

DR. JOSÉ NABOR CRUZ MARCELO
SECRETARIO EJECUTIVO
CONEVAL
P R E S E N T E

Por este medio manifiesto *mi voluntad* para ser considerado candidato al CONEVAL como *investigador académico*. Para abordar "la importancia de la evaluación y la medición de la pobreza para el desarrollo de la política social y los desafíos para su consolidación", desarrollaré dos puntos. Para ahorrar espacio las citas a mi producción aparecen en el CV, y subrayo que no ignoro que algunas de mis propuestas requerirían cambios legales.

I. El estudio de las políticas sociales tiene un fuerte componente ideológico. A muchos científicos se les ha enseñado que su sentido se limita a reparar las "fallas de mercado", y que su evaluación es imprescindible por las, aún peores, "fallas de gobierno" (Weimer y Vining, 2005). Se parte así del supuesto de que el funcionario público realiza su trabajo deficientemente, y que más vale que sea sometido a evaluaciones externas realizadas por organismos autónomos. Esta ideología ha permeado en muchos profesionistas y agentes sociales y políticos.

Dicho lo anterior debemos señalar que la *evaluación externa*, y tanto o más importante, la *auto evaluación*, representan *piezas clave* de los programas sociales y de las políticas de desarrollo social (Aguilar, 2003a, 2003b, 2003c). Para *valorar* su papel, proponer algunas ideas para su *consolidación*, y *enfocar correctamente el tema de la medición* (ligado a la evaluación de impacto, entre otras, o evidentemente a la propia metodología multidimensional), propongo como telón de fondo un par de citas de Jerry Z. Muller (2018, *The Tyranny of Metrics*, Princeton University Press, Kindle):

"Used properly, measurement, as we'll see, can be a good thing... (But measurement) can also distort, divert, displace, distract, and discourage. While we are bound to live in an age of measurement, we live in an age of mismeasurement, over-measurement, misleading measurement, and counter-productive measurement... And measurement may provide us with distorted knowledge –knowledge that seems solid but is actually deceptive." (pos. 241-7)

"For if doctors or hospitals are remunerated by government agencies or private insurers based on their success rates in keeping patients alive, then such measurements should create incentives for better care... What could go wrong? A good deal, as we have already seen. When their scores are used as a basis of reward and punishment, surgeons, as do others under such scrutiny, engage in creaming, that is, they avoid the riskier cases. When hospitals are penalized based on the percentage of patients who fail to survive for thirty days beyond surgery, patients are sometimes kept alive for thirty-one days, so that their mortality is not reflected in the hospital's metrics. In England, in an attempt to reduce wait times in emergency wards, the Department of Health adopted a policy that penalized hospitals with wait times longer than four hours. The program succeeded—at least on the surface. In fact, some hospitals responded by keeping incoming patients in queues of ambulances, beyond the doors of the hospital, until the staff was confident that the patient could be seen within the allotted four hours of being admitted." (pos. 252-63)

Como muchos otros científicos Muller (2018) nos llama a cautela cuando medimos en las ciencias sociales en general, y para la *evaluación* de los programas sociales en particular. Al respecto añadiré cuatro consideraciones de fondo:

- Primera consideración. Los científicos debemos hacer una distinción entre la definición teórica de una variable, su *valor medido* y su valor verdadero; esto es, en primer lugar determinar el grado de correspondencia entre una definición teórica y una variable observada, y en segundo preguntarnos sobre la *calidad de la medición* a partir de la comparación del *valor observado*, que contiene, siempre, *errores y sesgos*, con el valor verdadero –el así llamado “*sesgo de medición*”. Por cierto, éste depende, recordemos, de la calidad del *instrumento de medición*.
- Segunda. Adicionalmente, y con mucha perspicacia, Boumans (2005, p. 121) advierte que un “problema relevante de los instrumentos utilizados para hacer visible lo no observable es cómo distinguir entre los hechos sobre el fenómeno y los artefactos creados por el instrumento”. Consecuentemente, es necesario tener en cuenta que *los instrumentos de medición –entre otros los “experimentos estadísticos” para realizar evaluaciones– dan forma a nuestro objeto de estudio*, y es mejor que los usemos con cautela y prudencia al momento de hacer recomendaciones sobre los programas y las políticas sociales.
- Tercera. Si bien recientemente los galardonados con el Nobel de Economía consideran que la realización de “experimentos” representa un mecanismo útil para *diseñar y evaluar* programas sociales, las ciencias sociales son, por definición, *no experimentales*. Al respecto puedo citar, si hace falta, la perspectiva de los Nobeles Christopher Sims y Angus Deaton. El meollo es que, epistemológicamente hablando, *la realización de inferencia* basada en “experimentos” es limitada. *Si minimizamos lo anterior nuestra perspectiva sobre la aportación de la evaluación para el mejoramiento de las políticas sociales será inapropiada*. Como ejemplo baste la siguiente cita del Nobel Sims (2010, p. 59):

“Without apparent irony, Angrist and Pischke quote Griliches (1986): ‘If data were perfect, collected from a well-designed randomized experiments, there would hardly be room for a separate field of econometrics.’ The fact is, economics is not an experimental science and cannot be. ‘Natural’ experiments and ‘quasi’ experiments are not in fact experiments... Providing empirically grounded policy advice necessarily involves confronting these difficult central issues. If applied economists narrow the focus of their research and critical reading to various forms of pseudo-experimental analysis, the profession loses a good part of its ability to provide advice about the effects and uncertainties surrounding policy issues.”

- Cuarta consideración. La aplicación de instrumentos estadísticos (simulaciones, diferencias en diferencias, modelos computacionales, etc.) y las *prácticas no científicas* van de la mano. El ejemplo más crudo lo escribieron Lequiller y Blades (2014, p. 44) en un libro publicado por la OCDE:

“National accounts could better be called ‘national accounts statistics’ because without this qualifier users may think they are as reliable as the business accounts of a company. This is not true. In particular, while GDP for technical reasons is often expressed in millions of units of the National currency, users should be aware that they are very, very far from being accurate at the level of millions... It is not even possible to give a summary figure of the accuracy of the GDP. Indeed, national accounts, and in particular GDP, are not the result of a single big survey for which one might compile a confidence interval. They are the result of combining a complex mix of data from many sources, many of which require adjustment to put them into a national accounts database and which are further adjusted to improve coherence, often using non-scientific methods.”

II. Por todo lo anteriormente dicho quiero afirmar lo siguiente.

En primer lugar, que las evaluaciones deben contribuir genuinamente al mejoramiento de los programas y las políticas, y que la medición *no* es un componente imprescindible en todos los casos. Baste recordar que los agentes ajustan su comportamiento a las (imperfectas) evaluaciones (externas) por lo que su efecto puede ser inocuo y, peor aún, contra productivo. Recomiendo *revisar al conjunto de metodologías para realizar los distintos tipos de evaluación acotando el peso de la medición y re-ponderando la cultura de la auto evaluación*. Si el Consejo tiene éxito en esta dirección entonces en el futuro: 1) en el caso de algunos programas cedería su papel como evaluador, por lo menos en algunas de sus modalidades, y 2) lo conservaría para el caso de otros programas y políticas.

En segundo lugar, en administraciones pasadas se organizó un mercado ligado a la realización de evaluaciones (a niveles nacional y de los estados) que, creo, generó una burbuja (de precios). Me parece clave cuidar cuántas evaluaciones se realizan, quiénes las elaboran, y cuánto se paga. *Cada peso gastado en una evaluación externa representa un peso no gastado en los programas sociales*.

En tercer lugar, la consolidación de la evaluación requiere *disponer de mejores instrumentos de medición, y corregir los disponibles al momento* (entre otros las propias metodologías para la medición de la pobreza). Algunos ejemplos.

- La Ley obliga al CONEVAL a medir la pobreza a nivel municipal. Pero el instrumento, la ENIGH, no es representativo a ese nivel. Si bien el Consejo ha puesto en práctica soluciones creativas, los modelos econométricos actualmente utilizados constituyen un área de oportunidad.
- Los precios utilizados para valorar las canastas son *no observados*. Debe solicitarse atentamente al INEGI que para el caso del Índice Nacional de Precios al Consumidor (IPC) aplique, como el BLS de los EEUU, una *Encuesta de Puntos de Compra* por deciles de gasto, para así contar, entre otros usos, con un punto de referencia contra el cual comparar los precios implícitos (Guerrero, 2017a). Subrayemos que los lugares en los que se compra son distintos entre deciles.
- Para actualizar los valores de las canastas (y por tanto de las líneas de bienestar), del ingreso corriente monetario y del poder adquisitivo del ingreso laboral, se utiliza al IPC. Muellbauer (1976, p. 32), coautor del Nobel Deaton (2007), nos recuerda que los números índices *no* son indicadores estadísticos "*neutrales*". El IPC depende de los ponderadores y de los propios precios que lo componen. El método que actualmente se utiliza para compilarlo se conoce como *plutocrático* ya que refleja el patrón de gasto de los deciles más ricos (Guerrero, 2017a). A este respecto Deaton (1998, p. 43) estimó que los ponderadores del IPC estadounidense reflejan al patrón de gasto de los hogares del percentil 75 de la distribución del gasto; en el caso de España el percentil aplicable es el 61 (Izquierdo, Ley y Ruiz-Castillo, 2003, p. 149), y para el IPC mexicano el percentil en cuestión es el 86 (Guerrero, 2010, p. 2). *En verdad resulta absolutamente inapropiado estimar el número de pobres utilizando un índice plutocrático*. Como solución el Nobel Arrow (1958) propuso la compilación de un índice de precios para cada decil. Guerrero y Lara (2011, 2012 y 2013) ya propusieron variaciones a las mediciones oficiales de pobreza utilizando otros índices de precios. *En el mismo sentido el Banco de México no protege el*

poder de compra de todos, sino en primer lugar de los deciles más ricos (Guerrero, 2011). Debe solicitarse atentamente al INEGI que compile siguiendo a Arrow (1958), o a Prais (1959) –quien propuso una agregación *democrática* en la que *no* cada peso gastado sino *cada familia representa un voto*.

- La *cohesión social* tiene un papel menor en la metodología de la medición multidimensional. La microeconomía que la inspiró no incorpora adecuadamente el vínculo entre los individuos y la sociedad. Existe teoría económica que afirma que no existe independencia entre las preferencias de los consumidores (Duesenberry, 1962), lo que implica que el nivel de bienestar individual depende tanto de la canasta adquirida por el individuo como por las adquiridas por el decil al que pertenece y por otros deciles. *Adicionalmente el Gini de ingreso apenas representa la punta del iceberg para analizar la cohesión social*. Los países que miden la *distribución de la riqueza* reportan Ginis que ascienden a entre 0.6 y 0.8. En México no disponemos de esa valiosa información. Ante esta limitación, con base en las ENIGH, específicamente utilizando algunos bienes riqueza, Guerrero (2015a, 2015b, 2012) obtuvo Ginis de riqueza de alrededor de 0.75. En EEUU la FED levanta dos encuestas al año para medir el nivel y la distribución de la riqueza. Creo que el Consejo deberá poner en la mesa de la discusión pública la necesidad nacional de contar con esa información.

Aguilar, L.F. (2003a, 2003b, 2003c), *El Estudio de las Políticas Públicas, La Hechura de las Políticas, La Implementación de las Políticas*, estudios introductorios y edición, Miguel Ángel Porrúa, terceras ediciones, primeras reimpressiones.

Arrow, K.J. (1958), *The measurement of price changes, en Relationship of Prices to Economic Stability and Growth: Hearings before the Joint Economic Committee Congress of the United States*, U. S. Government Printing Office.

Boumans, M. (2005), *How Economists Model the World into Numbers*, Routledge.

Deaton, A. (1998), Getting prices right: What should be done?, *Journal of Economic Perspectives*, 12(1), pp. 37-46.

Deaton, A. y Muellbauer, J. (2007), *Economics and Consumer Behavior*, CUP, 22^{da} edición. Primera edición 1980.

Duesenberry, J.S. (1962), *Income, Saving and the Theory of Consumer Behavior*, CUP.

Izquierdo, M., Ley, E. y Ruiz-Castillo, J. (2003), The plutocratic gap in the CPI: Evidence from Spain, *IMF Staff Papers*, 50(1).

Lequiller, F. y Blades, D. (2014), *Understanding National Accounts*, second edition, revised and expanded, OECD Publishing.

Muellbauer, J. (1976), The political economy of price indices, *Birbeck Discussion Paper*, número 22.

Prais, S.J. (1959), Whose cost of living?, *The Review of Economic Studies*, 26(2), pp. 126-34.

Sims, Ch.A. (2010), But Economics is not an experimental science, *Journal of Economic Perspectives*, 24(2), pp. 58-68.

Weimer, D.L. y Vining, A.R. (2005), *Policy Analysis: Concepts and Practice*, cuarta edición, Pearson Prentice-Hall.


Dr. Carlos Guerrero de Lizardi