

8

Cuadernillo técnico
de evaluación educativa

Análisis y uso de resultados

8

Cuadernillo técnico
de evaluación educativa

Análisis y uso de resultados



Centro UC
Medición - MIDE



Textos de
divulgación

INEE

Instituto Nacional para la
Evaluación de la Educación
México

Análisis y uso de resultados

© Centro de Medición MIDE UC

Av. Vicuña Mackenna 4860
Macul, Santiago, Chile, cp 7820436

© Instituto Nacional para la Evaluación de la Educación INEE

Barranca del Muerto 341, col. San José Insurgentes,
Alcaldía Benito Juárez, Ciudad de México, cp 03900

Autores

Alex Torres Peña, MIDE UC
Johana Contreras Contreras, MIDE UC

Editora

María Rosa García González, MIDE UC

Corrección de estilo

Arturo Cosme Valadez, INEE
Lissette Sepúlveda Cepeda, MIDE UC

Coordinación General

Adriana Guadalupe Aragón Díaz, INEE
Marcela Cuevas Ossandón, MIDE UC
Marcela Ramírez Jordán, INEE

Diseño

www.iunta.cl

Índice

Presentación	1
Resumen	2
Introducción	3
Relevancia del uso y análisis de datos y su relación con la validez del proceso evaluativo	4
Nociones generales sobre análisis de datos	5
Características específicas de las técnicas de análisis para un plan de análisis cualitativo y/o cuantitativo	10
La interpretación y reflexión en el ciclo de uso de resultados evaluativos	26
Consideraciones finales: ideas fuerza	36
Referencias	37

Presentación

El Instituto Nacional para la Evaluación de la Educación de México, INEE, y el Centro de Medición MIDE UC, de la Pontificia Universidad Católica de Chile, han gestado una colaboración para el desarrollo y fortalecimiento de capacidades en evaluación educativa, en profesionales del Instituto y de los equipos responsables de los Programas Estatales de Evaluación y Mejora Educativa (PROEME) y del Proyecto Nacional de Evaluación y Mejora Educativa de Escuelas Multigrado (PRONAEME), en el marco del Sistema Nacional de Evaluación Educativa (SNEE), en México.

El documento que a continuación presentamos constituye un material de consulta que forma parte de una serie de nueve cuadernillos, cuyo propósito es orientar la comprensión de los conceptos centrales de la medición y la evaluación educativas y su impacto en el diseño de instrumentos; considerando que el proceso evaluativo es una suma de decisiones que deben cuidar la coherencia de cada uno de los elementos y fases que lo componen.

Este material se ha organizado en una serie de cuadernillos con base en las siguientes temáticas:

1. Nociones básicas en medición y evaluación en el contexto educativo.
2. Confiabilidad, validez e imparcialidad en evaluación educativa.
3. Definición del marco de referencia de la evaluación.
4. Desarrollo de instrumentos de evaluación: pruebas.
5. Desarrollo de instrumentos de evaluación: cuestionarios.
6. Desarrollo de instrumentos de evaluación: pautas de observación.
7. Desarrollo de instrumentos de evaluación: tareas de desempeño y rúbricas.
8. Análisis y uso de resultados.
9. Uso de resultados y retroalimentación.

Esperamos que este material resulte de utilidad para los profesionales que se desempeñan en el contexto de la medición y evaluación educacional. En los cuadernillos encontrarán nociones y conceptos fundamentales, además de recomendaciones prácticas, y sugerencias bibliográficas para quienes deseen profundizar en cada una de las temáticas trabajadas.

Análisis y uso de resultados

Resumen

Los distintos actores del sistema educativo se encuentran expuestos, actualmente, a un volumen considerable de información evaluativa y, por lo tanto, es importante que dispongan de las competencias necesarias para interpretarla y usarla efectivamente para la toma de decisiones orientada a la mejora de los procesos educativos.

En este contexto, el presente cuadernillo intenta ser una contribución al proceso de alfabetización y acercamiento a los datos que están en la base de las evaluaciones educacionales. Con tal fin, proponemos un breve recorrido sobre la importancia del análisis y usos de los datos, con especial énfasis en la noción de validez y en su evolución hacia un mayor aprecio de los usos y consecuencias de las evaluaciones.

Luego, se introduce al lector en los temas de cómo se estructura un análisis de datos, mediante la figura del plan de análisis, mostrando los tipos de datos que es posible producir mediante la evaluación y sus características específicas, detallando componentes y criterios de calidad para un buen plan de análisis, y deteniéndose en los métodos de análisis de datos cuantitativos y cualitativos, a nivel descriptivo.

Finalmente, se aborda la idea de cómo hacer uso de los hallazgos producidos, generando distinciones entre los distintos tipos de usos y destacando aquellos orientados a mejorar un programa o institución. Se enfatiza en la lectura e interpretación de los datos como un proceso esencial en el ciclo evaluativo destacando las ventajas de privilegiar el uso formativo de las evaluaciones.

Introducción

El presente cuadernillo aborda los usos y análisis de datos de las evaluaciones, un tema imprescindible en la medición educacional, que ha adquirido una creciente consideración en el ámbito académico, en la política pública y -aunque en menor medida- en las comunidades educativas.

En un contexto de expansión de las evaluaciones educacionales, a distintos niveles y con diversos grados de sofisticación de técnicas y métodos de análisis, los actores del sistema educativo se encuentran expuestos a un volumen considerable de información evaluativa. Sin embargo, ello no implica necesariamente un mayor conocimiento de los fenómenos y realidades educativas. Para que se logre producir este y se use con fines de mejora de los centros escolares, de las prácticas pedagógicas y de los aprendizajes de los estudiantes, es preciso recorrer un camino bastante más largo. Dicho de otro modo, entre la producción de los datos y su uso efectivo, existe una serie de mediaciones que van desde la difusión de los resultados, hasta la interpretación de los puntajes, su comprensión, integración, reflexión y toma de decisiones.

En consecuencia, desarrolladores de evaluaciones y usuarios de la información deben asumir responsabilidades en torno a la interpretación y uso de los datos, pero también, en particular los usuarios, necesitan desarrollar competencias que les permitan interpretarlos y comprenderlos.

Una línea de investigación y acción en el ámbito educativo, llamada toma de decisiones basada en evidencia, sugiere que para enfrentar los desafíos que imponen estos cambios se requiere generar capacidades y competencias en los equipos pedagógicos y de gestión, tanto para la lectura y comprensión de los datos derivados de las evaluaciones (Lai y Schildkamp, 2013; Mandinach, 2012; Mons, 2009) como para la generación de condiciones organizacionales propicias para el uso efectivo de los datos.

Por esta razón algunos autores sugieren que la primera etapa en el uso formativo de las evaluaciones y en la utilización de datos para la toma de decisiones, es la alfabetización orientada a la adquisición de competencias de lectura e interpretación de resultados y reportes de evaluaciones (Mandinach, 2012).

El material que se presenta en esta ocasión aspira, en cierta medida, a constituir una herramienta de alfabetización en el análisis y lectura de los datos, dirigida a usuarios de información evaluativa de distinta naturaleza.

Los objetivos de este cuadernillo son:

- Relevar la importancia del análisis y uso de los datos en el contexto de las evaluaciones educativas.
- Exponer las principales etapas del diseño y componentes de un plan de análisis en la medición educativa.
- Distinguir las interpretaciones y los usos posibles de los resultados de las evaluaciones educativas.

Para lograr dichos objetivos, se propone un recorrido que se estructura en torno a tres apartados. En primer lugar, se argumenta sobre la importancia del análisis y usos de los datos, con especial énfasis en las evaluaciones en el plano académico y en cómo se ha ampliado la noción de validez hacia una mayor consideración de los usos y consecuencias de las evaluaciones. En segundo lugar, se exponen los diversos elementos a considerar para el análisis de los datos generados en las mediciones educativas. Se mencionan las especificaciones del plan de análisis, los métodos, técnicas y sus principales componentes. En tercer lugar, se exponen los tipos de usos de las evaluaciones, así como las condiciones cognitivas necesarias para la interpretación de la información.

I. Relevancia del uso y análisis de datos y su relación con la validez del proceso evaluativo

En primer lugar, ¿por qué los usos y análisis de datos se han convertido en un tema ineludible en el ámbito de las evaluaciones educativas? Porque la evaluación no termina al momento de construir, aplicar o, incluso puntuar un instrumento de medición. La evaluación es un ciclo más extenso que incluye un juicio respecto de los datos obtenidos. Luego, es necesario que esos resultados se lean, interpreten, comprendan y usen para la reflexión y/o la toma de decisiones. Su punto final es el uso de la información como retroalimentación para la mejora de procesos educativos, en los distintos niveles de toma de decisiones, que van desde el aula hasta la evaluación de programas y políticas públicas.

En segundo lugar, ¿en qué sentido los análisis y usos de datos han adquirido una relevancia creciente en la medición y evaluación educativa? En el ámbito académico, este proceso se ha hecho visible a través de la evolución del concepto de **validez**, el cual ha pasado de ser entendido como una propiedad de los instrumentos de medición (de la forma en que están contruidos, de su capacidad para mensurar el constructo que pretenden evaluar) a ser definido como una relación entre los instrumentos y sus usos

o, dicho de otro modo, la coherencia entre el propósito para el cual fue diseñada una evaluación y las interpretaciones, usos y consecuencias que se derivan de ella.

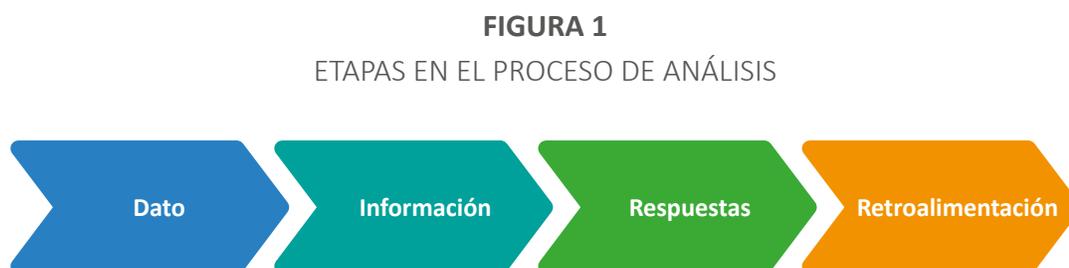
De acuerdo a la convención internacional sobre la medición educacional, materializada en los **Estándares para la medición educativa y psicológica**, los usos son parte central de un instrumento de medición: no existe un instrumento separado de un uso determinado y sus puntajes deben ser interpretados a la luz de sus usos promovidos (American Educational Research Association [AERA], American Psychological Association [APA] & National Council on Measurement in Education [NCME], 2018). La validez se entiende como el “grado en que la evidencia y la teoría respaldan las interpretaciones de los puntajes de una prueba para usos propuestos de las pruebas” (AERA, APA y NCME, 2018, p. 11).

De esta forma, un instrumento de medición es inseparable de su uso, y un uso adecuado solo podrá serlo en la medida en que sea coherente con el propósito de la evaluación y repose en información producida mediante análisis robustos técnicamente, de aquí la relevancia del análisis de datos.

II. Nociones generales sobre análisis de datos

¿Cómo se estructura el análisis de datos?

Con la mera aplicación de los instrumentos no obtenemos en rigor ninguna respuesta; por esto, luego de las definiciones acerca de cómo se producirán los datos necesarios, es preciso determinar qué se hará con ellos en términos de análisis, lo que se muestra en la figura 1.



Al aplicar los instrumentos obtenemos **datos**, los cuales no son interpretables, pues serían, por ejemplo, el conjunto de respuestas de los individuos a una encuesta, o las transcripciones de las entrevistas de cada participante. Estos datos son sometidos a distintos tipos de análisis, transformándose en **información** (Lind, Marchal y Wathen, 2015); sin embargo, la información obtenida puede ser numerosa, proveniente de distintas fuentes, e incluso contradictoria. Cuando se filtra y se pone en relación con el objeto de evaluación, obtenemos **respuestas**. Y una respuesta, acerca de un acierto o un desacierto, comunicada de cierta manera específica, constituye **retroalimentación**.

Hay que pensar este proceso como un mapa de ruta que nos dirigirá desde los datos en su estado bruto hacia la producción de hallazgos significativos para la evaluación de nuestro objeto; lo que llamaremos **plan de análisis** y es una herramienta importante de un proceso de evaluación.

¿Qué es un plan de análisis?

Un plan de análisis es un **documento** que reúne el **conjunto de decisiones** tomadas en torno a las distintas etapas del tratamiento de los datos. Es necesario recalcar que el plan de análisis **si no está escrito, no existe**, ya que debe servir como herramienta de consulta y de socialización de lo que se piensa hacer, para ser fiscalizado, y dividir y delegar tareas, entre otros cometidos.

Determinar el **conjunto de decisiones** implica anticipar el modo y las técnicas con que serán analizados los datos. En ese sentido, es un proceso planificado cuidadosamente. No debe sucedernos que no se produzcan los datos necesarios para responder a los indicadores. Complementariamente, tampoco debe pasar que los datos sean analizados de una forma que no permita responder a las necesidades iniciales.

De esta manera, la realización del plan de análisis implica:

- Un regreso a los indicadores, para cuestionarnos cómo pueden ser respondidos.
- Un regreso a las técnicas de producción de información, ya que debe existir congruencia con lo realizado.
- Una propuesta de metodologías de análisis (cuantitativa o cualitativa) y sus técnicas de análisis de datos asociadas. Esto supone un conocimiento mínimo del tipo de respuestas que nos entregan dichas técnicas. De esta manera, el plan de análisis

determina la forma que tendrá la información producida y, por tanto, las respuestas que obtendremos.

¿Que contiene un plan de análisis?

De forma general, lo siguiente:

1. **Qué se desea obtener o responder:** Dado que se trata de un documento de trabajo, es importante registrar todos los elementos que permitan entender la propuesta de análisis. Por ello, lo primero que conviene señalar son los objetos de evaluación y su indicadores.
2. A través de qué **variables o fenómenos** en particular se mide lo anterior: Es importante acotar específicamente las variables que se considerarán a la hora de medir. En el caso de que sea una evaluación con métodos cualitativos, basta con definir el fenómeno, ya que no constituye una variable numeralizable.
3. Los **instrumentos o fuente de los datos** utilizados: Se debe señalar cuáles son los instrumentos creados específicamente para la evaluación o escogidos entre los existentes para dar cuenta de las variables especificadas de un modo ideal, no solo se mencionan, sino que se adjuntan el test y las preguntas a utilizar.

También existe la posibilidad de usar datos previamente producidos, por ejemplo, pruebas de conocimiento ya realizadas localmente, pruebas aplicadas por entes gubernamentales (como PLANEA), o registros de distintas instituciones especializadas (por ejemplo, obtener el número de consultas por cierta enfermedad en los centros de salud locales); esta modalidad se llama **uso de datos secundarios**.
4. **Qué técnicas de análisis de datos** se ocuparán: Luego, se determina sobre qué variables o conjuntos de variables se empleará cada una de ellas. Por ejemplo, si se realizarán análisis descriptivos y/o inferenciales en el caso de utilizar metodologías cuantitativas. También se explicita el tipo de gráficos que se usará para visualizar la información.
5. **Criterios para cumplir metas:** En el caso de que se conozca o se haya determinado el nivel que se debe alcanzar para cumplir una meta en una determinada variable, se debe dejar consignado. Por ejemplo, "al menos el 50% de los niños y niñas evaluados debe alcanzar treinta puntos o más en el test de autoestima".

¿Qué caracteriza un buen plan de análisis?

Un buen plan de análisis (Secretaría Distrital de Salud de Bogotá, 2013) debe ser:

- **Específico:** Incluir toda la información y técnicas de análisis que se requieren para entender el fenómeno y/o problemática de interés.
- **Eficiente:** Permitir que se analice (e idealmente se produzca) solamente información útil; que no se pierda tiempo recopilando ni analizando información ajena a los objetivos del plan de análisis, que luego no se utilizará. Tampoco es recomendable analizar “todo contra todo” con el fin de encontrar algún resultado.
- **Factible:** Los datos que se plantea analizar (cuantitativa y/o cualitativamente) deben ser asequibles y estar disponibles en los plazos que se requieren para llevar a cabo los análisis (en especial si la evaluación es producida por terceros o por instituciones externas).
- **Estandarizado:** Utilizar tipos de análisis conocidos, cuyos pasos sean repetibles por otros, que puedan conversar con otras evaluaciones, y sean comunicables.

¿Qué tipos de datos hay?

Los datos que podamos producir pertenecen a dos grandes tipos: cuantitativos y cualitativos. Cada uno de ellos permite acceder y reflejar distintos aspectos de la realidad evaluada y se asocia a una metodología distinta de análisis, que contiene técnicas específicas para llevar a cabo su cometido (Canales, 2006).

Una metodología es una forma de reducir la complejidad estudiada, nos orienta a propósito de cómo hacemos foco y construimos conocimiento a partir de muchos datos inicialmente inconexos, los cuales asociamos a través de relaciones de semejanza, diferencia, causalidad, etcétera. Una determinada evaluación puede requerir de una metodología o de ambas simultáneamente y, por lo tanto, su plan de análisis debe reflejar esta necesidad (Krause, 1995).

Datos cuantitativos

Son datos altamente estructurados, suelen provenir de formatos de respuesta cerrada, en que las opciones de respuesta están limitadas (Canales, 2006). Por ejemplo, atributos de la realidad resumidos en números (edad), en etiquetas que se pueden contar (cantidad de mujeres u hombres en un grupo), ocurrencias de un evento en un período observado, etcétera.

Estos datos se transforman en información útil a través de “medidas de resumen”: Las edades, notas, desempeños, etcétera, de muchos individuos o casos se condensan en unos pocos números que delimitan el comportamiento o características del grupo. Además, bajo ciertas condiciones, sus técnicas de análisis permiten establecer “diferencias estadísticamente significativas” (Lind, Marchal y Wathen, 2015).

Datos cualitativos

Son datos de carácter inestructurado que provienen de situaciones e instrumentos con formatos de respuesta abierta, como narraciones y descripciones individuales o grupales, observaciones (sin pautas que los transformen en números), etcétera. La complejidad de los datos se reduce mediante distintas técnicas, la más común es sintetizar el contenido, generando categorías y conceptos a partir de los datos (Canales, 2006; Flick, 2004).

La pregunta general a la que responde cada metodología de análisis se muestra en la tabla 1.

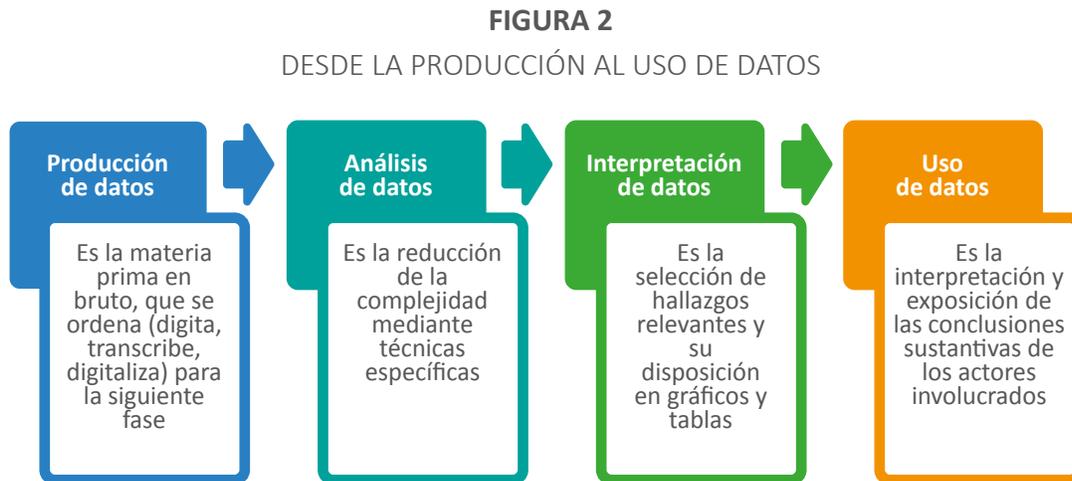
TABLA 1
PREGUNTAS ASOCIADAS CON METODOLOGÍAS CUANTITATIVAS Y CUALITATIVAS

	Cuantitativas	Cualitativas
Pregunta global	¿Cuánto es u ocurre una determinada cosa?	¿Cómo es algo?
Derivaciones comunes	<ul style="list-style-type: none"> • ¿Cuánto es...? • ¿Cuánto cambia (aumenta o disminuye)? • ¿Cuánto se asocia a con b? • ¿Cuánto influye/causa una cosa en otra? 	<ul style="list-style-type: none"> • ¿De qué tipo es...? • ¿Cuáles/cómo son sus procesos? • ¿Cuáles son sus causas (atribuidas)? • ¿Cuáles son sus consecuencias (atribuidas)? • ¿Cuáles son las estrategias de las personas al respecto?

Fuente: elaboración propia.

¿Qué significa analizar los datos?

Esta tarea consiste en reducir los datos para producir información que tenga sentido y sea manejable por los actores interesados. El proceso que siguen los datos se describe en la figura 2.



III. Características específicas de las técnicas de análisis para un plan de análisis cualitativo y/o cuantitativo

El plan de análisis es transversal a la metodología y tipo de datos que se contempla usar. Sin embargo, dadas las diferencias en el tratamiento de los datos y sus respectivos análisis, describiremos por separado las particularidades de un plan de análisis cuantitativo y las de uno cualitativo.

En particular, nos situamos dentro del cuarto componente que recomendamos para un plan de análisis: **técnicas de análisis**. Al describir sus pasos y tareas, haremos especial hincapié en el análisis descriptivo de cada metodología.

Plan de análisis para lo cuantitativo

Las técnicas de análisis son variadas, altamente específicas y dependen fundamentalmente del nivel de medición de las variables que nos interesa trabajar. El nivel de medición es la definición de la forma en que se medirá un determinado atributo de nuestro objeto (Lind, Marchal y Wathen, 2015). Algunos atributos pueden ser medidos en más de una forma, mientras que otros solo tienen una manera correcta. Por ejemplo, ¿cómo podríamos registrar la edad (atributo) de nuestros participantes? Veamos la tabla 2.

TABLA 2
EJEMPLO DE REGISTRO DE LA VARIABLE EDAD

Opción A	Opción B	Opción C
Con categorías: Menor de edad / Mayor de edad	Con intervalos: 0 a 9 / 10 a 19 / 20 a 29 / 30 a 39 / 40 a 49 / etc.	Con la edad en años: 1, 2, 3, 4..., 50, 51..., 90, etc.

En algunos casos, ciertas formas de medir se pueden transformar en otras, por ejemplo, desde la opción C podemos deducir las opciones A y B, ¡pero no al revés! Las distintas maneras de medir un atributo se denominan **variables categóricas, ordinales y cuantitativas** (ver tabla 3). Es importante registrar la que se elija en el plan de análisis, con el fin de entender la técnica de análisis que se propone.

TABLA 3
TIPOS DE VARIABLES

Tipo de variable	Descripción	Ejemplo
Categórica	La variable se estudia para distinguir categorías entre sí, no es relevante o posible medir su cantidad o magnitud, sino su definición o presencia.	El sexo de las personas o el estado donde viven: no puede ser numeralizado y, por lo tanto, no se puede sumar ni promediar el atributo, aunque se puede contar su ocurrencia o su proporción respecto del total.
Ordinal	El atributo se expresa en una escala discontinua, pero representa un orden de magnitud de la variable, es decir, se puede ordenar de alguna manera.	La satisfacción laboral de los docentes puede ser medida y expresada en los niveles de: insatisfactorio, aceptable, satisfactorio y excelente; donde sabemos que satisfactorio es mejor que aceptable, pero no cuánto mejor.
Cuantitativa	El atributo se trata como un continuo y se expresa en una escala numérica.	La edad medida en años, la cantidad de horas de televisión que ve una familia.

Muchas veces, en un plan de análisis se señalan solo las técnicas que producirán los resultados (análisis descriptivos e inferenciales). Sin embargo, es recomendable explicitar cómo se tratarán los datos inicialmente (pasos 1 y 2). Considerando eso, las etapas se muestran en la figura 3.

FIGURA 3

ETAPAS DEL ANÁLISIS DE DATOS EN UN PLAN CUANTITATIVO



Ordenamiento de los datos

En esta etapa se registran los procedimientos mediante los cuales los distintos instrumentos, aplicados en papel o forma digital, deben ser convertidos en datos usables. Dentro de las especificaciones más comunes están:

- Los datos deben estar digitados en tablas o bases de datos (por ejemplo, una planilla Excel o archivo similar) para ser analizados posteriormente. Algunas preguntas a resolver son: ¿quiénes digitarán?, ¿seguirán algunas reglas o protocolos?, ¿dónde se almacenarán los archivos con los datos?
- Se debe definir qué valores se considerarán válidos y cuáles no válidos para las distintas variables.
- Cada evaluado (sea persona, escuela, etc.) debe tener un identificador único que permita reconocerlo en distintos instrumentos y evitar confusiones con otros datos.

Validación de los datos

En esta etapa se señalan los procedimientos mediante los cuales se revisarán los datos producidos o recibidos. ¡Revisa tus bases de datos! Si no estamos seguros de la calidad del insumo, los resultados de los análisis y las conclusiones extraídas podrían ser completamente erróneas.

Dentro de las validaciones que pueden ser consignadas se deben señalar nuevamente el qué y el cómo de nuestros procedimientos (ver tabla 4).

TABLA 4
VALIDACIONES A REALIZAR A LOS DATOS PRODUCIDOS

¿Qué revisar?	¿Cómo?
¿Están todas las variables necesarias?	Revisión visual. Se explora la base de datos, identificando la presencia de las variables relevantes tanto de identificación de los casos como de respuestas de los participantes.
¿Los datos están presentes o perdidos?	Frecuencias de cada variable. Es importante saber si la base de datos presenta respuestas completas o parciales de los casos, y en qué magnitud. Un sujeto que no respondió nada o una pregunta que no fue respondida por nadie pueden ser situaciones problemáticas.
¿Los datos están en el rango esperado?	Frecuencia de cada variable. Es importante identificar que en cada variable solo están los valores que corresponden, errores en esto se pueden producir por problemas en la digitación de los datos.
¿Se tienen casos únicos o repetidos?	Frecuencia de la variable de identificación, revisando si los identificadores están duplicados. En general, solo es válida una respuesta por cada participante.

Análisis descriptivo de los datos

Para cada fenómeno que hemos decidido medir, se debe escoger la mejor forma de describirlo. Los análisis descriptivos son métodos numéricos que se utilizan para resumir y procesar los datos, y transformarlos en información (Lind, Marchal y Wathen, 2015). En este sentido, la descripción no es solo un paso intermedio para transformar los datos en información, puede comprender información suficiente para responder las preguntas formuladas.

El propósito es reducir la cantidad de elementos irrelevantes y particulares, conservando elementos centrales. Así, el resultado de un análisis descriptivo puede ser un valor que resume distintas mediciones de una misma persona, o que condense la información de muchos sujetos en un puntaje que represente el desempeño del grupo.

Los análisis descriptivos son válidos específicamente para el grupo o caso en que se midió, no pueden ser extendidos a poblaciones (Canales, 2016). Muchas veces no se requiere más que una buena selección de análisis descriptivos para caracterizar a un grupo. A continuación, detallaremos los análisis descriptivos más usuales.

a) Frecuencias y porcentajes

La frecuencia es el conteo de las veces que ocurre un determinado valor en el conjunto de casos. Se representa en una **tabla de frecuencias**. Cuando se divide en el total de casos se obtiene una **proporción** (Lind, Marchal y Wathen, 2015). Si, a su vez, la proporción se multiplica por cien hablamos de un porcentaje. Por ejemplo, si tenemos una base de datos de 380 casos y en ella tenemos la variable sexo, la frecuencia indica que 230 de los participantes son mujeres y 150 son hombres, como se observa en la tabla del ejemplo a continuación. En las columnas siguientes podemos observar la proporción y porcentaje asociados con dichas frecuencias.

Ejemplo de tabla de frecuencias			
Sexo	Frecuencia	Proporción	Porcentaje
Mujer	230	0,61	61%
Hombre	150	0,39	39%
Total	380	1	100%

b) Medidas de tendencia central

Permiten conocer dónde está **el centro** de un conjunto de datos, según distintos criterios (ver tabla 5).

TABLA 5
DESCRIPCIÓN DE LAS MEDIDAS DE TENDENCIA CENTRAL

Medida	Descripción
Moda	Corresponde al valor que se repite la mayor cantidad de veces en un conjunto de datos.
Mediana	Es el valor central de un conjunto de datos que están ordenados de menor a mayor.
Media o promedio	Corresponde, en un conjunto finito de números, a la suma de todos sus valores dividida entre el número de sumandos. Es sensible a los valores extremos, sobre todo en casos de pocas observaciones.

A continuación, supongamos que tenemos el dato de la edad de nueve personas:

Ejemplo de medidas de tendencia central de una serie de datos	
23] La moda es 23, es el valor que más se repite
23	
25] La mediana es 27, el cuarto en la serie de siete valores, es el valor central
27	
35	
38	Promedio del grupo es de 36,5 años, ¡la persona adulta mayor (85 años) eleva el promedio!
85	

c) Medidas de dispersión

Además de conocer el valor del **centro** de un conjunto de datos, también necesitamos describir cuán diferentes son los valores entre sí, es decir, su variabilidad o dispersión (ver tabla 6).

TABLA 6
DESCRIPCIÓN DE LAS MEDIDAS DE DISPERSIÓN

Medida de dispersión	Descripción
Mínimo	Valor mínimo de la variable en el conjunto de datos.
Máximo	Valor máximo de la variable en el conjunto de datos.
Rango	Es la diferencia (resta) entre el valor máximo y el valor mínimo de la variable en el conjunto de datos.
Desviación estándar	Una medida de cuánto, en promedio, se alejan los valores del promedio del grupo.

A continuación, supongamos que tenemos el dato de la edad de nueve personas:

Ejemplo de medidas de dispersión de una serie de datos	
23	→ Mínimo
23	
25	El rango es de 62 años, hay una importante dispersión de valores
27	
35	
38	
85	→ Máximo

Una medida de dispersión muy relevante es la **desviación estándar**. Es una medida de la variabilidad de los datos que requiere del promedio para su cálculo. De hecho, permite matizar la información obtenida por el promedio, ya que refleja cuán cercanos o lejanos a este se hallan los puntajes observados.

Una forma de conceptualizar la desviación estándar es “una medida de cuánto, en promedio, se alejan los valores del promedio del grupo”. Actualmente, muchas herramientas de fácil alcance permiten calcular las medidas de dispersión (y las de tendencia central): Excel, Google Spreadsheets, etcétera, por lo que no profundizaremos en su fórmula específica.

El siguiente ejemplo ilustra lo relevante que es considerar el promedio en conjunto con la desviación estándar. Los tres grupos de personas tienen el mismo promedio de edad, pero sus dispersiones son muy distintas.

Ejemplo de desviación estándar de tres grupos de igual promedio					
Grupo A		Grupo B		Grupo C	
23	Promedio = 36,5 años	30	Promedio = 36,5 años	34	Promedio = 36,5 años
23		30		35	
25	Rango = 62	32	Rango = 20	36	Rango = 5 años
27		33		37	
35		38		37	
38	Desviación estándar = 20,5 años	43	Desviación estándar = 7 años	38	Desviación estándar = 1,6 años
85		50		39	

Si solo nos fijamos en el promedio, un error muy común, parece que los tres grupos son iguales, ya que coinciden en 36,5 años... Pero, si a eso agregamos los valores de la dispersión de cada grupo, vemos que son muy distintos. El primer grupo es muy diverso, presenta una desviación estándar de 20,5 años; mientras que el grupo C es muy homogéneo, ya que la edad de los participantes se desvía, en promedio, 1,6 años del promedio de 36,5 años. En consecuencia, ¡no se puede confiar solo en el promedio!

d) Medidas de posición

Son medidas que informan acerca de la ubicación relativa de los evaluados en relación con su grupo de referencia, es decir, con el grupo completo de personas evaluadas. Se llaman según la cantidad de grupos en que dividen al conjunto de casos: quintiles, dividen al grupo en cinco subgrupos; deciles en diez; percentiles en cien, etcétera. Es importante notar que las medidas de posición refieren a los **cortes** y no a los grupos. Es decir, los quintiles que separan al conjunto de casos en cinco grupos son siempre cuatro cortes. Las medidas de posición más utilizadas son las que se muestran en la tabla 7.

TABLA 7
DESCRIPCIÓN DE LAS MEDIDAS DE POSICIÓN

Medida de posición	Descripción
Mediana	Ya fue descrita en las medidas de tendencia central. También es una medida de posición al dividir al conjunto de participantes en dos grupos.
Cuartiles	Cada uno de los tres valores que dividen al conjunto de datos en cuatro grupos, cada uno con 25% de los casos aproximadamente.
Quintiles	Cada uno de los cuatro valores que dividen al conjunto de datos en cinco grupos, cada uno con 20% de los casos aproximadamente.
Percentiles	Cada uno de los 99 valores que dividen al conjunto de datos en cien grupos, cada uno con 1% de los casos aproximadamente.

Los percentiles son un caso especial, ya que, por un lado, requieren una variable que tenga sentido dividir en cien segmentos, y por otro, una cantidad de evaluados no menor para que puedan ocupar dichos espacios. La potencialidad de un percentil estriba en que puede leerse asociado con un porcentaje, por ejemplo: si un participante se ubica en el percentil 34, quiere decir que está por encima del 34% de los sujetos del grupo y por debajo del 66%. Estos porcentajes hacia abajo o arriba especifican una posición de los casos dentro del grupo.

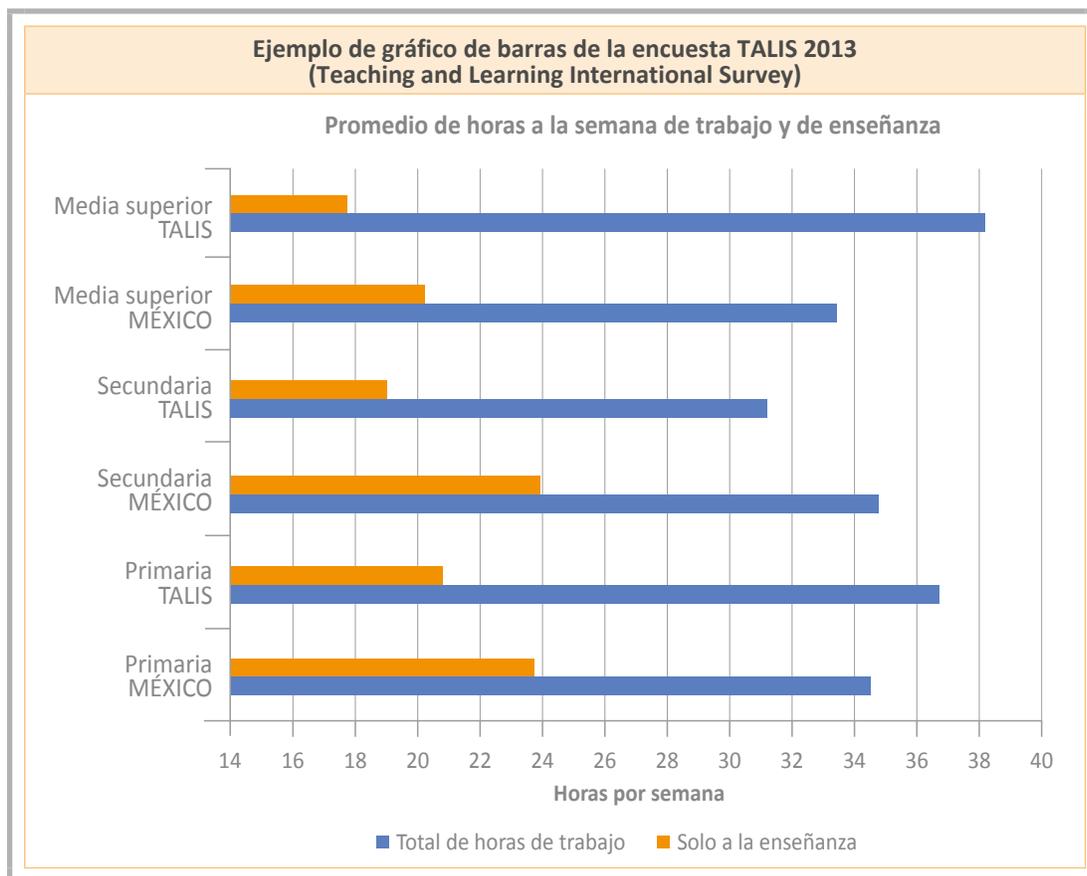
Como se deduce de lo anterior, el puntaje o valor de una determinada variable no tiene un vínculo irrompible con un determinado percentil, sino que depende del grupo en el que se encuentre. Por ejemplo, si una persona corre cien metros en dieciocho segundos ¿Es un buen tiempo o muy bajo? Si lo comparo con atletas que tienen promedio de once segundos, es muy malo (y, probablemente, obtendría percentil 1 o 2) y si lo comparo con un grupo de 75 años es un tiempo excelente (y, probablemente, obtendría percentil 98 o 99).

e) Los gráficos: representaciones visuales

Al igual que en las técnicas de análisis, los gráficos se asocian con el nivel de medición de las variables. Los más comunes son: gráfico de barras o columnas, gráfico circular o de pastel (*pie chart*) e histograma, mismos que describiremos a continuación.

• Gráfico de barras o columnas

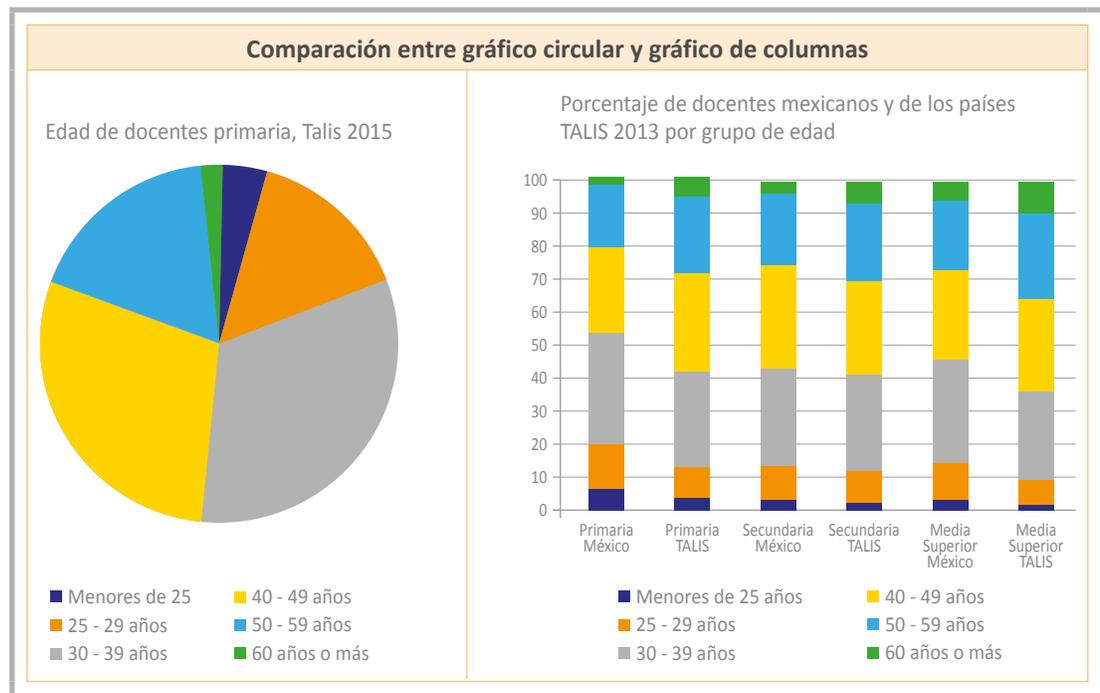
Es un gráfico que permite sintetizar frecuencias (ya sea como conteo o como porcentaje). También se utiliza para graficar el promedio alcanzado por varios grupos simultáneamente, lo que hace más fácil su comparación. Puede estar dispuesto de forma vertical (columnas) u horizontal (barras).



Fuente: Backhoff y Pérez-Morán (2015).

• **Gráfico circular o de pastel (pie chart)**

Es un gráfico usado para mostrar porcentajes de un determinado atributo y enfatiza su proporción dentro del total. Su defecto es que presenta poca información respecto de la cantidad de espacio que ocupa. Cuando se requiere mostrar información de porcentaje de varios grupos simultáneamente, se prefieren gráficos de barras o columnas apilados, donde cada barra representa 100%. En el siguiente ejemplo, se muestran dos gráficos, el circular de la izquierda equivale solo a la primera columna del gráfico de la derecha.



Fuente: elaboración propia.

Fuente: Backhoff & Pérez-Morán (2015).

• **Histograma**

Cuando se quiere mostrar un atributo medido como variable cuantitativa se usan histogramas. Se generan segmentos de igual magnitud, cada uno representado por una columna. La cantidad de segmentos queda a juicio de quien construye el gráfico. El ejemplo a continuación muestra la distribución de los ingresos promedio de México por hogar, dividido en diez segmentos, es decir, en deciles.



Fuente: Moy (2017).

¿Cuándo se usan cada uno de los elementos que hemos visto?

Tantos los análisis como las representaciones gráficas se realizan en función del nivel de medición que tiene nuestra variable. Ahora que ya conocemos el abanico de opciones, presentamos, en la tabla 8, las relaciones entre el nivel de medición y lo que se puede realizar.

TABLA 8
RELACIONES ENTRE NIVEL DE MEDICIÓN, TÉCNICAS Y GRÁFICOS

Nivel de medición	Técnicas que es posible usar	Técnicas que no es posible usar	Gráficos que es posible usar
Datos ordinales y categóricos	Frecuencias. Porcentajes. Moda.	Ni promedio ni desviación estándar. Tampoco medidas de posición.	De barras o de columnas (pueden ser apilados). Circulares (<i>pie chart</i>).
Datos cuantitativos	Son versátiles y permiten utilizar las distintas técnicas de análisis descriptivo. Con promedios y desviación estándar se puede representar de manera práctica la diversidad de este tipo de datos. También es muy relevante el uso de percentiles.	Frecuencias y porcentajes suelen ser menos útiles si la variable cuantitativa adopta muchos valores distintos. Pero, dado que los datos se pueden agrupar a voluntad, también se pueden ocupar.	Preferentemente histograma. No es pertinente el uso de gráficos circulares.

Como se señaló, además de lo que se desea medir y de los instrumentos para ello, hay que declarar los análisis a realizar. En particular, debe indicarse el atributo o variable específico a analizar (desempeño en matemáticas, edad, sexo, etc.) y, para cada uno de los atributos, la o las técnicas descriptivas que son necesarias con el fin de cumplir nuestro objetivo. Asimismo, se debe especificar el apoyo gráfico que se utilizará. Todo tiene que ser congruente con el nivel de medición en que estén las variables (categórico, ordinal, cuantitativo), el cual idealmente también debe puntualizarse.

Análisis inferenciales: algunas consideraciones

Existen ocasiones en que se requieren técnicas que nos permitan ir más allá de la descripción acabada de nuestros datos. En dichas situaciones se habla de **inferir el valor** de una cosa a partir de otra (Lind, Marchal y Wathen, 2015). Por ejemplo:

- a) Cuando queremos conocer el valor de un atributo de una población a través de una muestra; digamos, conocer el nivel de habilidades de lenguaje de un distrito a partir de una muestra de alumnos.
- b) Cuando queremos establecer que ciertas diferencias o asociaciones entre grupos no son producto del azar, si no que corresponden a la realidad.

La posibilidad de establecer esas diferencias o asociaciones no solo dependerá del uso de técnicas inferenciales, sino del diseño de nuestra evaluación. Se debe construir meticulosamente la muestra y cuidar que se cumplan con cuidado las intervenciones.

Plan de análisis para lo cualitativo

El análisis de datos cualitativos se refiere a las técnicas dirigidas a información que no está completamente estructurada. Por esto mismo, es especial para enfocarse en aspectos que se desconocen y requieren ser explorados. El otro fin con el que se suele ocupar este análisis es el de obtener información acerca de cómo piensan los actores involucrados en una determinada situación (Canales, 2006; Flick, 2004). De ahí que la técnica de recolección de datos que más se analiza de esta forma sea la entrevista.

Podemos medir el desempeño en matemáticas de un grupo de alumnos luego de una intervención con un conjunto de ítems específicos. Pero, al indagar con los docentes los elementos que consideran positivos y negativos de esta nueva metodología, un set de preguntas cerradas no será lo más fructífero, ya que probablemente se reducirá a verificar los presupuestos de los evaluadores y no estará abierto a elementos novedosos de la situación.

El plan de análisis para datos cualitativos puede seguir una estructura muy parecida a la expuesta. De la misma manera que en el apartado referido a los datos cuantitativos, al describir sus etapas se hará énfasis en la sección de análisis descriptivo (ver figura 4).

FIGURA 4
ETAPAS DEL ANÁLISIS DE DATOS EN UN PLAN CUALITATIVO



Ordenamiento de datos

El material para análisis cualitativos, suele ser recogido en archivos de audio (e incluso video), ya que lo constituyen principalmente entrevistas o interacciones con los involucrados. También puede tomar la forma de notas de campo o apuntes de lo observado por parte del evaluador. Los datos provenientes de la aplicación de los distintos instrumentos deben ser convertidos en datos usables (Flick, 2004): estar transcritos para ser utilizados en forma práctica y sin perder detalle (lo que no ocurriría si se trabajase directamente con el audio). De la misma manera, cada transcripción debe incluir los datos de identificación de la aplicación: fecha, lugar, institución, persona que realizó la técnica, algún registro breve de si hubo inconvenientes o problemas en su aplicación y persona entrevistada. Esto último se suele realizar a través de un identificador que no revele la identidad de la persona, con el propósito de resguardar su anonimato. Así, la forma como se registrarán los datos y los datos con que se identificará cada aplicación, deben quedar acordados y registrados de antemano.

Procesamiento de datos

La tarea de validación, en este caso, pasa por certificar la correcta transcripción de los datos. Este procesamiento debe ser cubierto en dos niveles: el primero es que, si la tarea de transcribir se encarga a terceros, se debe verificar al menos la primera vez que la transcripción sea fiel al audio, sin tergiversaciones. Igual revisión se debe hacer si hubo problemas para entender los audios y su transcripción no es del todo fidedigna.

El segundo nivel es que la transcripción cumpla con ciertas convenciones de correcto registro: a) deben estar presentes los datos de identificación señalados en el apartado anterior; b) debe diferenciarse lo que fue dicho por los participantes de lo que dicen los evaluadores o entrevistadores al inicio de cada párrafo, ya que no siempre es evidente; y c) cada párrafo debe numerarse, ya que esto sirve como estrategia de trazabilidad y verificación.

Enseguida, se muestra un extracto de transcripción en el que se observan los tres elementos señalados.

Ejemplo de transcripción de entrevista siguiendo convenciones de correcto registro

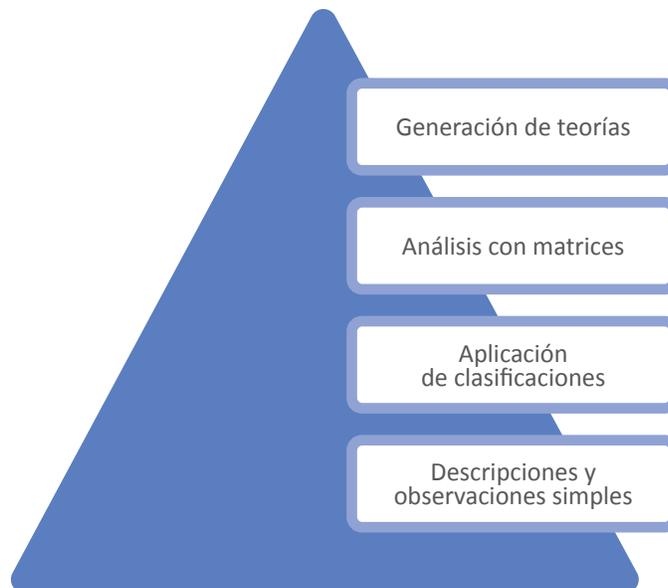
Sesión: 8 Fecha: 24 de octubre de 2002 Pacientes: Javier (P ₁), Cristóbal (P ₂), Cristóther (P ₃), Franco (P ₄), Rodrigo (P ₅) Terapeutas: Cristina (T ₁) y Ulises (T ₂)	
Parrafo	Discurso
1	T ₁ : alguien más?
2	P ₁ : yo ya hablé la vez pasada - - - - - ahí yo yo quiero pedir tiempo - - - - - lo que pasa es que
3	T ₁ : vamos a empezar con P ₁
4	P ₁ : bueno el viernes pasado yo tuve un paseo a Quintay con P ₂ : celebrando el día del profesor - - bueno y hasta las cinco de la tarde estuvo todo perfecto y cuando volví - - en el bus cuando volvía con los profesores y - - cuando íbamos a tomar once cuando llegamos a Santiago eh/h - - en realidad andaba con la idea hace tiempo de 'me voy a tomar una cerveza' - y me tomé la cerveza y' la encontré desagradable - después le eché un poco de fanta y - - no le encontré gracia - y eso me produjo al tiro un aceleramiento así una: - - una compulsión de tomar todo - entonces tomé vino blanco y tampoco le encontré gracia tomé vino tinto un vaso tampoco tomé una piscola +P ₁ : era once+ claro pero estaban todos los profesores tomando
5	P ₁ : están todos tomando?
6	P ₁ : claro - y en el fondo estaba con el grupo de profesores que habíamos estado un rato me agradaba - - o sea que nosotros volvimos en el bus y en el bus: - bueno ahí conversamos hasta Santiago e: seguimos tomando pisco con bebida - pero ya - - decía 'ya se embarró' e: 'sonó' - y como una cuestión como: con todo el sentimiento de culpa de haber perdido todos estos meses - y: no haber

Análisis descriptivo

En el campo de la evaluación, el tipo de análisis usado por excelencia es el análisis de contenido, que se encarga de sistematizar y trabajar lo que las personas efectivamente dicen (y, por tanto, pensarían) sobre un determinado fenómeno. Las diversas vertientes de análisis de contenido oscilan entre aquellas que utilizan categorías prefijadas y las que emplean categorías emergentes/construidas a partir de los datos.

Respecto al análisis de contenido existen variantes sobre cómo realizarlo, con distintas utilidades y potencialidades (ver figura 5).

FIGURA 5
VARIANTES DE ANÁLISIS DE CONTENIDO



a) Descripciones y observaciones simples

Un primer tipo de análisis consiste en la síntesis de los textos y las transcripciones. Básicamente, se toman notas acerca de los documentos en análisis y se detallan los elementos relevantes en torno al tema que se está indagando (Barton y Lazarsfeld, citado en Sánchez-Gómez *et al.*, 2017). No agrega mucho valor respecto de la transcripción, salvo la reducción de material.

b) Aplicación de clasificaciones

Se establecen temas y subtemas relevantes para la indagación. El material producido por las entrevistas y otros medios es sometido a un proceso de clasificación sistemática de su contenido, con base en los temas y subtemas establecidos. Estos se vinculan con segmentos de las entrevistas, lo que permite resumir todo el material a partir de los temas iniciales y tener a mano los ejemplos de las entrevistas para respaldarlo (Barton y Lazarsfeld, citado en Sánchez-Gómez *et al.*, 2017).

c) Análisis con matrices

Similar al anterior, pero en este caso el listado de temas corresponde a un eje de una matriz o cuadrícula, y en el otro eje se encuentran individualizadas las distintas entrevistas. En cada cuadrante de la matriz se ubica de forma condensada lo dicho por ese actor en

ese tema específico. Su utilidad es que permite sistematizar lo que dijo cada entrevistado y comparar sus aportes a lo largo de un mismo tema (Barton y Lazarsfeld, citado en Sánchez-Gómez *et al.*, 2017).

d) Generación de teorías

El esquema de clasificación en temas y subtemas va siendo construido a partir de la lectura del material. No se parte de un esquema previo. El trabajo está muy orientado a la exploración, por lo tanto, es más arduo inicialmente. Luego de identificar y generar los temas y subtemas, se buscan relaciones entre estos, las cuales permiten construir pequeñas teorías acerca del fenómeno evaluado (Barton y Lazarsfeld, citado en Sánchez-Gómez *et al.*, 2017; Strauss y Corbin, 2002).

Codificar: la acción principal del análisis de contenido

Codificar significa identificar pasajes, conceptos o formulaciones clave, que permitan describir el material de las entrevistas. Los datos son fragmentados (se escoge un pasaje relevante), conceptualizados (se les asigna un nombre, una etiqueta conceptual) y se relacionan entre ellos al ser incluidos en categorías. Las unidades básicas de significado consisten en una etiqueta verbal (un nombre) puesto sobre sucesos, eventos y otros aspectos de los fenómenos estudiados, señalados en las transcripciones (Strauss y Corbin, 2002).

Los conceptos se agrupan y clasifican en categorías temáticas de mayor nivel de abstracción; esta clasificación es **descubierta** cuando los conceptos son comparados unos con otros y aparecen como pertenecientes a fenómenos similares.

¿Qué caracteriza un análisis de contenido cualitativo de calidad?

El producto de este tipo de análisis cualitativo es un texto que describe lo encontrado y sistematizado en las categorías; a diferencia del análisis cuantitativo no presenta números ni gráficos, pero puede contener esquemas y, generalmente, hay en él citas textuales de las transcripciones.

Sin embargo, recordemos que un buen análisis no es solo un conjunto de citas, **es una categorización que aporta valor en la medida que reduce la cantidad de datos y genera abstracción** a partir de ellos (Flick, 2004; Krause, 1995). Por lo tanto, no se trata de una repetición textual de lo encontrado: solo se usan citas para ejemplificar y estas provienen de las distintas aplicaciones (y no siempre de la misma entrevista, observación o grupo focal).

El segundo atributo de un análisis de contenido de calidad es que es **trazable**: las citas y conclusiones más importantes pueden ser seguidas hasta el material mismo. Esto se logra **citando** el origen de la información (por ejemplo: E1, P17, nos permitiría revisar la información entregada en la entrevista 1, párrafo 17).

En tercer lugar, el análisis debe **estar orientado hacia los objetivos del estudio** y no se detiene en describir elementos irrelevantes, lo que es un riesgo latente si se entiende que el proceso de análisis es una tarea exploratoria, pero no sin rumbo.

Cuarto, el análisis fue **realizado por un equipo** y no por sujetos de forma individual, dado que el trabajo conjunto sobre una entrevista o una idea presente en el material ayuda a evitar sobreinterpretaciones.

En quinto lugar, la **robustez de los resultados no se determina en función de criterios numéricos**, como, por ejemplo, al decir "75% de los entrevistados señaló esto", ya que, al estar orientados al descubrimiento, algo dicho por un participante puede ser más relevante que una obviedad dicha por todos los entrevistados.

Como se señaló, es preciso especificar la técnica de recolección de información ocupada. En particular, se debe señalar que se trata de un análisis de contenido cualitativo, y qué tipo de análisis de contenido se pretende realizar en función de las necesidades de la evaluación, ya que generar teoría puede ser interesante, pero muchas veces innecesario en el contexto particular. Adicionalmente, si se va a usar un conjunto de temas y subtemas (categorías) preestablecido, es útil señalar cuál sería.

IV. La interpretación y reflexión en el ciclo de uso de resultados evaluativos

Ya tengo la información, ¿da lo mismo cómo la uso?

El empleo de la evaluación puede ser definido como la manera en la cual una evaluación, y la información que de ella emana, impacta en el programa o institución evaluado (Alkin y Taut, 2003). En esta sección veremos que tales usos son diversos en función de su intensidad, nivel de toma de decisiones, juicios de valor sobre la información y propósito de la evaluación. Es necesario, sin embargo, recordar que más allá de la importancia de usar efectivamente la información y de que, idealmente, las evaluaciones se apliquen para la mejora de las instituciones, prácticas y programas, entre otros, el buen o mal empleo de la información evaluativa depende del propósito específico para el

cual se diseñó la evaluación. Por esta razón los **estándares para la medición educativa y psicológica** exhortan a los desarrolladores de evaluaciones a delimitar con precisión los usos e interpretaciones que se desprenden, directa o indirectamente de las evaluaciones que diseñan.

Por supuesto, los límites no siempre son claros. En algunas ocasiones, una prueba diseñada para medir el dominio de un contenido curricular, es utilizada como selección para un programa de estudios universitario o como predicción del desempeño. Una evaluación diseñada con el fin de diagnosticar el nivel de aprendizaje de una cohorte de estudiantes, puede ser utilizada para la evaluación de resultados de una innovación educativa. En consecuencia, no es posible afirmar *a priori* si algunos usos que se acercan bastante al propósito de la prueba, pese a no coincidir exactamente con él, son totalmente desaconejados y en todas las circunstancias. En este sentido, parece ser que la recomendación no radica tanto en lo correcto o incorrecto de un uso en sí mismo, sino en que los usos se encuentren dentro de una zona perímetro de posibilidades definida por quienes diseñan la evaluación, de modo que mientras más alineados estén con ese propósito, más pertinentes serán.

Usos según su intensidad

Los empleos de la información evaluativa pueden ser diversos y, de hecho, en la literatura es posible encontrar distintas clasificaciones o tipologías. Una de las más conocidas, describe cinco tipos de usos (Fleischer y Christie, 2009).

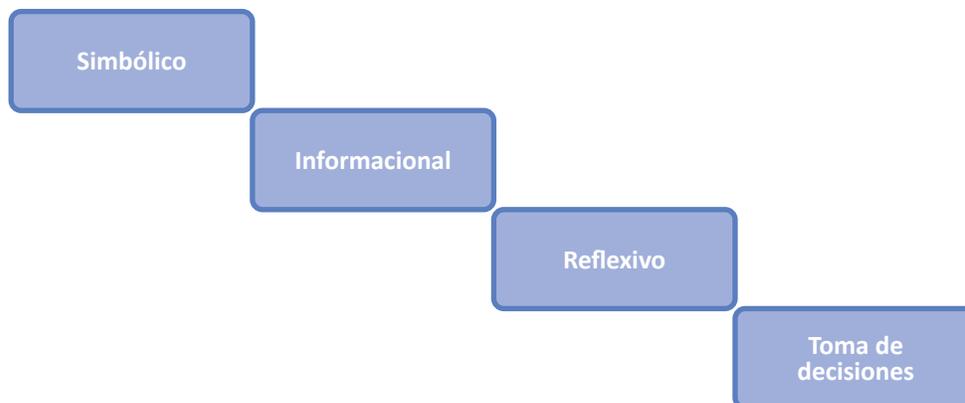
- 1. Instrumental:** Cuando los tomadores de decisiones usan los hallazgos de la evaluación para modificar, de alguna manera, a los evaluados o al objeto de la evaluación.
- 2. Conceptual:** Cuando los hallazgos de la evaluación ayudan a los equipos del programa a entenderlo de una manera distinta, aportando una visión diferente y nueva del programa.
- 3. Contribución a la producción de conocimiento:** Los hallazgos de la evaluación agregan conocimiento al campo y pueden ser usados por todos, no solo por los involucrados del programa.
- 4. Uso de procesos:** Propuesto por Patton *et al.* (1977), es la retroalimentación constante al proceso de implementación del programa. Se define el uso de proceso como los cambios de diseño, organizacionales, comportamentales y cognitivos producto de involucrarse en una evaluación o de pensar evaluativamente el propio desempeño institucional.

5. Uso persuasivo o simbólico: Sobre este último caso no existe consenso. Según algunos el uso es persuasivo o simbólico cuando se usan los resultados de la evaluación para persuadir a inversionistas o contratantes sobre el valor del programa (*accountability*). En una variante menos positiva, es cuando se usan los resultados para legitimar una decisión ya tomada.

Esta tipología de usos permite distinguirlos en términos de su intensidad o de la medida en la cual el uso de la información desemboca en decisiones o acciones concretas de cambio (uso instrumental), o bien, los resultados solo se invocan como una forma de presión sin que medie una reflexión ni incluso una lectura de los datos (uso simbólico). Entre ambos extremos se encuentran los usos que se orientan a la reflexión y aquellos que se limitan a la difusión de resultados a un público amplio (por ejemplo, a las familias de los estudiantes o a las escuelas). La figura 6, presenta una síntesis de los usos, siguiendo este continuo que va desde un *no uso* (inacción) hasta el empleo efectivo de la información para la toma de decisiones (acción).

FIGURA 6

TIPOS DE USOS DE LA INFORMACIÓN EVALUATIVA SEGÚN SU INTENSIDAD



La perspectiva denominada Data Driven Decision Making (DDDM) o toma de decisiones basada en datos, que se ha expandido considerablemente en los últimos años, postula la necesidad de basar las decisiones en evidencia (y no en apreciaciones o intuiciones) y de usar efectivamente los resultados generados por las evaluaciones educativas (Mandinach y Jimerson, 2016). En un contexto internacional en el cual se ha expandido con fuerza la exigencia de rendir cuentas a las escuelas e instituciones y, como corolario, la exposición de los actores del sistema escolar a un volumen creciente de información evaluativa de distinta índole, esta línea de investigación y trabajo ha producido conocimiento respecto de las

condiciones necesarias para el uso efectivo de los datos, en particular, para comprender e integrar la información. Por otro lado, ese uso instrumental no consiste en el simple hecho de tomar decisiones basadas en datos, sin importar cuáles sean, sino en promover aquellas que se orientan a la mejora de los procesos educativos y los aprendizajes de los estudiantes (Lai y Schildkamp, 2013). En tal sentido, la descripción de los usos desde la intensidad, si bien es necesaria, especialmente para distinguir el uso instrumental del “no uso”, no es suficiente. Debe acompañarse de otras clasificaciones que permitan profundizar en la naturaleza y diversidad de esos usos y decisiones.

Usos según el actor evaluado y el nivel de toma de decisiones

El empleo que se pueda dar a los resultados de una evaluación no será igual si los examinados son estudiantes o docentes, individuos o instituciones; tampoco será el mismo dependiendo del nivel de toma de decisiones. Por ejemplo, si la información evaluativa está destinada a un docente, se espera que el uso de la misma se produzca a nivel de aula y que las acciones derivadas de ella se orienten a revisar las prácticas pedagógicas, la planeación de clases, la gestión del grupo, etcétera. En cambio, si los usuarios objetivo de una evaluación son los equipos directivos, se espera que la información retroalimente la gestión escolar, el liderazgo, las intervenciones o proyectos a nivel del establecimiento. Si los usuarios son evaluadores de proyectos o programas a nivel de una entidad federativa, un municipio u otro servicio educativo desconcentrado, los resultados de la evaluación probablemente serán utilizados para el seguimiento o monitoreo de las intervenciones implementadas, juzgar los resultados y la continuidad de un programa o disponer de una visión global de los logros de los estudiantes en una división territorial determinada, entre otros.

El actor que es objeto de evaluación también va a condicionar en gran medida el uso de los resultados. La evaluación docente a menudo se emplea para tomar decisiones relacionadas con la afectación de profesores o guiar acciones de mejora de las prácticas pedagógicas; mientras que la evaluación a estudiantes suele usarse con fines tan diversos como la certificación individual o el seguimiento de programas cuyo fin último es impactar en los aprendizajes.

Usos según el juicio de valor respecto de la información producida

Las interpretaciones y usos de los resultados de las evaluaciones, de acuerdo con la literatura especializada (AERA, APA y NCME, 2018; Kane, 2008, 2016; Nkwake, 2015), deben estar alineados con los puntajes y reportes generados a partir del o los instrumentos de medición. Es la razón por la cual también es probable que el uso dado a la información varíe en función del tipo de datos producidos y de los juicios de valor que se formulan a partir de ellos.

Ravela (2006) distingue al menos tres formas de emitir juicios de valor en las evaluaciones educativas:

- **Enfoque normativo:** Orientado a ordenar a individuos, grupos, instituciones o divisiones territoriales, a la luz del constructo evaluado, con el fin de compararlos entre sí. Este enfoque es particularmente utilizado cuando el propósito de la evaluación es el establecimiento de *rankings* o procesos selectivos que deben determinar con precisión el lugar de los individuos o instituciones en una jerarquía de resultados.
- **Enfoque criterial:** El juicio evaluativo versa sobre la comparación del desempeño de un individuo o grupo con un criterio preestablecido, el cual fija un umbral o nivel de dominio del aspecto evaluado. Se sitúan en este grupo aquellas evaluaciones que establecen niveles de desempeño (por ejemplo, elemental, intermedio, avanzado) asociados con competencias y conocimientos mínimos que es necesario evidenciar para pasar de un nivel al siguiente. En este modelo, se espera conocer cuál es la brecha que separa el desempeño de cada individuo o grupo respecto del criterio impuesto, o bien, cuántos individuos y quiénes se sitúan en cada nivel de desempeño.
- **Enfoque de progreso:** Consiste en el juicio de valor respecto del avance, aprendizaje o cambio de un atributo medido en dos puntos distintos de tiempo. Supone una comparación del individuo, grupos o instituciones desde un punto de vista temporal. Este enfoque no es incompatible con los anteriores y puede dar lugar, además, a un juicio de criterio (cuando se define el progreso como el paso de un nivel de desempeño a otro) o normativo (cuando se comparan posiciones relativas o progresos entre individuos).

¿Qué implicaciones tienen estos juicios de valor en la interpretación y usos de los resultados de las evaluaciones? En primer lugar, los usuarios de información evaluativa lograrán comprenderla, interpretarla y usarla de una manera apropiada (coherente con el propósito de la prueba) si conocen el enfoque evaluativo subyacente, pues sabrán qué esperar o qué tipo de comparaciones cabe realizar a partir de los resultados.

Si la evaluación arroja resultados en una escala de puntaje estándar, que ordena con alta confiabilidad a los individuos de acuerdo a su desempeño, una interpretación pertinente será el juicio respecto de quiénes manifiestan más o menos dominio del atributo medido. Usar esta información como una de las fuentes de selección en un programa de becas de excelencia académica, podría ser pertinente si no se distancia de los propósitos iniciales del instrumento. Por el contrario, si una evaluación con estas características se emplea para determinar el desempeño específico de cada individuo examinado respecto del logro

de un determinado umbral de competencias, las interpretaciones probablemente serán equívocas. Del mismo modo, si lo que se desea es conocer el progreso de los aprendizajes o los cambios derivados de una intervención educativa, no es posible emitir juicios a partir de evaluaciones en un momento del tiempo (especialmente si son normativas).

En síntesis, distinguir entre tipos de juicios de valor es un primer paso para que la lectura e interpretación de los datos sea adecuada y, por lo tanto, sus posibles usos. Es más, como usuarios de evaluaciones nacionales e internacionales, es conveniente que al ver un reporte nos preguntemos, en primer lugar, cuál es el constructo que mide la prueba (no es lo mismo que se mida el dominio del currículo nacional de Ciencias o un concepto amplio, como la cultura científica) y qué enfoque de juicios de valor predomina. A partir de ese cimiento será posible ir construyendo, fehacientemente, las interpretaciones siguientes.

Usos según el propósito de la evaluación

Además de los empleos antes especificados, es importante que interpretaciones y usos sean coherentes con el propósito de la evaluación. Ahora bien, ¿cuáles son los propósitos que pueden tener las evaluaciones? En la teoría de la evaluación y medición, a partir de los trabajos de Scriven (1967), los autores suelen distinguir dos tipos de evaluación a este respecto (Martínez, 2012; Ravela, Picaroni y Loureiro, 2017):

- **Evaluación sumativa (o para la certificación):** Tiene lugar al final de un período de enseñanza y su propósito es calificar o certificar a los examinados.
- **Evaluación formativa:** Se realiza durante el proceso de enseñanza-aprendizaje y su propósito es guiar las estrategias pedagógicas de acuerdo al nivel de conocimiento que evidencian los estudiantes y a sus necesidades específicas.

Estos dos propósitos, si bien parecen opuestos, suelen presentarse imbricados en la realidad. Por ejemplo, las evaluaciones de los aprendizajes de los estudiantes, realizadas en aula por un profesor, suelen cumplir un doble propósito: guiar la toma de decisiones pedagógicas y calificar al estudiante.

Sin embargo, esta distinción impacta en los usos de las evaluaciones y es necesario tener en cuenta los propósitos para que el uso dado por los beneficiarios no se aleje del propuesto por los diseñadores. Así, por ejemplo, puede ocurrir que una evaluación planeada para ser formativa y de bajas consecuencias (sin que se le asocien decisiones

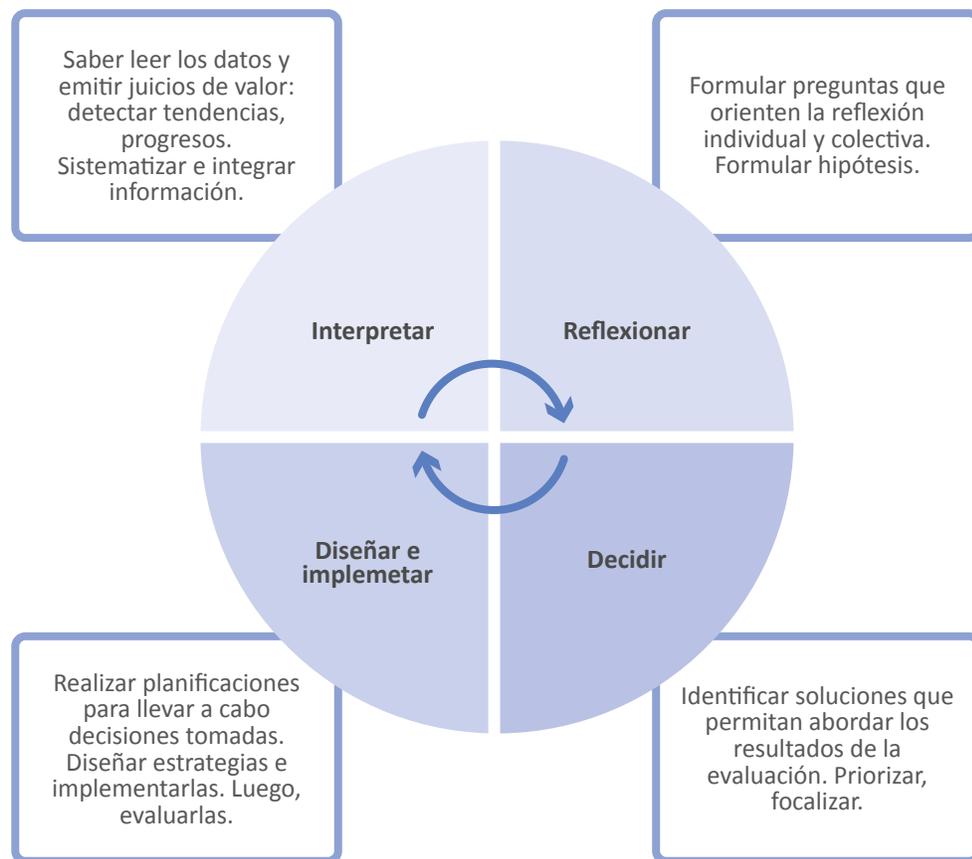
de alto impacto) se convierta en una sumativa o de alto impacto, porque los usuarios la utilizaron para calificar a los estudiantes o para otorgar incentivos o castigos a los docentes. De la misma manera, una evaluación diseñada con el fin de informar sobre los aprendizajes de los estudiantes, puede ser usada como mecanismo de rendición de cuentas de las escuelas y docentes, lo que no solo es riesgoso porque la evaluación se distancia de su propósito inicial (y, por ende, los resultados no permiten extraer ese tipo de conclusiones), sino también (y sobre todo) porque existe evidencia respecto de los efectos perversos de usar las evaluaciones de aprendizaje de ese modo (Darling-Hammond, 2007; Falabella y De la Vega, 2016; Koretz, 2010, 2017; Mons, 2009). Los ejemplos extraídos al respecto son numerosos: reducción del currículum a los sectores y formatos que se evalúa en la prueba; inflación de resultados (producto de estrategias como excluir de la evaluación a los estudiantes con dificultades escolares o falsear los resultados); y predominio de la motivación extrínseca por sobre el placer de aprender, entre otros.

Ya tengo la información, ¿cómo la leo e interpreto?

Como se señaló, la lectura e interpretación de los resultados de las evaluaciones educativas comienza por una comprensión adecuada de los tipos de datos existentes y los usos asociados a ellos. El conocimiento de nociones básicas de estadística, como las que se presentaron más arriba, favorece una interpretación correcta y más completa de los resultados, que, en muchos casos se presentan en un lenguaje técnico poco accesible para el público general. De hecho, las investigaciones sobre los usos y toma de decisiones basada en datos, revelan que una de las razones por las cuales la información evaluativa no se usa es por desconocimiento de su sentido y por incomprensión de los reportes de resultados (Mandinach, 2012; Manzi, Bogolasky, Gutiérrez, Grau y Volante, 2014). Por eso se sugiere la alfabetización en datos a los equipos pedagógicos y de gestión escolar, con el fin de que adquieran las competencias necesarias para aprovechar adecuadamente la información de la cual disponen (Kippers, Poortman, Schildkamp y Visscher, 2018).

La lectura e interpretación de datos constituyen la etapa inicial en el ciclo de usos de la información evaluativa (ver figura 7).

FIGURA 7
CICLO DEL USO DE LA INFORMACIÓN EVALUATIVA



Fuente: adaptado de Gill, Borden, & Hallgren (2014); Kippers *et al.* (2018); Mandinach (2012).

Los modelos del ciclo de uso de los datos para la toma de decisiones son múltiples y ofrecen variados componentes en el proceso, que van desde la lectura al uso instrumental de los datos (Gill, Borden, y Hallgren, 2014; Kippers *et al.*, 2018; Mandinach, 2012). En esta ocasión, se ha optado por distinguir analíticamente cuatro grandes fases o procesos: interpretar, reflexionar, decidir (o tomar decisiones) y diseñar e implementar las acciones escogidas. Estas fases relevan que un adecuado uso de la información evaluativa no es posible si de la lectura de los resultados se pasa directamente a la acción, sin que medien la reflexión y la toma de decisiones e, incluso, a un nivel interpretativo, la integración de la información proveniente de diversas fuentes e instrumentos.

En este cuadernillo nos hemos focalizado en la primera fase del ciclo, inmediatamente posterior a la publicación de resultados de la evaluación¹. Interpretar esa información supone, además de poner en práctica los conocimientos en medición, evaluación y estadística aplicada, integrarla con los resultados de otras evaluaciones disponibles y darles un sentido. De acuerdo con Mandinach (2012), esta etapa consiste en el proceso cognitivo que hace posible pasar desde el acceso a los **datos**, a la organización y resumen de **información** que conduce, luego, a la integración y priorización, permitiendo producir **conocimiento** sobre el fenómeno educativo de interés.

Ya tengo la información y la comprendo, ¿cuál es el siguiente paso para que sea formativa?

Algunas páginas arriba, se explicó la distinción entre evaluación sumativa y formativa, sin hacer alusión explícita a las ventajas de uno u otro propósito en términos educativos. Sin embargo, a partir de los argumentos esgrimidos en torno a la toma de decisiones basada en evidencia y el uso efectivo de la información, el lector podrá haber inferido que es el propósito formativo de la evaluación el que se promueve.

¿Por qué preferir el uso formativo de las evaluaciones?

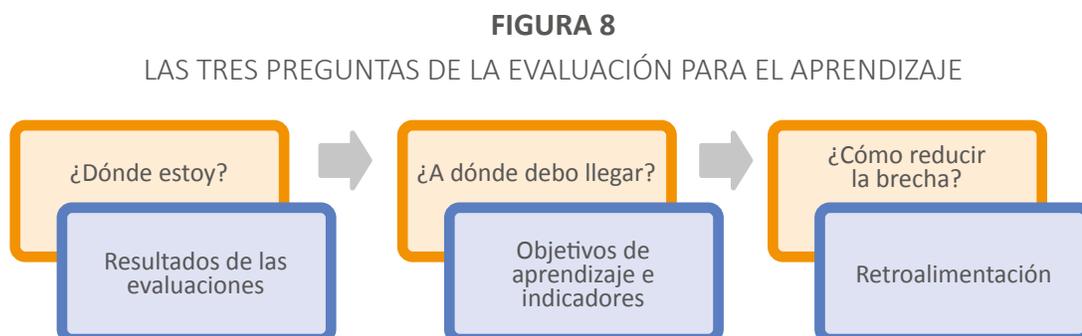
Porque si bien todo sistema educativo -por el hecho de cumplir funciones de selección y certificación de competencias- depende en cierta medida de la existencia de dispositivos evaluativos de carácter sumativo, puede optar por no limitarlos a ese uso y aprovecharlos para la retroalimentación de los procesos de enseñanza-aprendizaje, de gestión e intervención educativa. Dicho de otra manera, la evaluación constituye una herramienta de cambio en la medida en que se usa formativamente, es decir, para la mejora y la toma de decisiones (no solo para juzgar un ciclo terminado con miras a la rendición de cuentas o a la aplicación de medidas punitivas).

¿Cómo hacer para que la evaluación se use formativamente?

Es necesario que se generen condiciones individuales y colectivas que van desde la adquisición de competencias cognitivas para comprender e interpretar la información, hasta el desarrollo de actitudes favorables al cambio, espacios de reflexión individual y trabajo colaborativo en torno a los datos. Es central pasar de la interpretación a la reflexión en torno a los resultados.

¹ Las otras fases del ciclo serán abordadas en el cuadernillo 9 de esta serie, sobre los usos y la retroalimentación.

Chappuis (2009), Black y Wiliam (2009) y otros autores señalan que, para que una evaluación promueva el aprendizaje, debe orientarse a responder tres preguntas (ver figura 8).



Estas tres preguntas, que guían el proceso reflexivo, a su vez se emplearán como herramienta de retroalimentación a los actores e instituciones evaluados y serán, por tanto, retomadas en el cuadernillo de retroalimentación y uso de datos, cerrando así el ciclo de evaluación.

Consideraciones finales: ideas fuerza

Para concluir, algunas ideas centrales que hemos revisado son las siguientes:

- Las evaluaciones educacionales se han expandido en los distintos niveles de los sistemas educativos, exigiendo competencias más profundas y generalizadas de lectura e interpretación de los datos a los diversos actores.
- La evaluación no cumple sus objetivos si, aun cuando los resultados son leídos y comprendidos, no se usan para la reflexión y/o la toma de decisiones.
- La noción de validez ha recogido esta preocupación por los usos de la información, pasando de ser una propiedad del instrumento de medición a un proceso de recolección de evidencia, que apunta a establecer vínculos entre los propósitos de las evaluaciones y sus usos.
- Con tal fin, es importante que el análisis de datos sea un proceso planificado e integrado con las demás etapas, y promueva la existencia de un diseño orgánico de evaluación. Una importante herramienta para ello es el plan de análisis.
- Se debe escoger cuidadosamente lo que requiere nuestra pregunta para ser respondida, tanto en términos de los instrumentos como, de manera muy importante, de los análisis. Ahora bien, en este último campo más no siempre es mejor: muchas veces un buen análisis descriptivo es más que suficiente y mucho más comunicable.
- Se deben considerar siempre la existencia y potencialidad de ambos tipos de metodologías, cuantitativa y cualitativa, para responder nuestra pregunta. Usar una en vez de otra, basándonos en la familiaridad con una metodología o en la falta de recursos, simplemente nos impedirá llegar a las respuestas que necesita la evaluación.
- En vista de que los usos de las evaluaciones difieren en función de su intensidad, nivel de toma de decisiones, enfoque evaluativo y propósito, su uso correcto depende, en primer lugar, de la lectura e interpretación adecuada de la información; en segundo lugar, de su coherencia con los propósitos de la evaluación; y finalmente, de su orientación hacia la mejora de los procesos educativos y los aprendizajes.

Referencias

- ALKIN, M. C. y Taut, S. (2003). Unbundling evaluation use. *Studies in Educational Evaluation*, 29(1), 1-12. [http://dx.doi.org/10.1016/S0191-491X\(03\)90001-0](http://dx.doi.org/10.1016/S0191-491X(03)90001-0)
- AMERICAN EDUCATIONAL RESEARCH Association, American Psychological Association y National Council on Measurement in Education. (2018). *Standards for Educational and Psychological Testing*. Washington: American Educational Research Association.
- BACKHOFF, E., y Pérez-Morán, J. C. (coords.) (2015). *Segundo estudio internacional sobre la enseñanza y el aprendizaje (TALIS 2013). Resultados de México*. Ciudad de México: INEE.
- BLACK, P., y Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, 21(1), 5-31.
- CANALES, M. (ed.). (2006). *Metodologías de investigación social: Introducción a los oficios*. Santiago de Chile: LOM Editores.
- CHAPPUIS, J. (2009). *Learning Team Facilitator Handbook: A Resource for Collaborative Study of Classroom Assessment for Student Learning*. Melbourne, Australia: Hawker Brownlow.
- DARLING-HAMMOND, L. (2007). Race, inequality and educational accountability: The irony of 'No Child Left Behind'. *Race Ethnicity and Education*, 10(3), 245-260, DOI: 10.1080/13613320701503207.
- FALABELLA, A. y De la Vega, F. (2016). Políticas de responsabilización por desempeño escolar: Un debate a partir de la literatura internacional y el caso chileno. *Estudios Pedagógicos*, 42(2), 395-413. <https://dx.doi.org/10.4067/S0718-07052016000200023>
- FLEISCHER, D. N. y Christie, C. A. (2009). Evaluation use: Results from a survey of US American Evaluation Association members. *American Journal of Evaluation*, 30(2), 158-175.
- FLICK, U. (2004). *Introducción a la investigación cualitativa*. Madrid: Morata.
- GILL, B., Borden, B. C. y Hallgren, K. (2014). *A Conceptual Framework for Data-Driven Decision Making*. Final report of research conducted by mathematica policy research. Princeton: Bill y Melinda Gates Foundation.
- KANE, M. (2008). Terminology, emphasis, and utility in validation. *Educational Researcher*, 37(2), 76-82.

- KANE, M. (2016). Validation strategies: Delineating and validating proposed interpretations. En S. Lane, M. Raymond y T. Haladyna (eds.). *Handbook of Test Development*. (2nd ed., pp. 64-80). Nueva York: Routledge.
- KIPPERS, W. B., Poortman, C. L., Schildkamp, K. y Visscher, A. J. (2018). Data literacy: What do educators learn and struggle with during a data use intervention? *Studies in Educational Evaluation*, 58, 21-31. DOI: 10.1016/j.stueduc.2017.11.001
- KORETZ, D. (2010). *El ABC de la evaluación educativa* (Measuring Up). Ciudad de México: Ceneval.
- KORETZ, D. (2017). *The Testing Charade: Pretending to Make Schools Better*. Chicago: University of Chicago Press.
- KRAUSE, M. (1995). La investigación cualitativa: un campo de posibilidades y desafíos. *Temas de Educación*, 7(7), 19-36.
- LAI, M. K., y Schildkamp, K. (2013). Data-based decision making: An overview. En K. Schildkamp, M. K. Lai, y L. Earl (eds.), *Data-based Decision Making in Education* (pp. 9-21). Dordrecht: Springer.
- LIND, D., Marchal, W. y Wathen, S. (2015). *Estadística aplicada a los negocios y economía*. Ciudad de México: McGraw-Hill / Interamericana de España, S. A.
- MANDINACH, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, 47(2), 71-85.
- MANDINACH, E. B., & Jimerson, J. B. (2016). Teachers learning how to use data: A synthesis of the issues and what is known. *Teaching and Teacher Education*, 60, 452-457.
- MANZI, J., Bogolasky, F., Gutiérrez, G., Grau, V. & Volante, P. (2014). *Análisis sobre valoraciones, comprensión y uso del SIMCE por parte de directores escolares de establecimientos subvencionados*. Santiago: FONIDE.
- MARTÍNEZ, F. (2012). La evaluación formativa del aprendizaje en el aula en la bibliografía en inglés y francés. *Revista Mexicana de Investigación Educativa*, 17(54), 849-875.
- MONS, N. (2009). Effets théoriques et réels des politiques d'évaluation standardisée. *Revue française de pédagogie*, 169, 99-140. DOI : 10.4000/rfp.1531
- MOY, V. (2017, 29 de agosto). *La nueva ENIGH. México Cómo Vamos*. Recuperado de <http://www.mexicocomovamos.mx>
- NKWAKE, A. (2015). *Credibility, Validity, and Assumptions in Program Evaluation Methodology*. Aarau: Springer.

- PATTON, M. Q., Grimes, P. S., Guthrie, K. M., Brennan, N. J., French, B. D., y Blyth, D. A. (1977). In search of impact: An analysis of the utilization of federal health evaluation research. En C. H. Weiss (ed.), *Using Social Research in Public Policy Making* (pp. 141–163). Lexington: Lexington Books.
- RAVELA, P. (2006). *Fichas didácticas para comprender las evaluaciones educativas*. Montevideo: PREAL.
- RAVELA, P., Picaroni, B. y Loureiro, G. (2017). *¿Cómo mejorar la evaluación en el aula?* Montevideo: Grupo Magro.
- SÁNCHEZ-GÓMEZ, M. C., Martín-Cilleros, M. V., García-Peñalvo, F. J., Muñoz Sánchez, J. L., Pinto Bruno, Á., Parra, E., y Franco, M. (2017). Análisis de contenido cualitativo: Estudio de la satisfacción de los usuarios sobre la presentación de un nuevo medicamento en la salud pública. En A. P. Costa, M. C. Sánchez-Gómez y M. V. Martín-Cilleros (eds.), *La práctica de la investigación cualitativa: Ejemplificación de estudios* (pp. 57-92). Aveiro: Ludomedia.
- SCRIVEN, M. (1967). The methodology of evaluation. En R. W. Tyler, R. M. Gagne, y M. Scriven (eds.), *Perspectives of Curriculum Evaluation* (pp. 39-83). Chicago: Rand McNally.
- SECRETARÍA DISTRITAL DE SALUD DE BOGOTÁ (2013). *Guía para elaboración de plan de análisis*. Recuperado de <http://www.saludcapital.gov.co>
- STRAUSS, A. y Corbin, J. (2002). *Bases de la investigación cualitativa*. Antioquia: Universidad de Antioquia.

