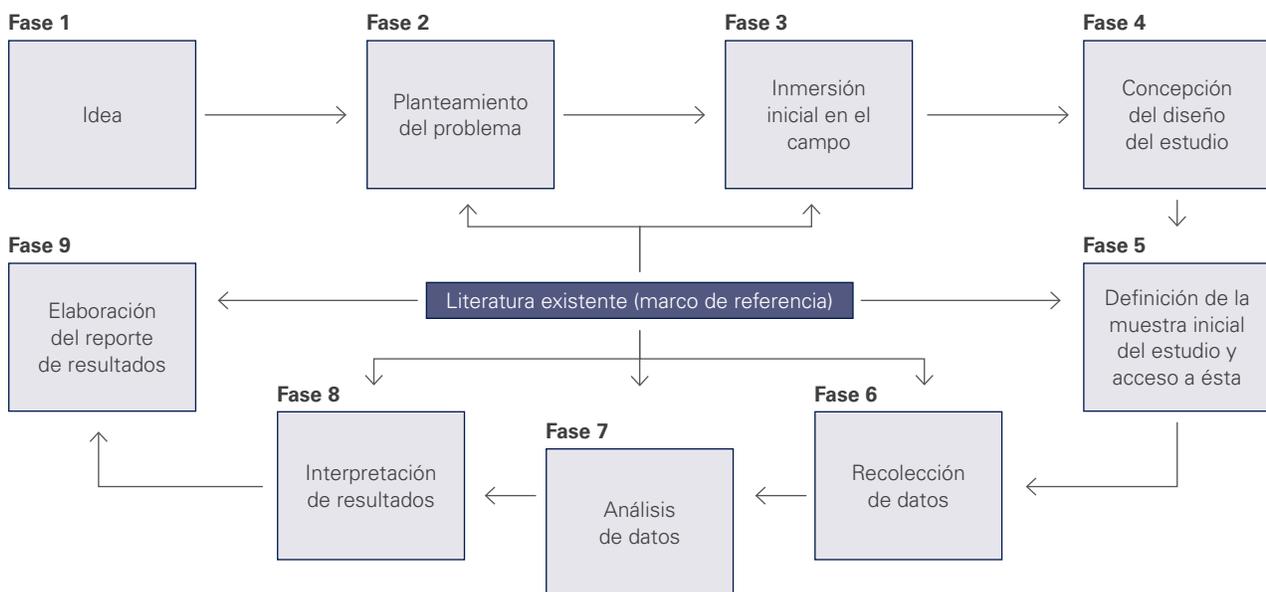
Adicionalmente, el enfoque cualitativo tiene las siguientes características:

- Se plantea un problema que no sigue un proceso claramente definido. Sus planteamientos no son tan específicos como en el enfoque cuantitativo y las preguntas de investigación no siempre se han conceptualizado ni definido por completo.

- En lugar de iniciar con una teoría particular, se comienza examinando el mundo social, y en este proceso desarrolla una teoría coherente con los datos de acuerdo con lo que se observa, es decir, va de lo particular a lo general, y aplica un proceso inductivo.
- En la mayoría de los estudios cualitativos no se prueban hipótesis, éstas se generan durante el proceso y van refinándose conforme se recaban más datos o son el resultado del estudio.
- Utiliza métodos de recolección de datos no estandarizados ni completamente pre-determinados. No se efectúa una medición numérica, por lo cual el análisis no es estadístico. La recolección de datos consiste en obtener las perspectivas y puntos de vista de los participantes.
- Utiliza técnicas para la recolección de datos, como la observación no estructurada, entrevistas abiertas, revisión de documentos, discusión en grupo, evaluación de experiencias personales, registro de historias de vida, e interacción e introspección con grupos o comunidades.
- El proceso de indagación es más flexible y se mueve entre las respuestas y el desarrollo de la teoría. Su propósito es "reconstruir" la realidad, tal como la observan los actores de un sistema social previamente definido. A menudo se le llama *holístico*, porque se precia de considerar el "todo" sin reducirlo al estudio de sus partes.

De acuerdo con Hernández Sampieri, Fernández Collado y Baptista (2010, p. 8), el enfoque cualitativo es un proceso integrado por nueve fases, que se muestran en la figura A.1.

Figura A.1. Proceso cualitativo



Fuente: Hernández Sampieri, Fernández Collado y Baptista Lucio, 2010, p. 8.

En el proceso de evaluación cualitativa destacan los siguientes aspectos:

- Aunque hay una revisión inicial de la literatura, ésta puede complementarse en cualquier etapa del estudio y apoyar desde el planteamiento del problema hasta la

elaboración del reporte de resultados (la vinculación teoría-etapas del proceso se representa mediante las líneas curvas).

- Con frecuencia, es necesario regresar a etapas previas; por ello, las líneas, desde la fase 3 a la 9, van en ambos sentidos.
- La inmersión inicial en el campo significa sensibilizarse con el ambiente o entorno en el cual se llevará a cabo la evaluación; identificar informantes que aporten datos y nos guíen por el lugar; adentrarse y compenetrarse con la situación de evaluación, además de verificar la factibilidad del estudio.
- La muestra, la recolección y el análisis son fases que se realizan prácticamente de forma simultánea.
- El planteamiento del problema debe considerar los objetivos; preguntas de investigación; justificación y viabilidad de la evaluación; exploración de las deficiencias en el conocimiento del problema, y definición inicial del ambiente. Asimismo, se debe establecer una muestra inicial para definir las personas que participarán en la evaluación y establecer el contexto o ambiente, es decir, los lugares para la recolección de los primeros datos a fin de precisar el tiempo y lugar de la evaluación.
- El muestreo en la evaluación cualitativa se determina durante o después de la inmersión inicial; se puede ajustar en cualquier momento, no es probabilístico y no busca generalizar resultados.
- La recolección de datos implica que el investigador es su instrumento más importante, se realiza en ambientes naturales y no se miden variables. Para realizarlo utiliza herramientas como: biografías e historias de vida, documentos, registros, artefactos, grupos de enfoque, entrevistas, observación, anotaciones y bitácoras de campo.
- El análisis de la información inicia con la organización de los datos recopilados y la transcripción del material, además de una bitácora de análisis para documentar el proceso. Los criterios de rigor, validez y confiabilidad para el análisis del material son: Dependencia, Credibilidad, Transferencia y Confirmación.

Para concluir, la tabla A.1 contiene un cuadro comparativo de las etapas de la evaluación y las características del enfoque cuantitativo y el enfoque cualitativo.

Previamente se mencionó que para la recolección de datos se requiere de técnicas e instrumentos para su registro. A continuación se describen la entrevista y los grupos de enfoque, que son las técnicas más representativas del enfoque cualitativo.

Tabla A.1 Cuadro comparativo entre el enfoque cuantitativo y cualitativo

Características cuantitativas	Procesos fundamentales de la evaluación	Características cualitativas
<ul style="list-style-type: none"> <li>▪ Orientación hacia la descripción, predicción y explicación.</li> <li>▪ Específico y acotado.</li> <li>▪ Dirigido hacia datos medibles u observables.</li> </ul>	Planteamiento del problema de evaluación.	<ul style="list-style-type: none"> <li>▪ Orientación hacia la exploración, la descripción y el entendimiento.</li> <li>▪ General y amplio.</li> <li>▪ Dirigido a las experiencias de los participantes.</li> </ul>
<ul style="list-style-type: none"> <li>▪ Rol fundamental.</li> <li>▪ Justificación para el planteamiento y la necesidad de la evaluación.</li> </ul>	Revisión literaria	<ul style="list-style-type: none"> <li>▪ Papel secundario.</li> <li>▪ Justificación para el planteamiento y la necesidad de la evaluación.</li> </ul>
<ul style="list-style-type: none"> <li>▪ Instrumentos predeterminados.</li> <li>▪ Datos numéricos.</li> <li>▪ Número considerable de casos.</li> </ul>	Recolección de datos	<ul style="list-style-type: none"> <li>▪ Los datos emergen poco a poco.</li> <li>▪ Datos en texto o imagen.</li> <li>▪ Número relativamente pequeño de casos.</li> </ul>
<ul style="list-style-type: none"> <li>▪ Análisis estadístico.</li> <li>▪ Descripción de tendencias, comparación de grupos o relación entre variables.</li> <li>▪ Comparación de resultados con predicciones y estudios previos.</li> </ul>	Análisis de los datos	<ul style="list-style-type: none"> <li>▪ Análisis de textos y material audiovisual.</li> <li>▪ Descripción, análisis y desarrollo de temas.</li> <li>▪ Significado profundo de los resultados.</li> </ul>
<ul style="list-style-type: none"> <li>▪ Estándar y fijo.</li> <li>▪ Objetivo y sin tendencias.</li> </ul>	Reporte de resultados de la evaluación	<ul style="list-style-type: none"> <li>▪ Emergente y flexible.</li> <li>▪ Reflexivo y con aceptación de tendencias.</li> </ul>

Fuente: adaptado de Hernández Sampieri, Fernández Collado y Baptista Lucio, 2010, p. 16.

## Entrevista

De acuerdo con Best (1982, p. 159), la entrevista es, en cierto sentido, un tipo verbal de cuestionario. En lugar de escribir las respuestas, el sujeto proporciona la información necesaria verbalmente en una relación personal. Para Tenbrink (1999, p. 142), la entrevista es una técnica de recogida de información muy valiosa, porque permite obtener información que sería muy difícil conseguir de otro modo.

Características de la técnica de entrevista:

- Es muy flexible.
- Se pueden repetir las preguntas con otras palabras (cuidando no cambiar la intención de la pregunta).
- Aclarar términos.
- Cambiar la línea de las preguntas.

### Aspectos críticos de la entrevista

Si el entrevistador no tiene cuidado con el desarrollo de la entrevista, podría:

- Plantear cuestiones inadecuadas.
- Confundir a su interlocutor con preguntas mal formuladas.
- Frustrar al entrevistado saltando de un tema a otro.

- Amenazarlo siendo demasiado personal y provocando que se “cierre” totalmente.
- Influnciar al entrevistado obteniendo las respuestas que quiere “oír”.

Como técnica de recogida de datos, la entrevista tiene ventajas particulares. En áreas que se refieren a la motivación humana, según se revela por las razones para actuar, sentir y mostrar determinadas actitudes, la entrevista puede ser lo más eficaz. En manos de un hábil interrogador, mediante interestimulación de sí mismo y del sujeto, es posible una profundidad en las respuestas, una penetración que no se logra con ningún otro medio.

Sin embargo, esta técnica ocupa mucho tiempo y es una de las más difíciles de realizar con éxito. El peligro de las falsas interpretaciones es constante. Como la objetividad, sensibilidad y penetración del interrogador son cruciales, este procedimiento requiere un nivel de habilidad que de ordinario no poseen los investigadores inexperimentados (Best, 1982, p. 161).

## Recomendaciones para realizar entrevistas

Algunas sugerencias para realizar la entrevista y conseguir información útil son:

- *Estar preparado para la entrevista.* Dada la flexibilidad de la técnica, algunas entrevistas incluso pueden surgir en el momento menos esperado, por lo que además de tener claridad en el propósito de evaluación se debe tener claridad del tipo de información que se quiere conseguir con la entrevista.
- *Establecer un “rapport”<sup>17</sup> (y mantenerlo).* Un buen entrevistador establece este *rapport* comenzando por un tema no conflictivo y no amenazante. Después, durante la entrevista, es amistoso, tranquilo y paciente; siempre que sea posible, es positivo y muestra su aprobación ante el hecho de que el interlocutor está hablando. No siempre es posible aprobar o estar de acuerdo con lo que dice, pero se puede aprobar el hecho de que esté diciendo algo. Un apoyo constante, con asentimientos de cabeza o dar ánimo verbalmente para que continúe, puede ayudar mucho a establecer y mantener ese *rapport*.
- *Guiar al entrevistado.* Un entrevistador puede mantener al entrevistado en el tema por una serie de preguntas cuidadosamente planeadas. Con todo, a veces éste puede “bloquearse” en una pregunta concreta porque se siente amenazado por ella o porque no se siente capacitado para responder. Cuando esto pasa, es legítimo buscar información relacionada con ello, pero no lo es perseguir al interlocutor hasta el extremo de que se sienta irritado. Si continúa sin querer hablar sobre un tema dado, el entrevistador debe sacar preguntas sobre temas relacionados ligeramente con él, y

<sup>17</sup> El *rapport* está basado en la confianza, el respeto y la aceptación mutuos. Sattler (1992, citado por Aragón y Silva, 2004, p. 183) dice que establecer el *rapport* se facilita si el entrevistador mantiene un contacto visual apropiado; si muestra una postura natural, relajada y de atención; si habla de manera lenta, clara y amistosa, con un tono cálido y expresivo; si no juzga al entrevistado; si hace que la entrevista sea una empresa conjunta, y si demuestra interés, comprensión y reconocimiento. Por supuesto, es importante permitir la espontaneidad natural del entrevistado. Por su parte, el entrevistador debe emplear vocabulario apropiado, aclarar términos al entrevistado y formular preguntas de manera adecuada y directa, con varias alternativas y en el momento oportuno, evitando hacer aquellas que se respondan sólo con sí o no, así como preguntas largas y múltiples que sugieran respuestas embarazosas.

poco a poco llevar al entrevistado al tema de nuevo. Si después de dos o tres veces de realizar este intento todavía no ha tenido éxito, es mejor dejar esa cuestión y pasar a otra o terminar la entrevista. Normalmente, todo lo que se necesita para que el interlocutor siga hablando y dando información es ser receptivo y de vez en cuando pedir más detalles o una explicación. Pequeños comentarios como: "Sí, sigue"; "Y entonces, ¿qué pasó?"; o "¿Se dijo algo más?"; pueden ayudar mucho al entrevistado a recordar información importante.

- *Obtener una información exacta.* Por supuesto, la información que se obtiene de un entrevistado será tan exacta como él quiera, pero si no está registrada cuidadosamente, será todavía menos exacta. Por tanto, es importante anotar lo que se ha dicho. Una técnica muy útil es repetir la información clave dada por el entrevistado y ver si está de acuerdo con ello. Esto ayudará a determinar si lo ha anotado con exactitud o no. A veces se conseguirá una información inexacta a partir de una entrevista, no porque el entrevistado distorsione la información intencionalmente, sino porque no puede responder. Hay que hacer todo lo que se puede durante la entrevista para averiguar si el entrevistado está capacitado para dar la información que se le está pidiendo (Tenbrink, 1999, pp. 165-166).

### Instrumento para el registro de información

La entrevista es una técnica de interrogación, es un procedimiento que requiere de un instrumento para el registro de la información de la evaluación. El instrumento que guía al entrevistador es el plan de la entrevista. Los elementos que pueden incluirse en el plan de entrevista son:

- Propósito de la entrevista.
- Orientaciones para establecer un *rapport* con el entrevistado.
- Preguntas de la entrevista.
- Guion de la estructura de la entrevista.
- Espacio para el entrevistado, fecha, hora y lugar de su realización.
- Espacio para el registro de las notas durante la realización de la entrevista.

Un plan de entrevista está normalmente escrito para que se pueda hacer referencia a cada uno de los elementos enumerados anteriormente durante la misma. Sin embargo, al prepararse para entrevistas cortas y menos formales, esos elementos no se necesitan escribir; con todo, se debe pensar un poco sobre cada uno de esos elementos al prepararse para la entrevista (Tenbrink, 1999, pp. 283-285).

#### *Un enunciado del propósito de la entrevista*

Una vez que se ha decidido lo que se espera lograr en una entrevista, es importante mantener ese fin en la mente al desarrollarla. Si lo primero en el esquema es un enunciado del propósito, serviría de recuerdo y será más probable que se obtenga la información que se quiere; hay que especificar primero el propósito primordial de la entrevista, y preguntarse: "¿Qué tipo de información es más probable que pueda obtener de ella?" Las preguntas que se hacen durante la entrevista estarán determinadas en parte por la respuesta a esta primera pregunta.

### *Un conjunto de orientaciones para establecer un rapport*

Un buen entrevistador puede establecer rápidamente un *rapport* con el interlocutor. Un plan previo, cuando sea posible, puede ser efectivo en el primer momento de una entrevista. Lo primero que se debe hacer es identificar algunos temas no amenazantes con los que se pueda abrir la conversación, y, entonces, pensar en un modo de hacer el propósito de la entrevista significativo y aceptable para el interlocutor. Si, por alguna razón, esto no se debe dar a conocer al entrevistado (porque podría sentirse amenazado o no ser objetivo y por tanto no responder honestamente y con objetividad), se debe inventar un seudopropósito que sea plausible y legítimo. En tercer lugar, es necesario establecer razones sólidas, significativas, por las que el entrevistado deba ser totalmente sincero. También, hay que ofrecer argumentos convincentes de que la información se mantendrá estrictamente confidencial y no se usará contra el entrevistado. Es preciso estar preparado para demostrar de qué modo la información que se obtenga se puede usar para ayudar al sujeto.

### *Un conjunto de preguntas*

El entrevistador debe estar seguro de que las preguntas son cortas y fáciles de entender. El que responde debe recordar la pregunta que se le hace al pensar en su respuesta; en una entrevista no se pueden releer las preguntas como en un cuestionario. Conviene utilizar palabras comúnmente comprensibles y no formular las preguntas en estructuras gramaticales complejas.

### *Un guion de cómo estaría estructurada la entrevista*

Una entrevista puede estar bien estructurada de varias maneras. La estructura particular a utilizar debe estar determinada por el objetivo de la entrevista y la naturaleza de la relación entre el entrevistador y el entrevistado. Algunas están altamente estructuradas. Las preguntas están casi completamente elaboradas por adelantado, se reservan un tiempo y lugar específicos, y el propósito formal está bien especificado.

El otro extremo es una entrevista completamente desestructurada. La mayoría de las preguntas no están formuladas por adelantado y el entrevistador lo hace "a ojo," dejando que una pregunta lleve a otra. Este tipo de estructura (o su ausencia) es a veces útil para obtener una información muy general. A menudo se descubren intereses inesperados, actitudes y opiniones durante estas entrevistas tan poco estructuradas; a causa de su tono informal de conversación, son buenas para abrir y mantener abiertas líneas importantes de comunicación; este tipo de entrevista es muy bueno para situaciones donde el entrevistado se sienta amenazado y dude en hablar.

### *Nota sobre la hora y lugar de la entrevista*

Una nota breve de la persona, la fecha, el lugar y la hora de la entrevista es muy útil para: 1) recordar a quién se va a entrevistar o se entrevistó, cuándo y dónde, y 2) cuando se

registre la información que se obtenga de una entrevista, se sabrá cuándo se recogió. Es fácil dejar para más tarde el registro de las notas de la entrevista. Hay que consignar al menos quién fue entrevistado y dónde.

### *Espacios para tomar notas*

Durante la mayoría de las entrevistas planeadas, el entrevistador querrá tomar notas breves. Tener un lugar disponible para esto en el plan le asegura de que cuando se archiven las notas, la información sobre por qué se tuvo la entrevista, con quién, en qué modo, etc., no se perderá.

## **Grupo de enfoque**

El grupo de enfoque<sup>18</sup> se considera una técnica acorde al enfoque cualitativo. “Es particularmente útil para explorar los conocimientos, las prácticas y las opiniones, no sólo en el sentido de examinar lo que la gente piensa sino también cómo y por qué piensa como piensa” (Kitzinger, 1995, citado por Kornblit, 2007, p. 77). Según Martínez-Miguélez (2017), el grupo focal “es un método de investigación colectivista, más que individualista, y se centra en la pluralidad y variedad de las actitudes, experiencias y creencias de los participantes, y lo hace en un espacio de tiempo relativamente corto”. Para Amezcua y sus colaboradores un grupo de enfoque es una sesión informal en la que participantes que representan una población diana discuten varios aspectos sobre un tema o asunto específico (1996, p. 40).

El grupo de enfoque orienta (focaliza) su atención e interés en un tema específico de estudio e investigación que le es propio, por estar cercano a su pensar y sentir; y es de discusión porque realiza su principal trabajo de búsqueda por medio de la interacción discursiva y la contratación de las opiniones de sus miembros.

## **Planificación de un grupo de enfoque**

La planificación e implementación de una serie de grupos de enfoque comprende los siguientes pasos (Centro de Investigación en pólizas de salud de UCLA, 2017):

1. *Reunir y revisar los datos existentes.*
2. *Determinar el propósito del grupo de enfoque.* Para determinar el propósito del grupo de enfoque, debe responder a la pregunta: ¿qué información se requiere del grupo de enfoque?

---

<sup>18</sup> También conocido como grupo focal o *focus group*.

3. *Elegir la audiencia.* Los participantes en el grupo de enfoque deben tener similitudes —en el sentido de que se sientan cómodos compartiendo sus puntos de vista con el grupo— y ser seleccionados de acuerdo a ciertos criterios que se determinarán con el propósito de la evaluación y el propósito de la realización del grupo de enfoque.

El grupo de enfoque debe estar integrado por entre 6 y 12 personas, que comparten las características de la población de la evaluación. Es conveniente invitar a 12 personas debido a que algunas no se presentan, pero tampoco invite demasiadas.

4. *Realizar la guía de discusión.* La guía de discusión es un instrumento que necesita para mantener el enfoque de la discusión del grupo. Debe tener un guion para la introducción y una lista de preguntas con respuestas abiertas que serán las generadoras de la discusión. La guía de discusión incluye:

- Guion para la introducción.
- Propósito.
- Reglas de conducta.
- Preguntas para el grupo de enfoque:
  - Preguntas de introducción.
  - Preguntas de transición.
  - Preguntas clave.
  - Preguntas para terminar.
  - Preguntas de resumen.
- Declaraciones para cerrar la discusión.

Al contar con una guía de discusión, se procura tener el mismo propósito, la misma estructura, reglas de conducta y preguntas, en caso de que se requiera la realización de varios grupos de enfoque.

5. *Elaborar el horario para el grupo de enfoque.* Escoja el lugar, la hora y la fecha en que realizará el grupo de enfoque.

6. *Reclutar a los participantes para el grupo de enfoque.*

- Los participantes de los grupos de enfoque deben ser cuidadosamente reclutados, evitando reunir en el mismo grupo a amigos, familiares, vecinos, jefes, etc. Además, se debe invitar a gente con experiencia en este tipo de grupos.
- Las estrategias de reclutamiento de los participantes pueden realizarse a partir de listas existentes de grupos; otras veces se piden nombres a otros grupos; también se reclutan participantes pidiendo nombres de posibles nuevos colaboradores.
- Una vez que se elige a los participantes (Amezcu Viedma, Jiménez Lara, Bermejo Ascorbe, García Alonso y Ramalle Gómara, 1996, pp. 43 y 44):
  - Hay que lograr que los participantes vengan a la sesión y colaboren.
  - Explicar a los participantes de qué se trata y para qué se hacen los grupos de enfoque.

- Fijar fechas, horarios y lugares para la sesión que no entren en conflicto con fechas, horarios y lugares del calendario comunitario.
- Contactar con los posibles participantes 12 a 14 días antes de la sesión.
- Enviar una invitación personalizada una semana antes de la reunión.

7. *Reclutar un moderador y una persona para tomar apuntes.* Identifique un moderador para tratar la discusión del grupo. La guía de discusión que se elaboró será la herramienta que orientará al moderador para obtener la información deseada. Durante el grupo de enfoque, el moderador es responsable de establecer las reglas (aplicar las reglas establecidas en el guion), mantener la discusión centrada en el tema, lograr comentarios de los participantes callados, y preguntar, cuando se necesita, para clarificar u obtener información adicional de los participantes.

El papel de la persona que toma apuntes es: grabar la sesión del grupo de enfoque, escuchar con atención, observar las interacciones del grupo y hacer un resumen de los puntos y observaciones que no fueron capturados por la grabadora.

8. *Conducir el grupo de enfoque.* Los grupos de enfoque duran entre 1 y 2 horas, aunque depende de la edad de los participantes y de su propósito. Una reunión normal, para una investigación corriente, puede durar una o dos horas. En algunos casos, puede durar hasta cuatro y, excepcionalmente, puede requerir un fin de semana. En todo caso, mucho dependerá de la dinámica particular del grupo y del tema tratado, y será el director de la evaluación el que decidirá cuándo un tema ha sido suficientemente cubierto y saturado para los fines que persigue la evaluación.

La conducción del grupo tiene tres momentos:

*Inicio.* El moderador da la bienvenida al grupo, se realizan las presentaciones y se da una breve explicación del proyecto y el objetivo del grupo de enfoque haciendo énfasis en los beneficios que tendrán de los resultados que se generen con la realización del grupo de enfoque. Se explican los procedimientos y reglas básicas (respetar las opiniones de otros, asegurar la participación compartida del grupo y finalizar cada pregunta antes de pasar a la siguiente). Se destaca que no hay respuestas correctas ni incorrectas. Adicionalmente, es necesario asegurar a los participantes que no se usarán sus nombres o cualquier otra forma de información que los identifique, y que la información individual no será compartida con nadie.

*Desarrollo.* El moderador usará la guía de discusión para realizar el planteamiento de las preguntas establecidas, aunque es importante ser flexible si el orden de éstas cambia o si se termina el tiempo. Se debe estar preparado para plantear las preguntas clave, antes de que se concluya el tiempo del grupo de enfoque.

*Cierre.* Agradezca la participación de los asistentes y enfatice nuevamente la confidencialidad que se dará a los resultados de la discusión que se tuvo.

9. *Compilar y organizar los datos.*

---

## Listas de cotejo

- Lista de cotejo del paso 8. Evaluación de la métrica del instrumento (documento LC\_F4\_P8\_P0\_000.docx).
- Lista de cotejo del paso 9. Resultados de las puntuaciones obtenidas a partir del instrumento (documento LC\_F4\_P9\_P0\_000.docx).
- Lista de cotejo para el análisis de resultados en el enfoque cualitativo (documento LCAC\_PABC\_P0\_000.docx).
- Lista de cotejo general (adjunta al final de este trabajo y en el documento LCG\_F4\_P0\_000.docx).

## Lista de Cotejo General. Fase 4. Análisis de resultados del instrumento de evaluación. Paso 8. Evaluación de la métrica del instrumento

### I. Datos de identificación del PROEME.

#### Instrucciones:

1. Registre los datos de identificación del proyecto.
2. Seleccione con X la celda de categoría y tipo educativo al que corresponde el PROEME:

**Entidad:** \_\_\_\_\_

**Nombre del PROEME:** \_\_\_\_\_

Categoría	Tipo Educativo		
	Educación básica (EB)	Educación media superior (MS)	Educación obligatoria (EO)
C1. Currículo, materiales y métodos educativos			
C2. Organización escolar y Gestión del aprendizaje			
C3. Condiciones de la oferta educativa			
C4. Docentes, directivos, supervisores y ATP			
C5. Logro educativo			
C6. Políticas, programas y sistemas de información			

**II. Entrega de las evidencias establecidas en los Criterios Técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación.**

**Instrucciones:**

1. Seleccione la columna según corresponda, de acuerdo a la entrega de evidencias de cada paso.
2. En caso de no entregar justifique, en la columna de observaciones, la razón de su decisión y si no aplica a su tipo de evaluación registre NA.

Evidencias	Sí	No	Observaciones
<b>Paso 8. Evaluación de la métrica del instrumento</b>			
1. Ficha técnica 21. Protocolo de lectura.			
2. Ficha técnica 22. Protocolo para el análisis de la métrica.			
3. Ficha técnica 23. Resultados del análisis.			
<b>Paso 9. Resultados de las puntuaciones obtenidas a partir del instrumento</b>			
4. Ficha técnica 24. Protocolo de puntuación.			
5. Ficha técnica 25. Resultado de la evaluación.			
6. Ficha técnica 26. Protocolo de escalamiento.			
7. Ficha técnica 27. Protocolo para establecer niveles de desempeño.			
8. Ficha técnica 28. Protocolo de escalamiento de resultados.			

**Observaciones generales:** \_\_\_\_\_

---



---

## DIRECTORIO

### **JUNTA DE GOBIERNO**

Teresa Bracho González  
CONSEJERA PRESIDENTA

Bernardo Naranjo Piñera  
CONSEJERO

Sylvia Schmelkes del Valle  
CONSEJERA

Patricia Vázquez del Mercado Herrera  
CONSEJERA

### **UNIDADES ADMINISTRATIVAS**

Miguel Ángel de Jesús López Reyes  
UNIDAD DE ADMINISTRACIÓN

Jorge Antonio Hernández Uralde  
UNIDAD DE EVALUACIÓN DEL SISTEMA EDUCATIVO NACIONAL

Rolando Erick Magaña Rodríguez (encargado)  
UNIDAD DE INFORMACIÓN Y FOMENTO DE LA CULTURA DE LA EVALUACIÓN

Francisco Miranda López  
UNIDAD DE NORMATIVIDAD Y POLÍTICA EDUCATIVA

José Roberto Cubas Carlín  
COORDINACIÓN DE DIRECCIONES DEL INEE EN LAS ENTIDADES FEDERATIVAS

Tomislav Lendo Fuentes  
COORDINACIÓN EJECUTIVA DE LA JUNTA DE GOBIERNO

José de la Luz Dávalos (encargado)  
ÓRGANO INTERNO DE CONTROL

Dirección General de Difusión y Fomento de la Cultura de la Evaluación  
José Luis Gutiérrez Espíndola

Dirección de Difusión y Publicaciones  
Blanca Gayosso Sánchez

