MÉTRICA Y ANALÍTICA

Consultora de métrica y análisis de datos sociales

Evaluar el impacto de un curso o taller

Una guía para que su organización construya, con sus propios datos, la teoría de cambio de su programa, defina sus indicadores, diseñe el cuestionario y ejecute la evaluación de principio a fin.

Para colegios, parroquias, fundaciones y organizaciones que realizan cursos, talleres, escuelas para padres y programas formativos de 3 a 6 meses, presenciales, online o mixtos.

1 Teoría de cambio

2 Indicadores

3 Instrumento

4 Ejecución y lectura

Francisco Olmos, sociólogo · metricayanalitica.com

ANTES DE EMPEZAR

Cómo usar esta guía

La guía tiene cuatro partes. Cada una explica un paso y termina en una o más fichas para completar con los datos de su programa. Al llegar al final, las fichas completas son el plan de evaluación.

0	Ideas de partida: qué es evaluar el impacto	p. 3
1	Teoría de cambio: del problema al resultado final · Fichas 1 a 5	p. 5
2	Indicadores: cómo saber si cada eslabón ocurrió · Ficha 6	p. 12
3	Instrumento: del indicador a la pregunta · Ficha 7	p. 15
4	Ejecución: diseño, calendario, análisis y lectura · Fichas 8 a 10	p. 19
R	Fuentes y lecturas	p. 26

Qué necesita tener a mano

- Una descripción del curso o taller: objetivos, sesiones, contenidos, duración y fechas.
- Quiénes participan: cuántos, edades, si hay más de un grupo y si se los puede contactar por correo o WhatsApp.
- Una o dos horas con el equipo que diseña o ejecuta el programa. La teoría de cambio se construye mejor conversando que escribiendo solo.

Un ejemplo que acompaña toda la guía

EJEMPLO

Taller de habilidades socioemocionales y convivencia, Colegio San José, La Serena. Cuatro meses, una sesión semanal de 90 minutos (16 sesiones), presencial. Participan 60 estudiantes de 14 a 17 años de dos cursos (1° medio A y B).

En los recuadros con este formato verá cómo se completa cada paso para este taller. Úselo como modelo y ajústelo a la lógica propia de su programa.

GLOSARIO MÍNIMO

- **Actividad:** lo que el programa hace (una sesión de role playing, una charla, un proyecto).
- **Producto:** el resultado directo e inmediato de la actividad (sesiones realizadas, participantes que asistieron, material entregado).
- **Resultado intermedio:** el cambio en las personas que el programa busca producir: actitudes, conocimientos, capacidades y conductas.
- **Resultado final:** el cambio de más largo plazo en la situación de la población, directamente ligado al problema.
- **Supuesto:** condición externa al programa que tiene que cumplirse para que la cadena funcione.
- **Riesgo:** efecto negativo imprevisto que el programa podría producir.
- **Indicador:** medida observable que dice si un eslabón de la cadena ocurrió y en qué grado.
- **Línea base:** la medición antes de que el programa empiece.

PARTE 0

Ideas de partida

Tres distinciones que ordenan todo lo que viene después.

Evaluar el impacto es medir el cambio en los participantes

La satisfacción indica si el taller gustó. La evaluación de impacto mide **qué es distinto en los participantes después del programa**, en qué valoran, saben, pueden hacer y efectivamente hacen, y cuánto de ese cambio se puede atribuir al programa.

Por eso se mide **antes y después**, con las mismas preguntas. La línea base es el punto de referencia de toda cifra final.

El programa y su entorno

Un curso es una secuencia de sesiones que se sostiene durante un tiempo dentro de una organización (el colegio, la parroquia) y que busca producir efectos fuera de ellas: en la sala de clases, en la familia, en el grupo de pares, en la comunidad.

Esa definición permite separar con claridad tres planos:

Plano	Qué incluye	Qué lugar tiene en la evaluación
El programa	Lo que el equipo hace y controla: sesiones, contenidos, ejercicios, materiales, acompañamiento.	Indicadores de proceso y de producto
La organización que lo aloja	El colegio o la parroquia, con sus horarios, calendario, reglas y expectativas.	Condiciones y espacios donde se espera ver el cambio
El entorno	Familias, barrio, redes sociales, situación económica, otros programas.	Supuestos y riesgos: se registran como contexto

> Si el programa puede operar sobre algo, ese algo puede ser indicador. Si el programa solo puede ser afectado por algo, ese algo es un supuesto.

Así, cada cifra de la evaluación corresponde a algo que el programa efectivamente puede cambiar, y los factores del entorno quedan registrados como contexto para leer los resultados. Esta regla se usa en las partes 1 y 2.

Lo que se puede afirmar depende de cómo se mida

Cada diseño permite afirmar algo distinto. Comparar a los participantes antes y después muestra cuánto cambiaron; para atribuir ese cambio al programa, se agrega un grupo de comparación. La parte 4 explica las opciones. Lo importante desde ya es que **el diseño se decide antes de empezar**, porque la línea base se toma antes de la primera sesión.

PARTE 1

Teoría de cambio

Una teoría de cambio explica cómo y por qué las actividades de un programa deberían producir una serie de cambios que llevan al resultado final buscado. Es una explicación que se puede contrastar con datos.

Se construye en seis pasos. Los tres primeros arman la cadena; los tres últimos la ponen a prueba antes de empezar a medir.



Figura 1. La cadena causal. Supuestos y riesgos acompañan cada flecha.

PASO 1

Analizar la situación

Toda teoría de cambio parte de un buen diagnóstico: cuál es el problema que el programa enfrenta, a quiénes afecta, cuáles son sus causas y sus consecuencias. Aunque el programa ya esté en marcha, vale la pena volver a preguntarse si está intentando resolver el problema correcto.

- **El problema** se escribe como una carencia de la población. Por ejemplo: "los estudiantes resuelven sus conflictos con agresiones".
- **La población** tiene que quedar especificada: quiénes, qué edad, dónde.
- **Las causas** importan porque el programa actuará sobre algunas de ellas. Conviene separar las causas sobre las que el programa puede operar (falta de herramientas, creencias) de las causas externas (situación familiar, barrio).

EJEMPLO

Problema: los estudiantes de 1° medio resuelven sus conflictos con descalificaciones y agresiones, en la sala, en los recreos y en redes sociales, lo que deteriora la convivencia del curso.

Causas sobre las que el taller puede operar: dificultad para reconocer y regular emociones; pocas herramientas para conversar un desacuerdo; normas de grupo que valoran "no dejarse".

Causas externas al taller: conflictos familiares, exposición a violencia en el barrio.

PASO 2

Definir el propósito y el resultado final

El **propósito** es el cambio general que se busca, la razón de ser del programa. Se parece a la misión de la organización. El **resultado final** es su expresión concreta para esta población: el cambio de más largo plazo que el programa busca contribuir a producir.

Una prueba sencilla: **el resultado final es lo opuesto al problema**, escrito como una situación lograda. Si en el cuadro del resultado final aparece el problema ("violencia escolar") o una actividad ("realizar el taller"), está mal formulado.

Agregue también **una señal observable**: qué se vería distinto si el resultado se logra. Esa señal será la base del indicador de resultado final.

EJEMPLO

Propósito: una convivencia escolar respetuosa. **Resultado final:** estudiantes que resuelven sus desacuerdos conversando y con respeto. **Señal observable:** menos conflictos con agresión entre compañeros, reportados por ellos mismos y registrados por convivencia escolar.

PASO 3

Completar la cadena causal

Entre las actividades y el resultado final hay una secuencia de cambios que tienen que ocurrir. Lo más útil es construirla **desde el final hacia el inicio**: para que los estudiantes resuelvan sus conflictos conversando, ¿qué tiene que haber cambiado antes en ellos? Y para que eso cambie, ¿qué tiene que haber hecho el programa?

Los resultados intermedios se ordenan en cuatro dimensiones. Son el foco principal de la mayoría de las evaluaciones de impacto de programas formativos:

Dimensión	Qué cambia	Ejemplo de cambio esperado
Actitud	Lo que las personas valoran, creen o están dispuestas a hacer	Valoran el diálogo como forma de resolver conflictos
Conocimiento	Lo que saben y comprenden	Reconocen sus emociones y las de otros
Capacidad	Lo que son capaces de hacer	Aplican una técnica para calmarse ante la rabia
Conducta	Lo que efectivamente hacen, fuera de las sesiones	Conversan sus desacuerdos en vez de pelear

La conducta es casi siempre el eslabón decisivo y el más difícil: para actuar distinto, además de saber y querer, hace falta la ocasión de hacerlo. Por eso cada actividad debería apuntar explícitamente a una dimensión, y al menos una actividad debería trabajar la conducta.

Los **productos** son el resultado directo de las actividades (sesiones realizadas, participantes que asistieron, campaña difundida). A veces son casi una reformulación de la actividad; está bien, porque son lo que permite verificar que el programa se ejecutó.

EJEMPLO

Actividad	Producto	Dimensión	Cambio esperado
Talleres de reconocimiento emocional	8 sesiones realizadas	Conocimiento	Identifican y nombran sus emociones
Práctica de técnicas de autorregulación	Ejercicios aplicados en sesión	Capacidad	Aplican una técnica para calmarse
Dinámicas de empatía y colaboración	Dinámicas realizadas	Actitud	Valoran el respeto entre pares
Círculos de diálogo sobre conflictos reales del curso	Círculos realizados	Conducta	Resuelven desacuerdos conversando

PASO 4

Explicar el mecanismo y dónde se verá el cambio

Una teoría explica **por qué** cada flecha de la cadena debería funcionar. Una buena teoría de cambio tiene dos partes: una idea de cómo se produce el cambio en las personas, y una idea de cómo el programa lo desencadena.

Para cambios de conducta, una pauta útil es revisar tres condiciones previas: **conocimiento** (saben qué hacer), **motivación** (quieren hacerlo) y **oportunidad** (tienen ocasión y medios para hacerlo). Un programa puede lograr las dos primeras y fallar en la tercera, y la evaluación tiene que poder detectarlo.

Pregúntese también **en qué espacios se espera ver el cambio**. Las sesiones son el lugar donde ocurre el programa, pero el cambio que importa tiene que aparecer en otros lados. Llamamos a esos espacios **acoplamientos**: los lugares de contacto entre el programa y la vida de los participantes.

	Colegio	Parroquia
Espacio cotidiano	La sala de clases y el curso	El grupo juvenil
Espacio institucional	Convivencia escolar, consejo de curso, protocolos	Servicios de la comunidad: liturgia, catequesis, acción social
Entorno externo	La familia, el grupo de pares, las redes	La familia, el colegio o el trabajo

Cada organización observa a sus participantes desde sus propios códigos: el colegio, desde el rendimiento, la asistencia y la convivencia; la parroquia, desde la participación y el compromiso. Un programa que quiera ser leído por la organización tiene que traducir sus logros a esos términos y mantener a la vez sus propios objetivos.

PASO 5

Identificar supuestos y riesgos

Los **supuestos** son condiciones externas al programa que tienen que cumplirse para que la cadena funcione. Los **riesgos** son efectos negativos imprevistos que el propio programa podría producir.

Dos condiciones que parecen supuestos y en realidad dependen del propio programa, por lo que se miden:

- "La necesidad existe": eso es el diagnóstico (paso 1), y hay que hacerlo bien.
- "Los facilitadores hacen bien su trabajo": eso es la implementación, y se mide con indicadores de proceso.

EJEMPLO

Supuestos: el colegio mantiene el horario del taller durante los cuatro meses; las pruebas y vacaciones interrumpen como máximo dos sesiones seguidas; el profesor jefe respalda el taller ante el curso.

Riesgo: que hablar de conflictos reales del curso en los círculos exponga a algunos estudiantes y aumente las tensiones. Mitigación: normas de confidencialidad y facilitación capacitada.

PASO 6

Resumir la hipótesis causal

Toda la teoría tiene que caber en una frase. Una frase completa es señal de una cadena completa.

PLANTILLA

"Si **[actividades]** producen **[productos]**, esto debería llevar a **[resultados intermedios]**, lo que al final mejorará **[resultado final]**, contribuyendo a **[propósito]**. Esto ocurrirá siempre que **[supuestos principales]**."

EJEMPLO

"Si 16 sesiones de reconocimiento emocional, autorregulación, empatía y círculos de diálogo llegan a la mayoría de los estudiantes de 1° medio, esto debería llevarlos a reconocer y regular sus emociones, valorar el diálogo y practicarlo, lo que al final hará que resuelvan sus desacuerdos conversando y con respeto, contribuyendo a una convivencia escolar respetuosa. Esto ocurrirá siempre que el colegio sostenga el horario y el profesor jefe respalde el taller."

Errores frecuentes al construir una teoría de cambio

Error	Cómo se ve y cómo corregirlo
Mecanismo implícito	Flechas que unen actividades con resultados y omiten la razón. Corrija agregando, en cada flecha, por qué debería funcionar (paso 4).
Demasiado simple	"Taller → conocimiento → conducta → mejor convivencia", como si el programa actuara aislado. Agregue supuestos y el eslabón de oportunidad.
Demasiado complicada	Tantos cuadros que la idea general se pierde. Agrupe actividades que apuntan al mismo cambio.
El problema en el lugar del resultado	"Resultado final: violencia escolar". El resultado es el cambio buscado: "convivencia respetuosa".
Una teoría archivada	Se hace al inicio y queda archivada. Úsela para decidir qué datos se recogen: cada nodo de la cadena tiene un indicador (parte 2).

FICHA 1 · PASO 1 Análisis de la situación

Complete con el equipo del programa. Escriba en el lenguaje de su organización.

Nombre del curso o taller	
Organización	
Tipo de organización	☐ Colegio ☐ Parroquia ☐ Fundación ☐ Municipio ☐ Otra: _____
Tipo de intervención	☐ Curso ☐ Taller ☐ Programa ☐ Escuela para padres ☐ Otro: _____
Modalidad	☐ Presencial ☐ Online ☐ Mixta
Duración y frecuencia	____ meses · ____ sesiones · _____
Fechas	Inicio: ___ / ___ / _____ Término: ___ / ___ / _____
Participantes	N° estimado: ____ · Edad: _____ · N° de grupos: ____

¿Cuál es el problema que el programa busca enfrentar?

Escríbalo como una situación de la población.

¿A quiénes afecta? ¿Dónde?

Causas sobre las que el programa puede operar

Falta de información, de herramientas, creencias, normas del grupo, etc.

Causas externas al programa

Irán a los supuestos (ficha 4).

Consecuencias del problema si la situación se mantiene

FICHA 2 · PASO 2 Propósito y resultado final

Propósito del programa

El cambio general que se busca. Suele parecerse a la misión de la organización.

Resultado final esperado

Lo opuesto al problema, escrito como una situación lograda en esta población.

Señal observable

¿Qué se vería distinto si el resultado se logra? ¿Quién lo vería?

Horizonte

☐ Al terminar el programa ☐ Algunos meses después ☐ Más de un año después

FICHA 3 · PASO 3 Cadena causal

Una fila por actividad. Dimensión: A = actitud, Co = conocimiento, Ca = capacidad, Cd = conducta. Procure que al menos una actividad apunte a la conducta.

Actividad	Producto	Dim.	Cambio esperado en los participantes

FICHA 4 · PASOS 4 Y 5 Mecanismo, acoplamientos, supuestos y riesgos

¿Por qué estas actividades deberían producir el cambio?

Explique la flecha más importante de su cadena. ¿Qué ocurre en las personas?

Condición para el cambio de conducta	¿Qué actividad la trabaja?	¿Está cubierta?
Conocimiento: saben qué hacer		☐ Sí ☐ En parte ☐ No
Motivación: quieren hacerlo		☐ Sí ☐ En parte ☐ No
Oportunidad: tienen ocasión de hacerlo		☐ Sí ☐ En parte ☐ No

¿Dónde se espera ver el cambio? (acoplamientos)

Espacio	¿Qué se vería distinto ahí?	¿Quién lo puede observar?

Supuestos y riesgos

Supuestos: condiciones externas necesarias	Riesgos: efectos negativos posibles y cómo mitigarlos

FICHA 5 · PASO 6 Hipótesis causal

"Si [actividades] producen [productos], esto debería llevar a [resultados intermedios], lo que al final mejorará [resultado final], contribuyendo a [propósito]. Esto ocurrirá siempre que [supuestos]."

PARTE 2

Indicadores

Un indicador es una medida observable que dice si un eslabón de la cadena ocurrió y en qué grado. Los indicadores salen de la teoría de cambio de cada programa.

> Cada nodo de la cadena tiene al menos un indicador, y cada indicador corresponde a un nodo.

Así, todo lo que se mide es algo que el programa se propuso, y si el resultado final queda por debajo de lo esperado, los indicadores muestran en qué eslabón se rompió la cadena.

Cuatro niveles de indicadores

Nivel	Qué responde	Ejemplo	Fuente típica
Proceso	¿Se ejecutó el programa como estaba previsto? ¿Llegó a quienes debía?	% de sesiones realizadas; asistencia promedio; % que completa el programa	Registros del equipo
Producto	¿Se entregó lo que cada actividad debía entregar?	Círculos de diálogo realizados / planificados	Registros del equipo
Resultado intermedio	¿Cambiaron las actitudes, conocimientos, capacidades y conductas?	Cambio en el puntaje de autorregulación entre el inicio y el cierre	Cuestionario antes y después
Resultado final	¿Cambió la situación que dio origen al programa?	% que en el último mes resolvió sus desacuerdos conversando	Cuestionario; registros de la organización

Se puede agregar un quinto tipo, de **percepción**: la satisfacción de los participantes y su opinión sobre la utilidad del programa. Sirve para mejorar el programa y complementa a los anteriores.

Cómo se ve un buen indicador

Criterio	Pregunta de control
Específico	¿Mide un solo cambio, el de su nodo?
Observable	¿Se puede registrar de forma directa con una pregunta, una lista o un registro?
Sobre algo que el programa opera	¿Mide algo sobre lo que el programa actúa?
Sensible al plazo	¿Es razonable que cambie en 3 a 6 meses? Para cambios de más largo plazo, use un resultado intermedio como indicador de avance.
Factible	¿Se puede obtener con los recursos y el acceso que se tiene?

De la cadena a la matriz de indicadores

Para cada indicador se definen seis elementos: qué mide, cómo se calcula, de dónde sale el dato, cuándo se mide, quién lo registra y cuál es la meta.

- **Fórmula:** casi siempre un porcentaje (cuántos de cuántos) o un promedio. Escríbala completa: "N° de estudiantes que asistieron al 75% o más de las sesiones / N° de estudiantes inscritos × 100".
- **Meta:** para los indicadores de proceso y producto se puede fijar desde el inicio (por ejemplo, 90% de sesiones realizadas). **Las metas de resultado se fijan después de la línea base**, cuando se conoce el punto de partida de los participantes.

EJEMPLO · MATRIZ DE INDICADORES (EXTRACTO)

Nivel	Indicador	Fuente	Cuándo	Meta
Proceso	% de sesiones realizadas según lo planificado	Registro de sesiones	Continuo	≥ 90%
Proceso	% de estudiantes que asiste al 75% o más	Lista de asistencia	Continuo	≥ 70%
Producto	Círculos de diálogo realizados / planificados	Registro de sesiones	Mensual	100%
Resultado Ca	Cambio en el puntaje promedio de autorregulación (escala 1 a 5)	Cuestionario	Inicio y cierre	Tras línea base
Resultado Cd	Cambio en la frecuencia de desacuerdos resueltos conversando	Cuestionario	Inicio y cierre	Tras línea base
Final	% que en el último mes resolvió sus desacuerdos conversando	Cuestionario; registro de convivencia	Inicio, cierre, seguimiento	Tras línea base

Del indicador a las preguntas

Los indicadores de resultado se miden con preguntas. Una sola pregunta es frágil: puede malentenderse o responderse al azar. Por eso cada indicador de resultado se mide con **dos a cuatro ítems** que apuntan al mismo cambio desde ángulos distintos, y el indicador es el promedio de esos ítems. La parte 3 explica cómo redactarlos.

EJEMPLO · INDICADOR DE CAPACIDAD

Indicador: "Aplican una técnica para calmarse ante la rabia". Ítems (escala 1 = muy en desacuerdo a 5 = muy de acuerdo):

1. Cuando me enojo, sé qué hacer para calmarme.
2. Puedo esperar antes de responder cuando alguien me molesta.
3. Puedo decir lo que pienso con respeto.

FICHA 6 · PARTE 2 Matriz de indicadores

Complete al menos: 2 indicadores de proceso, 1 de producto por actividad, 1 de resultado por cada dimensión trabajada y 1 de resultado final. Revise que cada fila corresponda a un nodo de su cadena (ficha 3).

Nivel	Indicador y fórmula	Fuente	Cuándo	Quién	Meta
Proceso					
Proceso					
Producto					
Producto					
Producto					
Resultado					
Resultado					
Resultado					
Resultado					
Final					
Percepción					

Control de la matriz

- ☐ Cada nodo de la cadena (actividad, producto, resultado intermedio, resultado final) tiene al menos un indicador.
- ☐ Todos los indicadores miden algo sobre lo que el programa actúa.
- ☐ Cada indicador de resultado tiene entre dos y cuatro ítems asignados.
- ☐ Las metas de resultado quedan "por fijar tras la línea base".

PARTE 3

El instrumento: del indicador a la pregunta

Un buen cuestionario es un instrumento diseñado para que las personas respondan con exactitud, en poco tiempo y con confianza.

Cómo responde una persona una pregunta

Responder exige al menos cuatro operaciones mentales. Un error en cualquiera de ellas produce una respuesta equivocada, aunque la persona responda de buena fe:

Operación	Qué hace la persona	Qué hacer al redactar
Comprender	Entiende qué se le pregunta	Lenguaje de los participantes; una idea por pregunta; redacción afirmativa
Recordar	Busca en su memoria la información	Períodos de referencia cortos ("en el último mes"); ejemplos que ayuden a recordar
Juzgar	Evalúa, estima o suma lo que recordó	Preguntar por hechos concretos más que por promedios ("¿cuántas veces...?")
Responder	Elige una alternativa	Alternativas claras, exhaustivas y excluyentes

Reglas de redacción

1. **Una idea por ítem.** "Respeto a mis compañeros y a mis profesores" son dos preguntas.
2. **El lenguaje de quien responde.** Pruebe los ítems con jóvenes de la edad de los participantes: ellos confirman si la redacción es clara.
3. **Conductas con período de referencia.** Las preguntas de conducta piden hechos de un período corto y definido: "En el último mes, ¿cuántas veces...?". A menor período, mayor exactitud. Use el mismo período al inicio y al cierre.
4. **Cuidado con las escalas de acuerdo.** Las escalas "de acuerdo / en desacuerdo" invitan a decir que sí (aquiescencia). Combine ítems redactados en ambos sentidos, o use escalas de frecuencia (nunca, a veces, casi siempre, siempre) cuando pregunte por conductas.
5. **Opción "no sé", solo cuando corresponde.** Si se ofrece, la gente la usa. Inclúyala en preguntas sobre hechos que la persona podría desconocer; evítela en preguntas de opinión sobre uno mismo.
6. **Preguntas sensibles con redacción neutra.** Consumo, violencia o conflictos se responden con más honestidad si el cuestionario es autoadministrado, anónimo, respondido en privado y con una redacción comprensiva ("A muchos jóvenes les ha pasado que... ¿Le ha pasado a usted?").
7. **Lo más corto posible.** Quien responde quiere minimizar el esfuerzo. Si el cuestionario es largo, empieza a responder cualquier cosa. Vuelva a la matriz de indicadores: cada pregunta alimenta un indicador.

> **Preguntas filtro.** Cuando un "sí" lleva a cinco preguntas más, las personas aprenden rápido a responder "no" para terminar antes. Si necesita preguntas filtro, haga primero todas las preguntas filtro juntas y después las de seguimiento.

La estructura del cuestionario

Parte	Qué contiene

1. Presentación	Quién hace el estudio y para qué; cuánto demora; que la participación es voluntaria y las respuestas son confidenciales; a quién escribir si hay dudas. En un cuestionario autoadministrado todo esto tiene que estar escrito.
2. Código anónimo	Un código que la persona construye y que permite unir su respuesta del inicio con la del cierre de forma anónima (ver recuadro).
3. Preguntas iniciales	Fáciles y relacionadas con el tema, para que la persona sienta que vale la pena responder.
4. Módulos	Los ítems agrupados por tema, con una frase de transición antes de cada módulo ("Ahora queremos preguntarte sobre cómo resuelves los desacuerdos").
5. Preguntas sensibles	Hacia el final, cuando ya hay confianza.
6. Datos de caracterización	Edad, curso, género, etc. Al final, y solo los que se usarán en el análisis.
7. Agradecimiento	Siempre. Quien responde le está haciendo un favor al programa.

EL CÓDIGO ANÓNIMO PARA UNIR EL ANTES Y EL DESPUÉS

Pida a cada participante que construya el mismo código en ambas mediciones con datos que recuerde y que son estables, por ejemplo: **las dos primeras letras del nombre de su madre + el día de su nacimiento + la última letra de su primer apellido**. "MA" + "07" + "Z" = MA07Z. Así se comparan las respuestas de la misma persona y se mantiene su anonimato.

Detalles que evitan errores

- **Numere las preguntas** (P1, P2… o A1, B1… por módulo) y las alternativas, con códigos consistentes: si "No" es 2 en una sección, es 2 en todas.
- **Saltos explícitos.** Indique siempre a qué pregunta pasar. El error más común es que la gente responda preguntas que le corresponden a otros.
- **La misma versión en todas las mediciones.** Cambiar el orden o la redacción entre el inicio y el cierre hace que las diferencias puedan deberse al cuestionario. Mantenga la misma versión para que reflejen el efecto del programa.

Elegir la forma de aplicación

La forma de aplicación depende de cómo se puede llegar a los participantes, de lo sensible del tema y de los recursos. Cada forma tiene sus ventajas y sus riesgos:

Forma	Ventajas	Riesgos	Conviene cuando
Papel en la sesión	Responde casi todo el grupo; funciona offline	Hay que digitar; saltos y respuestas en blanco quedan a criterio de quien responde	Grupos presenciales con conexión limitada
Formulario en línea en la sesión (QR)	Alta respuesta; saltos y validaciones automáticos; datos listos	Requiere celular o equipo y conexión en la sala	Grupos presenciales con conexión. **Opción recomendada.**
Formulario enviado a distancia	Barato; libera tiempo de sesión	Menor respuesta y más pérdida entre el inicio y el cierre	Programas online con contacto de todos los participantes

Entrevista	El entrevistador aclara y motiva; apto para quienes leen con dificultad	Más cara; las personas responden lo socialmente aceptable	Participantes con baja lectoescritura; temas complejos y de baja sensibilidad

Para temas sensibles conviene que las personas respondan solas (autoadministrado), aunque el equipo esté presente para motivar y resolver dudas. Si se combinan formas de aplicación, use **la misma forma al inicio y al cierre** para cada persona: las personas responden distinto según el medio.

Si el cuestionario es en papel

- Diseño amable: si parece que a quien lo hizo le importa, quien responde siente que vale la pena hacerlo bien. Sobrio.
- Preguntas en una sola columna, alineadas, con cada pregunta y sus alternativas completas en la misma página.
- Agrupe visualmente las partes de una misma pregunta y deje más espacio entre preguntas distintas.
- Evite la impresión horizontal y los pliegues raros. Los datos de uso interno (código de grupo, fecha) van pequeños, en la esquina inferior derecha.

Probar antes de aplicar

Siempre, aunque sea con cinco a diez personas parecidas a los participantes. Pídales que respondan en voz alta y que digan qué entendieron en cada pregunta. Corrija y, si cambió mucho, vuelva a probar. Diseñar un cuestionario es un proceso de ida y vuelta, y la prueba se planifica en el calendario desde el inicio.

FICHA 7 · PARTE 3 Banco de ítems y revisión del cuestionario

Escriba 2 a 4 ítems por indicador de resultado. Código: dimensión + número (A1, Co1, Ca1, Cd1, F1). Escala sugerida: 1 = muy en desacuerdo a 5 = muy de acuerdo, o de frecuencia para conductas.

Código	Indicador al que alimenta	Redacción del ítem	Escala

Revisión de cada ítem

- ☐ Tiene una sola idea.
- ☐ Usa palabras de los participantes.
- ☐ Las conductas tienen período de referencia.
- ☐ Es neutral: deja abierta cualquier respuesta.
- ☐ Alimenta un indicador de la ficha 6.

Revisión del cuestionario

- ☐ Presentación completa: quién, para qué, duración, voluntario, confidencial, contacto.
- ☐ Código anónimo al inicio.
- ☐ Preguntas sensibles y caracterización al final.
- ☐ Preguntas numeradas y saltos explícitos.
- ☐ Se responde en menos de 15 minutos.
- ☐ Probado con 5 a 10 personas.

Forma de aplicación

☐ Papel en sesión ☐ QR en sesión ☐ Enviado a distancia ☐ Entrevista ☐ Combinada: _____

PARTE 4

Ejecución de la evaluación

Con la teoría, los indicadores y el instrumento listos, queda decidir el diseño, fijar el calendario, recoger los datos con cuidado, analizarlos y leer los resultados contra la teoría de cambio.

Elegir el diseño

Diseño	Cómo funciona	Qué permite afirmar
Antes y después, mismo grupo	Se aplica el cuestionario al inicio y al cierre a los mismos participantes.	Cuánto cambiaron los participantes. La atribución al programa es parcial.
Con grupo de comparación	Se mide también, en las mismas fechas, a un grupo parecido que participará más adelante (por ejemplo, una lista de espera o un curso paralelo que participará después).	Cuánto del cambio se debe al programa, si los grupos son comparables.
Asignación al azar	Entre los interesados, se sortea quién participa ahora y quién en la siguiente versión.	La atribución más sólida. Rara vez factible; ético si todos participan después.
Solo al cierre (retrospectivo)	Se pregunta al cierre cómo estaban antes y cómo están ahora.	Débil: el recuerdo tiende a exagerar el cambio. Úselo solo cuando la línea base ya pasó.

Para cursos y talleres de 3 a 6 meses, el diseño más común y razonable es **antes y después en el mismo grupo**, complementado con indicadores de proceso. Si hay lista de espera o un curso que participará el semestre siguiente, úselo como grupo de comparación: mejora mucho lo que se puede afirmar con poco costo adicional.

¿Cuántos participantes hacen falta?

Con pocos casos, solo se pueden detectar cambios grandes; los cambios moderados quedan dentro del margen de azar. El tamaño del cambio se expresa como **d**: aproximadamente 0,2 es un efecto pequeño, 0,5 moderado y 0,8 grande. La tabla indica el menor efecto que se puede detectar con confianza según el número de personas que responden **ambas** mediciones:

Casos emparejados	10	15	20	25	30	40	50	60	80	100
Efecto detectable (d)	0,89	0,72	0,63	0,56	0,51	0,44	0,40	0,36	0,31	0,28

Aproximación para una comparación antes y después en el mismo grupo, con 95% de confianza y 80% de probabilidad de detectar el efecto si existe ($d \approx 2{,}8 / \sqrt{n}$).

Ojo: los casos emparejados son menos que los inscritos. Entre la deserción y quienes faltan el día de la medición, es común perder un 20% a 40%. Si se inscriben 40 y se espera emparejar al 70%, quedan unos 28 casos: se detectan efectos de $d \approx 0{,}5$ o mayores.

EJEMPLO

60 estudiantes inscritos, 70% emparejados ≈ 42 casos. Efecto detectable ≈ 0,43: el taller puede detectar efectos moderados. Para conservar casos, el cuestionario se aplica con QR en la primera y en la última sesión, y se repite la sesión siguiente para quienes faltaron.

El calendario de la evaluación

Momento	Qué se hace
Antes del inicio	Teoría de cambio validada con el equipo; cuestionario probado; consentimientos firmados; códigos y formulario listos.
Línea base	Antes de la primera sesión o en sus primeros 15 minutos. Todas las dimensiones.
Durante	Registro de asistencia y de sesiones realizadas (indicadores de proceso y producto), en cada sesión.
Intermedia (opcional)	A mitad del programa, en programas de 4 meses o más. Actitud y conocimiento, que cambian primero. Permite corregir a tiempo.
Cierre	En la última sesión. Todas las dimensiones, el resultado final y la satisfacción.
Seguimiento (opcional)	Uno a dos meses después. Conducta: ¿el cambio se sostiene fuera del programa?
Informe	Dentro del mes siguiente al cierre, para que sirva a las decisiones del período siguiente.

Cuidados éticos

- **Consentimiento informado:** las personas saben para qué se usan sus respuestas y participan libremente.
- **Menores de edad:** consentimiento de madres, padres o apoderados y asentimiento de los propios jóvenes.
- **Confidencialidad:** las respuestas individuales quedan solo en manos del equipo evaluador. Los resultados se informan agrupados, en grupos suficientemente grandes para resguardar la identidad de cada persona.
- **Resguardo de los datos:** en una carpeta con acceso restringido, identificados solo por el código anónimo.

Dónde se puede equivocar una evaluación

Las cifras finales pueden alejarse de la realidad por cuatro razones. Conocerlas permite prevenirlas:

Error	Qué es	Cómo prevenirlo
Cobertura	Parte de quienes debían responder quedó fuera	Lista completa de inscritos; aplicar en la sesión
Pérdida de casos	Quienes faltan a una medición suelen ser distintos de quienes responden	Repetir la medición para ausentes; comparar a quienes desertaron con quienes siguieron
Medición	Las respuestas se alejan de lo que se quería medir	Buena redacción, prueba previa, anonimato
Procesamiento	Errores al digitar, codificar o unir bases	Formulario en línea con validaciones; revisar códigos duplicados

El análisis, paso a paso

1. **Ordenar los datos.** Una fila por persona y una columna por ítem, con el código anónimo como identificador. Revise códigos duplicados o mal escritos.
2. **Emparejar.** Una las respuestas del inicio y del cierre por código. Trabaje con quienes respondieron ambas.
3. **Puntaje por dimensión.** Promedio de los ítems de cada indicador, para cada persona y en cada medición. Si invirtió ítems, recodifíquelos antes (en una escala de 1 a 5: 6 − respuesta).
4. **Cambio promedio.** Diferencia entre el puntaje de cierre y el de inicio, para cada persona, y su promedio.
5. **Prueba estadística.** Prueba t para muestras pareadas (o Wilcoxon si hay pocos casos o datos muy asimétricos). Dice si el cambio es mayor de lo esperable por azar.
6. **Tamaño del efecto.** d = promedio de las diferencias / desviación estándar de las diferencias. Dice si el cambio es pequeño, moderado o grande.
7. **Porcentaje que mejora.** Proporción de participantes cuyo puntaje subió. Es fácil de comunicar.
8. **Proceso y producto.** Calcule los indicadores de proceso y producto con los registros.

> EN EXCEL Y EN R
>
> **Excel:** promedio por dimensión con `=PROMEDIO(B2:D2)`; prueba t pareada con `=PRUEBA.T.N(rango_inicio; rango_cierre; 2; 1)` (devuelve el valor p); d con `=PROMEDIO(dif)/DESVEST.M(dif)`.
>
> **R:** `t.test(cierre, inicio, paired = TRUE)`; `wilcox.test(cierre, inicio, paired = TRUE)`; `mean(cierre - inicio) / sd(cierre - inicio)`.

Leer los resultados contra la teoría de cambio

El valor de una teoría de cambio aparece al leer los resultados. Si el resultado final queda por debajo de lo esperado, la cadena muestra **dónde se rompió**, y eso define la recomendación. Hay dos casos distintos: la **falla de implementación** (el programa se ejecutó de forma incompleta) y la **falla de teoría** (el programa se ejecutó bien y el cambio esperado quedó pendiente).

¿Se ejecutó? (proceso)	Resultados intermedios	Resultado final	Lectura y recomendación
Incompleto	—	—	**Falla de implementación.** Antes de juzgar el programa, hay que mejorar su ejecución: asistencia, sesiones, facilitación.
Sí	Se mantienen	Se mantiene	**Falla de teoría en el primer eslabón.** Las actividades quedan lejos del cambio esperado. Revisar contenidos y metodología.
Sí	Cambian actitud y conocimiento; la conducta se mantiene	Se mantiene	**Se rompe entre saber y hacer.** Revisar la oportunidad: ¿hay espacios donde practicar? ¿El entorno lo permite?
Sí	Cambian	Se mantiene	**Plazo insuficiente o supuestos que fallaron.** Revisar el horizonte del resultado final y el registro de supuestos.
Sí	Cambian	Sí	**Coherente con la teoría.** Cuánto se atribuye al programa depende del diseño y de si incluye grupo de comparación.

> **Contraste las fuentes.** Si casi todos dicen que ahora conversan sus desacuerdos, pero convivencia escolar registra los mismos conflictos que antes, una de las dos fuentes se aleja de la realidad. Informe ambas cifras y explique la diferencia. A menudo indica que las personas saben lo que deberían hacer, pero aún les falta la oportunidad o el respaldo para hacerlo.

El informe

Ordene el informe según para qué se usarán los resultados (rendir cuentas, mejorar el programa, postular a fondos, comunicar a la comunidad). Una estructura que funciona:

1. **Resumen para quien decide**, en una página: qué se evaluó, qué cambió, qué se puede afirmar, con qué límites, y qué se recomienda.
2. **El programa y su teoría de cambio**, con el diagrama.
3. **Cómo se evaluó**: diseño, instrumento, fechas, cuántos respondieron y cuántos se emparejaron.
4. **Resultados por eslabón**: proceso, productos, cada dimensión y resultado final.
5. **Lectura**: dónde se sostuvo y dónde se rompió la cadena, y por qué.
6. **Límites**: el alcance de lo que el diseño permite afirmar.
7. **Recomendaciones**, separando las de implementación de las de diseño del programa.

FICHA 8 · PARTE 4 Diseño y calendario de la evaluación

Diseño	☐ Antes y después, mismo grupo ☐ Con grupo de comparación ☐ Asignación al azar ☐ Solo al cierre
Grupo de comparación (si hay)	
Participantes inscritos	_____ Casos emparejados esperados (≈ 70%): _____ Efecto detectable (tabla): d ≈ _____
Consentimientos	☐ Participantes ☐ Apoderados (menores de edad) ☐ Organización

Momento	Fecha	Qué se mide o se hace	Responsable
Preparación			
Prueba del cuestionario			
Línea base			
Registro continuo			
Medición intermedia			
Cierre			
Seguimiento			
Análisis e informe			

Registro de supuestos durante el programa

Anote cada vez que un supuesto de la ficha 4 falle. Servirá para leer los resultados.

Fecha	Supuesto afectado	Qué ocurrió y cómo afectó al programa

FICHA 9 · PARTE 4

Resultados por eslabón

Complete después del cierre. Una fila por indicador de la ficha 6.

Proceso y producto

Indicador	Meta	Logrado	Comentario

Resultados intermedios y final

Indicador	Inicio	Cierre	Diferencia	Valor p	d	% que mejora	Lectura
A							
Co							
Ca							
Cd							
Final							

Casos emparejados: _____ de _____ inscritos (____%).

¿Dónde se sostuvo y dónde se rompió la cadena?

Use la tabla de lectura de la parte 4: ¿falla de implementación, de teoría, entre saber y hacer, de plazo?

¿Qué dicen otras fuentes (registros de la organización, observación)? ¿Coinciden?

FICHA 10 Lista de control de la evaluación

Teoría de cambio

- ☐ El problema describe una carencia de la población.
- ☐ El resultado final es lo opuesto al problema.
- ☐ Cada actividad apunta a una dimensión.
- ☐ Al menos una actividad trabaja la conducta.
- ☐ Están explicados el mecanismo y los acoplamientos.
- ☐ Supuestos y riesgos identificados.
- ☐ La hipótesis cabe en una frase.
- ☐ El equipo del programa la revisó y la valida.

Indicadores

- ☐ Cada nodo tiene indicador.
- ☐ Todos los indicadores miden algo que el programa opera.
- ☐ Fórmula, fuente, momento y responsable definidos.
- ☐ Metas de resultado tras la línea base.

Instrumento

- ☐ 2 a 4 ítems por indicador de resultado.
- ☐ Presentación, código anónimo y agradecimiento.
- ☐ Probado con 5 a 10 personas.
- ☐ Igual al inicio y al cierre.

Ejecución

- ☐ Diseño elegido y su alcance explicado.
- ☐ Consentimientos firmados.
- ☐ Línea base tomada antes de la primera sesión.
- ☐ Asistencia registrada en cada sesión.
- ☐ Medición repetida para ausentes.
- ☐ Datos emparejados y revisados.
- ☐ Resultados leídos eslabón por eslabón.
- ☐ Fuentes contrastadas.
- ☐ Informe con límites y recomendaciones.

¿NECESITA APOYO?

Métrica y Analítica, consultora de métrica y análisis de datos sociales, puede revisar sus fichas, aplicar el cuestionario, procesar los datos y elaborar el informe de resultados. Escríbanos desde metricayanalitica.com.

REFERENCIAS

Fuentes y lecturas

Esta guía sintetiza, con fines prácticos, ideas de las siguientes fuentes. Para profundizar, se recomienda leerlas directamente.

Teoría de cambio

- Rogers, P. (2014). *La teoría del cambio*. Síntesis metodológicas: evaluación de impacto n.º 2. Centro de Investigaciones Innocenti de UNICEF, Florencia.
- Thibault, A. *Teoría de cambio*. Clase del curso JPAL101SPAx, Evaluación de Impacto de Programas Sociales. Abdul Latif Jameel Poverty Action Lab (J-PAL).
- Funnell, S. C. y Rogers, P. J. (2012). *Purposeful Program Theory: Effective Use of Logic Models and Theories of Change*. Jossey-Bass/Wiley.
- White, H. (2009). *Theory-Based Impact Evaluation: Principles and Practice*. 3ie Working Paper n.º 3.

Diseño de cuestionarios y métodos de encuesta

- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E. y Tourangeau, R. (2009). *Survey Methodology*. Wiley.
- Dillman, D. A. (2000). *Mail and Internet Surveys: The Tailored Design Method*. Wiley.
- Redline, C. y Dillman, D. A. (2002). The influence of alternative visual designs on respondents' performance with branching instructions in self-administered questionnaires. En Groves et al. (eds.), *Survey Nonresponse*. Wiley.
- Tourangeau, R., Kreuter, F. y Eckman, S. (2012). Motivated underreporting in screening interviews. *Public Opinion Quarterly*, 76(3), 453-469.
- Eckman, S., Kreuter, F., Kirchner, A., Jäckle, A., Tourangeau, R. y Presser, S. (2014). Assessing the mechanisms of misreporting to filter questions in surveys. *Public Opinion Quarterly*, 78(3), 721-733.
- Michigan Program in Survey Methodology y Joint Program in Survey Methodology. *Questionnaire Design*, unidades 1 a 6.

Análisis

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2.ª ed.). Lawrence Erlbaum.

Guía elaborada por Francisco Olmos, sociólogo · Métrica y Analítica, consultora de métrica y análisis de datos sociales · metricayanalitica.com.

Reproducción permitida para uso educativo, citando la fuente.