

Encuentro Besaldi, 23 de septiembre de 2026.

Resumen de las principales ideas

Ainhoa Vega-Bayo · Euskal Herriko Unibertsitatea

1. Mi vinculación con la evaluación

Mi trabajo se sitúa en la intersección de tres espacios. En la Universidad del País Vasco / Euskal Herriko Unibertsitatea investigo en microeconometría aplicada y enseño evaluación cuantitativa de políticas públicas, es decir, formo a quienes tendrán que evaluar mañana. En la Fundación Iseak participo en investigación aplicada sobre empleo, educación e inclusión, con datos y preguntas que vienen del propio territorio. Y con Besaldi, en el marco de BesaldiLab, he estado en contacto con el proceso de evaluación del Sistema Vasco de Garantía de Ingresos y para la Inclusión.

2. Evaluación causal de impacto

Evaluar no es fiscalizar. La evaluación integral es un proceso sistemático de generación de conocimiento sobre el diseño, la implementación y los efectos de una intervención pública. Es una herramienta de mejora y de buen gobierno, y debe recorrer todo el ciclo: pre-análisis y diagnóstico, diseño, implementación, medición de efectos, aprendizaje y ajuste. Evaluar también es aprender.

Hay muchos tipos de evaluación y todos son necesarios. Según el objeto (necesidades, diseño, proceso, resultados, impacto, evaluabilidad), según el momento (ex ante, durante, ex post) y según el agente (interna, externa, mixta). Los criterios de la OCDE ordenan el juicio: pertinencia, coherencia, eficacia, eficiencia, impacto y sostenibilidad. Conviene no confundirlos: la eficacia pregunta si funciona, la eficiencia si funciona al coste adecuado, y el impacto qué cambios reales y duraderos produce.

El problema fundamental es el contrafactual. Para atribuir un cambio a una política necesitamos saber qué habría ocurrido sin ella, y eso nunca se observa. En el marco de resultados potenciales (Rubin, 1974), de cada persona vemos solo una de las dos historias posibles. Por eso la comparación ingenua entre participantes y no participantes, o entre el antes y el después, mezcla el efecto causal con el sesgo de selección. Los tres falsos contrafactuales más frecuentes son el antes/después, que confunde tendencia temporal con efecto; el con/sin programa, porque quien participa es distinto de quien no participa; y la predicción sin tratamiento, que depende por completo de los supuestos del modelo.

Cada método compra credibilidad a cambio de un supuesto. La asignación aleatoria en los experimentos, la ausencia de confusores inobservables en la regresión y el emparejamiento, la continuidad en el umbral en los diseños de discontinuidad, la relevancia y exclusión en variables instrumentales, las tendencias paralelas en diferencias en diferencias. La diferencia entre las dos herramientas más habituales es nítida: diferencias en diferencias controla las diferencias inobservables constantes en el tiempo, mientras que el emparejamiento controla únicamente las observables. A menudo se combinan. La pregunta política decide además qué efecto queremos estimar: el efecto medio en la población o el efecto sobre quienes efectivamente participan.

Sin datos no hay evaluación de impacto. Registros administrativos enlazados, indicadores SMART definidos antes de poner en marcha la intervención, y una recogida de información planificada desde el principio. Si el sistema de indicadores y el grupo de comparación se piensan cuando el programa ya lleva tres años funcionando, casi siempre es tarde.

3. Lo cuantitativo y lo cualitativo

Las dos dimensiones responden preguntas distintas y ninguna sustituye a la otra. Lo cuantitativo mide cuánto, para quién y con qué grado de certeza. Lo cualitativo explica por qué y cómo, abre la caja negra de la implementación y detecta las historias que hay detrás, lo que no habíamos pensado medir o que es muy difícil

de captar con datos cuantitativos. La teoría del cambio es el puente entre lo cualitativo y lo cuantitativo: obliga a explicitar la cadena causal completa, desde los insumos hasta el impacto, y a decir qué supuestos deben cumplirse en cada eslabón. Un efecto nulo bien estimado no dice si el problema está en el diseño o en la ejecución; para eso hacen falta las dos miradas.

4. El ecosistema vasco: universidad y Besaldi

La alianza entre la universidad pública y el órgano de evaluación aporta lo que ninguna de las dos partes tiene por separado: independencia técnica y capacidad metodológica de un lado, conocimiento operativo y acceso a la información del otro. Tres rasgos que me gustaría ver consolidados. **Primero**, un acceso estable, seguro y protocolizado a los microdatos administrativos enlazados, con las garantías de confidencialidad correspondientes. **Segundo**, evaluabilidad desde el diseño: que ninguna política relevante se apruebe sin haber definido antes cómo se va a evaluar. **Tercero**, cultura de uso y de publicación, incluidos los resultados nulos o incómodos, porque también informan la decisión.

5. La evaluación en la inclusión y el empleo

Tres cosas en las que suelo centrar mis trabajos. **La heterogeneidad**: detrás de las cifras agregadas hay trayectorias muy distintas, y la media esconde a quienes peor están. **El acceso**: el diseño formal de una prestación y su alcance real difieren, y las barreras administrativas explican buena parte de esa distancia. Y **la brecha entre evaluar y decidir**, que la OCDE identifica como el principal reto de los sistemas de evaluación. La voz de las personas usuarias, expertas por experiencia, es evidencia, no decoración.