

## Orientaciones de EMpower para el Diseño de Pre y Post-Tests

El diseño de evaluación más simple es el pre y post-test, que se define como una evaluación antes y después de un programa para medir si los cambios esperados se produjeron entre las/os participantes de un programa. Se aplica una norma de prueba, encuesta o cuestionario antes que comienza la participación (**pre-test o línea de base**), y volver a aplicar después de un período determinado, o al final del programa (**post-test o línea de fondo**). Las pruebas se pueden administrar por vía oral o por escrito.

El objetivo de esta orientación es ayudar a los programas evitar algunos de los errores más comunes en el uso de pre-y post-evaluación. Una orientación más detallada está disponible en "Recursos útiles" que se enumeran a continuación.

Por ejemplo, si el lanzamiento de un programa después de la escuela que enseña habilidades de negocios coincide con la adición de estas habilidades en el plan de estudios de matemáticas de la escuela local, sería difícil saber qué resultados se deben al programa vs. la escuela. El programa podría responder mediante el análisis de donde el programa se superpone con el programa de estudios, y centrándose en los resultados del programa en aquellas áreas que el plan de estudios no aborda.

La principal debilidad de diseño pre y post-tests es que no se puede detectar otras posibles causas de los resultados positivos o negativos entre las/os participantes. Por esta razón, si un nuevo programa está bajo consideración para la expansión, otras explicaciones para los resultados deben descartarse mediante la recopilación de más información sobre cada posible explicación, y luego señalar a los resultados como resultado del programa.

### REVISIÓN DE DISEÑO

En los casos en que se está escribiendo un nuevo pre y post-test, o está adaptando un análisis existente, se recomienda encarecidamente la prueba piloto. Un buen método de pilotaje sería convocar a un grupo de asesores jóvenes - de las mismas comunidades de las/os jóvenes que estarán dentro del programa - para tomar la prueba, discutir entre sí usando esta lista de verificación, y sugerir modificaciones. Esta es una oportunidad perfecta para la participación juvenil y el liderazgo.

#### Consejos para el desarrollo y la revisión de pre y post preguntas

1. **RELEVANCIA DE CONTENIDO A OBJETIVOS:** ¿El contenido aborda claramente los objetivos del programa?
  - Si el programa se centra en el conocimiento, entonces iguale el contenido tan cerca como sea posible a los objetivos de aprendizaje.
  - Si el programa tiene como objetivo cambiar las actitudes o normas, ¿las preguntas cubren la gama de actitudes que su programa aborda?

2. **LONGITUD** - Cuanto más corto sea, mejor, sobre todo si hay preguntas abiertas.<sup>1</sup> Elimine las preguntas redundantes. Haga una prueba preliminar para asegurar que la toma de la prueba no va a durar más de media hora.<sup>2</sup>
3. **NIVEL EDUCACIONAL** – Asegúrese de que la lectura o el nivel de vocabulario es el nivel adecuado para las/os jóvenes participantes. Determine si los niveles de alfabetización exigen entrevistas orales. La retroalimentación de las/os jóvenes en palabras o frases desconocidas o ambiguas es útil.

En inglés, el nivel de lectura se puede comprobar en formato MS Word. Ejecutar "Ortografía / Gramática"; y el último artículo bajo "legibilidad" que aparece es el análisis de Flesch-Kincaid Grado escolar. Otras pruebas podrían ser más fiables en otros idiomas.
4. **ADAPTACIÓN CULTURAL O DE IDIOMA** – Cuando los que tomaron el piloto de la prueba no pueden ponerse de acuerdo sobre el significado de una pregunta, la adaptación es necesaria. Las preguntas sobre temas sensibles como la salud reproductiva y sexual a menudo se deben adaptar a la cultura juvenil local. Por otro lado, la misma prueba / cuestionario aplicado a adultos no deben usar los términos de la cultura juvenil que las/os adultas/os pueden encontrar ofensiva.
5. **EVITE LA PREGUNTAS DEMASIADO GENERALES O AMBIGUAS** Las preguntas que son demasiado generales están sujetas a una variedad de interpretaciones, dando resultados inconsistentes.
  - Por ejemplo, "¿Crees que las niñas y los niños deben ser tratados por igual?" No es tan clara como una pregunta específica sobre los objetivos del programa, tales como "¿Cree que las niñas deben poder jugar [deporte local] en lugares públicos?"
6. **EVITE LAS PREGUNTAS CAPCIOSAS O SESGADAS:**
  - Una pregunta capciosa puede provocar una respuesta predeterminada que no refleja con precisión su opinión. Por ejemplo: "¿Cómo ha cambiado su vida como resultado del programa?" Debería transformarse en "¿Ha cambiado su vida en algún modo como resultado del programa?" (sí, no) seguido de múltiples opciones o respuesta abierta.
  - Una pregunta sesgada conducirá al/a participante a dar una respuesta socialmente aceptable. Por ejemplo, "¿Usted bebe demasiado alcohol en las fiestas?" Debe ser cambiado a "¿Toma alcohol?" Y luego, si es así, dar una gama de opciones en frecuencia, la cantidad y marco.
7. **EVITE FORMULAR DOS PREGUNTAS EN UNA:** Por ejemplo, "¿Cómo calificaría sus conocimientos y habilidades financieras?" debe dividirse en dos preguntas separadas.
8. **MEZCLE DECLARACIONES NEGATIVAS Y POSITIVAS:** Al medir las actitudes o comportamientos a través de las declaraciones que piden si los encuestados están "de acuerdo" o "en desacuerdo". Mezcle al azar declaraciones que reflejan las actitudes promovidas por el programa frente a los que están desanimados. Por ejemplo, si el post-test de un programa de género sólo tiene declaraciones que favorecen la igualdad de género, los encuestados detectarán fácilmente la respuesta deseada.
9. **MUESTRA:** Cuando los programas sirven grandes cantidades de jóvenes, a menudo no hay suficiente personal o fondos para aplicar el pre y post-test a todos. En ese caso, las/os evaluadores generalmente deciden sobre el tamaño de la muestra,<sup>3</sup> y utilizan dos métodos para lograr una muestra representativa no sesgada, es decir, un conjunto más pequeño de jóvenes que son susceptibles de reflejar las características de un grupo más grande:

<sup>1</sup> Preguntas abiertas no definen un conjunto de respuestas, por lo que los participantes deben salir con sus propias respuestas. Un ejemplo podría ser: "¿Cuál es su actividad favorita?"

<sup>2</sup> Vea las sugerencias para eliminar preguntas en la página 32 de Barkman abajo "Recursos útiles" a continuación (en Inglés).

<sup>3</sup> <https://www.surveymonkey.com/mp/sample-size-calculator/>

- "La selección aleatoria" se asegura de que cada una/o de las/os jóvenes en el programa tiene la misma probabilidad de ser elegida/o. Métodos computarizados son comúnmente utilizados.<sup>4</sup> Un método no computarizado es como una lotería. Por ejemplo, en un programa que involucra a 500 estudiantes del grado 11, poner sus nombres en pedazos de papel en un recipiente, y elegir nombres uno por uno hasta que alcanza la muestra deseada de 50 para el pre y post-test.
- "Selección sistemática" es otra forma de evitar el sesgo. Divida la población total de jóvenes en el programa por el tamaño de la muestra que se haya decidido, a continuación, utilice el número resultante  $n$  para elegir cada estudiante enésimo ( $n^{\circ}$ ). Usando el mismo ejemplo, de la lista completa de las/os 500 estudiantes en el grado 11, elija cada décima/o estudiante en la lista para ser parte de la muestra, y termine con 50.

## Análisis y Comunicación de los Resultados

1. Análisis de los resultados del pre y post-test debe permitir identificar las preguntas que necesitan ser mejoradas la próxima vez que utilice la prueba. Señales comunes de necesidades de mejora son:
  - Muchas de las respuestas a preguntas específicas se quedan en blanco, respuestas incoherentes a las preguntas relacionadas, respuestas a las preguntas abiertas que reflejan falta de comprensión, e inhabilidad de responder a las preguntas hacia el final de la prueba (fatiga de las/os encuestadas/os debido a la duración).
2. Muchos instrumentos producen **datos cuantitativos** de preguntas de tipo cerrado<sup>5</sup> o valoraciones.
  - Estos datos se introducen habitualmente en un programa de computadora como Excel, y generalmente se analizan para las frecuencias, como porcentajes y promedios. Se necesitan análisis más sofisticados para observar los cambios en el tiempo, y las diferencias entre los grupos.
    - i. Si presenta los resultados en porcentajes en una tabla, siempre haga referencia entre paréntesis al número de encuestadas/os.
  - Si usted no tiene acceso a los conocimientos estadísticos, sobre todo cuando el número de participante es mayor que 50, sería de gran ayuda alistar un/a investigador/a local para analizar la significación.<sup>6</sup>
3. Preguntas abiertas producen datos cualitativos, que pueden ser analizados por tema, o por medio de aplicación de las categorías que se puede reportar cuantitativamente.<sup>7</sup>
  - Citas de preguntas abiertas se incluyen comúnmente en los informes y materiales de comunicación (sin identificación de las personas que proporcionaron la frase).

## Recursos

Susan J. Barkman, "A Field Guide to Designing Quantitative Instruments to Measure Program Impact." Una publicación excelente de Purdue University.

<http://www.northskynonprofitnetwork.org/sites/default/files/documents/Field%20Guide%20to%20Developing%20Quantitative%20Instruments.pdf>

El trabajo de Jane E. Davidson es una excelente guía para la evaluación en general, y es conocida por explicar cómo utilizar matrices de valoración para analizar la información cualitativa en la evaluación de los resultados.

<sup>4</sup> <https://www.randomizer.org/tutorial/> y <https://www.surveymonkey.com/blog/2012/06/08/random-sample-in-excel/>

<sup>5</sup> Preguntas de tipo cerrado proporcionan opciones de ajuste; tipos incluyen la opción múltiple, preguntas verdaderas-falsas, y preguntas utilizando escalas de calificación. Vea las páginas 20-25 de la guía Barkman sobre estas opciones, se hace referencia a continuación.

<sup>6</sup> Para identificar qué cambios desde el inicio a la línea de fondo se deben probablemente al programa y no al azar. La mayoría de las pruebas de significación apuntan a la probabilidad de resultados debido a la oportunidad de ser el 5% o menos.

<sup>7</sup> Véase E. Jane Davidson en Recursos útiles por debajo.

- a. Davidson, E. Jane, *Evaluative Reasoning* [http://www.unicef-irc.org/publications/pdf/brief\\_4\\_evaluativereasoning\\_eng.pdf](http://www.unicef-irc.org/publications/pdf/brief_4_evaluativereasoning_eng.pdf)
- b. Davidson, E. Jane mini-books: <http://realevaluation.com/read/minibooks/>
  - *Making the Important Measurable, Not the Measurable Important*
  - *Actionable Evaluation Basics*

I-Tech, *Guidelines for Pre- and Post-Testing*, tiene una breve y útil discusión de diseño y análisis de pre y post-tests, enfocada en sus ejemplos sobre preguntas de conocimiento. Disponible a petición de las/los oficiales de programa.

I-Tech tiene una caja de herramientas con muchos otros recursos para la evaluación de programas de entrenamiento. Consulte <http://www.go2itech.org/HTML/TT06/toolkit/evaluation/index.html>. La ficha de formularios tiene muchas herramientas que podrían utilizarse antes y después de la prueba, incluyendo listas de verificación escritas y de observación.