

Ámbitos críticos de los programas sociales en Chile 2001-2020

Brigitte Ortiz-Castro ^a 

Cristian Fuentes-Tobar ^a 

Fernanda Castañeda-Zúñiga ^a 

^a *Universidad de Chile*

Resumen | Diversas instituciones internacionales han valorado positivamente el sistema de evaluación de programas gubernamentales desarrollado en Chile. No obstante, aspectos de carácter técnico e institucional siguen representando un desafío para este sistema, sumado a la problemática respecto a la baja evaluación de desempeño que están recibiendo los programas. Este artículo presenta los resultados de una investigación que tuvo como propósito identificar, mediante una revisión sistemática a las Evaluaciones de Programas Gubernamentales realizadas en Chile durante los años 2001-2020, los ámbitos críticos de los programas sociales en su conjunto y a nivel sectorial, considerando que no poner atención a dichos aspectos podría ser un obstáculo tanto para la toma de decisiones de mejora y aprendizaje continuo sobre los programas, como también para la realización de evaluaciones de mejor calidad. La metodología utilizada fue el Análisis de Contenido Cuantitativo. Los hallazgos de esta investigación sugieren falencias prevalientes a nivel global en los ámbitos de resultados, diseño y coordinación institucional en los programas sociales, tendencias que en general se mantienen a nivel sectorial, además de evidenciarse problemáticas específicas.

Palabras clave: Programas sociales, DIPRES, Evaluación de Programas Gubernamentales.

Critical areas of social programs in Chile 2001-2020

Abstract | Several international institutions have positively assessed the evaluation system of government programs developed in Chile. However, technical and institutional aspects continue to represent a challenge for this system, in addition to the problem of the low performance evaluation that the programs are receiving. This article corresponds to a presentation of the results of a research aimed at identifying, through a systematic review of the Government Program Evaluations carried out in Chile during the years 2001-2020, the critical areas of social programs as a whole and at the sectoral level. This is based on the consideration that not paying attention to these aspects would constitute an obstacle, both for making decisions for improvement and continuous learning about the programs, as well as for carrying out better quality evaluations. The methodology used was Quantitative Content Analysis. The findings of this research suggest prevailing shortcomings at the global level in the areas of results, design and institutional coordination in social programs, trends that are generally maintained at the sectoral level, in addition to highlighting specific problems.

Keywords: Social programs, DIPRES, Evaluation of Government Programs.

Este trabajo tiene el patrocinio del Núcleo Interdisciplinario de Investigación Evaluativa orientada a la Decisión Pública (NIIE-DP) de la Universidad de Chile.

Dirección de correspondencia [Correspondence address]:

Brigitte Ortiz-Castro, Universidad de Chile

E-mail: brigitte.ortiz@ug.uchile.cl



1. Introducción

DIVERSAS instituciones supranacionales, como el Banco Mundial (2005) y el Banco Interamericano de Desarrollo, en conjunto con el Centro Latinoamericano para la Administración del Desarrollo (Ríos, 2007), han analizado el sistema de evaluación chileno y lo han valorado positivamente, destacando el avance y la robustez de su institucionalidad y sus procedimientos, en comparación con el resto de la región. Dentro de los aspectos relevados se encuentran el mantenimiento y continuidad del sistema de evaluación a través del tiempo, el hecho que la evaluación se realice por paneles de expertos externos seleccionados mediante concurso público, y la disposición de los informes ejecutivos de las evaluaciones para el uso público, contribuyendo a la transparencia del sistema. No obstante, y sin desmedro de los logros señalados, existen aspectos técnicos e institucionales que, según estas evaluaciones, representan aún desafíos para el sistema de evaluación de programas chileno.

Dentro de los desafíos institucionales mencionados se encuentra la promoción de un compromiso más genuino y activo por parte de las instituciones y ministerios con las evaluaciones (Banco Mundial, 2005) y la generación de una base de datos unificada de las evaluaciones realizadas, en sus diversas etapas (i.e., *ex-ante*, *ex-dure* y *ex-post*), a políticas, planes y programas a nivel local, regional y nacional (Olavarría y Peroni, 2012). Estos desafíos apuntan a que existiría una desarticulación entre los actores institucionales. Por ejemplo, entre el Congreso y la labor evaluativa de la Dirección de Presupuestos (DIPRES), tanto en el trabajo conjunto para definir las áreas o programas a evaluar, como en el escaso uso de la información producida en la evaluación para la discusión sobre la Ley de Presupuestos (Sepúlveda, 2019). En el estudio de Olavarría (2012) se evidencia que diversos actores, tales como directores de programas y parlamentarios, suelen considerar a las evaluaciones que realiza la DIPRES como parte de un proceso formal más que como un insumo para la toma de decisiones. Dentro de los desafíos técnicos, en tanto, se encuentra la formación de capital humano avanzado en metodologías de seguimiento y evaluación de mayor complejidad (e.g., evaluación de impacto, diseño de grupos de control, perspectivas cualitativas, metodologías participativas, entre otras), y la incorporación del seguimiento y evaluación en las etapas de diseño y formulación de políticas y programas con el objetivo de incorporar el levantamiento de líneas base de información de calidad, las que representan un insumo fundamental para la evaluación (Olavarría y Peroni, 2012), y cuya ausencia repercute negativamente en la calidad intrínseca de las evaluaciones y los fundamentos de sus conclusiones (Banco Mundial, 2005). Otras evaluaciones (Haefner y Villegas, 2019; Olavarría, 2012) han hecho hincapié en las falencias que presentan los diseños de marco lógico en cuanto a criterios de focalización e indicadores que permitan constatar el cumplimiento de objetivos y verificación de resultados de los programas.

A partir del año 2000, la DIPRES resolvió instalar por primera vez clasificaciones de rendimiento para los programas tras su evaluación. Respecto a estas categorías de clasificación (ver Tabla 1), durante la década

de los 2000 esta institución utilizó una clasificación en función de la magnitud de los cambios que requerían los programas evaluados. Mientras que a partir del año 2011, y hasta la fecha, una nueva clasificación estableció la importancia del desempeño de los programas en relación a sus resultados, lo que fue planteado como un cambio de foco del Sistema de Evaluación hacia el análisis de costo-efectividad (DIPRES, 2012), estableciéndose una diferencia cualitativa entre las clasificaciones. De esta forma, la evaluación transita desde una mirada que privilegia el carácter formativo, o de retroalimentación de la evaluación, a una con primacía y énfasis en la eficiencia.

Según la clasificación de las Evaluaciones de Programas Gubernamentales de DIPRES (EPG), de los 328 programas evaluados entre 2001-2020 (ver Figura 1), el 75,9% presenta un bajo desempeño global. Es decir, 249 programas son clasificados en una de las siguientes categorías: desempeño bajo, desempeño insuficiente, desempeño medio, mal desempeño, modificaciones en el diseño y/o procesos de gestión interna, rediseño sustantivo, finalización o reemplazo íntegro, resultados no demostrados y reubicación institucional. Mientras que sólo un 24,1% de los programas, es decir, 79 de los 328 presentan un desempeño global suficiente.

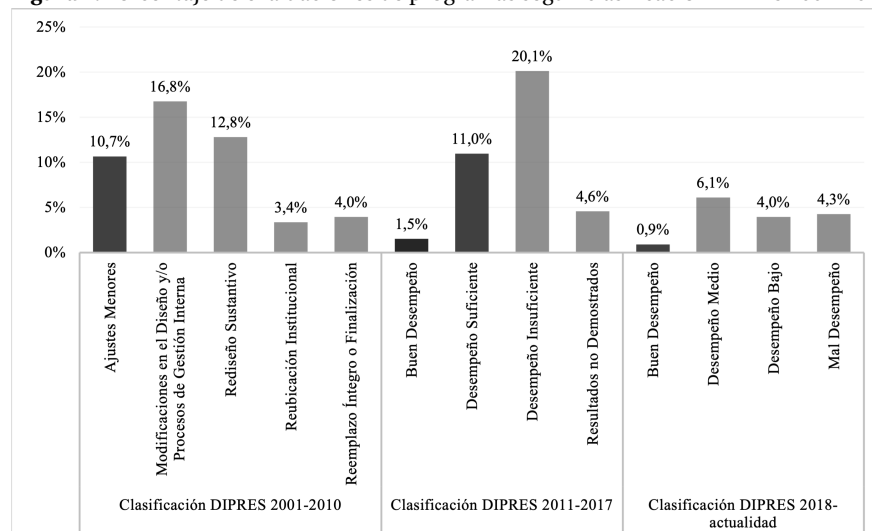
El cambio de perspectiva en la clasificación a partir de 2011 significó una importante tensión entre la evaluación de los programas y su función. Esto porque los programas que son establecidos por leyes permanentes o que atienden necesidades relevantes difícilmente pueden eliminarse o disminuir su presupuesto, a pesar de la evaluación de desempeño que se les otorgue (DIPRES, 2020). De esta forma, la evaluación no puede prescindir de su carácter formativo para corregir las falencias de los programas. Esto se basa en el hecho que, aunque una política pública tenga deficiencias no significa que no sea un beneficio para aquella población que recibe sus productos-servicios (Sepúlveda, 2019). Por lo tanto, si se entiende a la evaluación como información clave para generar un proceso de retroalimentación que permita entender las políticas implementadas y con ello generar aprendizajes para el debate de los nuevos temas de agenda pública, el diseño de nuevos instrumentos de política pública o, simplemente, una mejor implementación de ellos, como señalan Irrázaval et al. (2020), parece necesario aprovechar de mejor manera los resultados de la evaluación para la toma de decisiones en sus distintos niveles.

De acuerdo a lo anteriormente expuesto, y atendiendo especialmente a la problemática suscitada respecto a la baja clasificación de desempeño global que la DIPRES le ha otorgado a los programas, se realizó una investigación que tuvo como propósito identificar, mediante una revisión sistemática a las Evaluaciones de Programas Gubernamentales realizadas en Chile durante los años 2001-2020, los ámbitos críticos de los programas sociales en su conjunto y a nivel sectorial. Los resultados que se presentan en este artículo tienen la finalidad de generar un aporte al conocimiento sobre las razones que estarían incidiendo en la baja evaluación de desempeño de los programas. Esto basado en que no atender dichos aspectos constituiría un obstáculo, tanto para la toma de decisiones de mejora y

Tabla 1: Categorías de clasificación de las evaluaciones de la DIPRES según período

Clasificación DIPRES 2000-2010	Clasificación DIPRES 2011-2017	Clasificación DIPRES 2018-actualidad
Ajustes menores	Buen desempeño	Buen desempeño
Modificaciones en el diseño y/o procesos de gestión interna del programa	Desempeño suficiente	Desempeño medio
Rediseño sustantivo del programa	Desempeño insuficiente	Desempeño bajo
Reubicación institucional	Resultados no demostrados	Mal desempeño
Finalización o reemplazo íntegro del programa		

Fuente: Elaboración propia.

Figura 1: Porcentaje de evaluaciones de programas según clasificación DIPRES 2001-2020

Fuente: Elaboración propia en base a datos de la DIPRES.

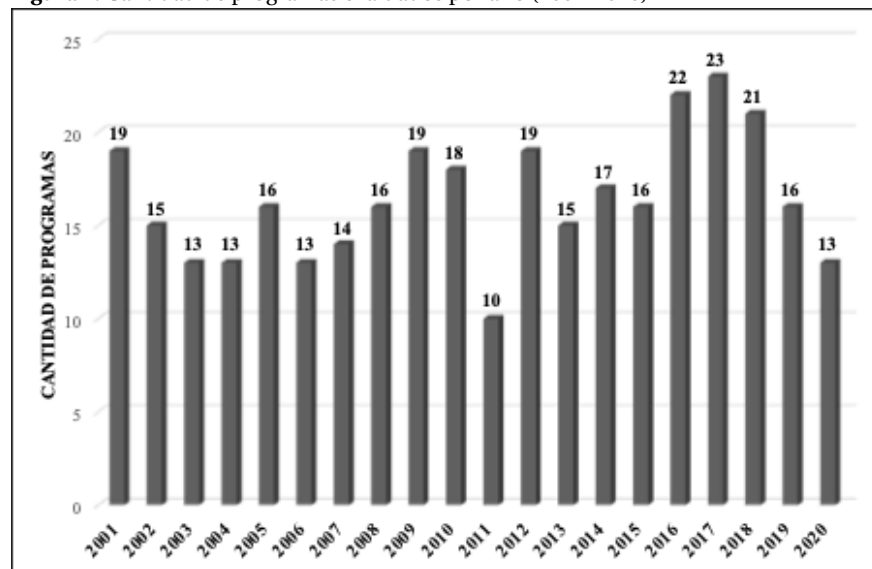
aprendizaje continuo sobre los programas, como también para la realización de evaluaciones de mejor calidad. Estudios previos han resumido la información disponible en base a las clasificaciones de desempeño definidas por la DIPRES (Libertad y Desarrollo, 2020), sin profundizar en los hallazgos sobre aspectos de diseño, implementación, eficiencia y resultados de los programas. Haefner y Villegas (2019) realizan un análisis sobre las falencias tras la clasificación de desempeño que afecta a los programas evaluados en la línea EPG en las áreas de Salud y Seguridad. En línea con aquella investigación, este artículo se propone avanzar hacia la búsqueda de los ámbitos de evaluación más problemáticos o críticos en los programas sociales en Chile, con el fin de aportar información, desde una mirada más formativa, que permita aprovechar de mejor manera la experiencia y el aprendizaje de las EPG en Chile.

1.1. Antecedentes de la Evaluación de Programas Gubernamentales

La evaluación de programas sociales y no sociales en Chile comenzó a gestarse en la década de los noventa con la definición de planes estratégicos e indicadores

de desempeño por parte de las instituciones públicas (DIPRES, 2008; Irrarázaval et al., 2020). Desde 1997 se implementó la EPG, que se enfoca en la evaluación de aspectos relacionados con “(...) la consistencia del diseño del programa, aspectos de su organización y gestión, y sus resultados a nivel de producto y de objetivos finales” (DIPRES, 2020: 5). A comienzos de la década del 2000 se promulgaron las normativas que institucionalizan la obligación de realizar evaluaciones de forma periódica a través del Sistema de Evaluación y Control de Gestión, facultando a la DIPRES para “(...) efectuar evaluaciones de los programas sociales, de fomento productivo y de desarrollo institucional incluidos en los presupuestos de los servicios públicos que se determinen mediante uno o más decretos del Ministerio de Hacienda” (DIPRES, 2015: 7).

En Chile se entiende como programa al “(...) conjunto de actividades necesarias, integradas y articuladas que proveen bienes y/o servicios (productos), tendientes a lograr un objetivo específico en una población determinada, de modo de resolver un problema o atender una necesidad que la afecta” (DIPRES, 2015: 9). Un programa se debe diseñar bajo el marco general de una política pública promovida por el Estado (Peroni, 2014), siendo una forma en que la política pública

Figura 2: Cantidad de programas evaluados por año (2001-2020)

Fuente: Elaboración propia en base a datos de la DIPRES.

se materializa (DIPRES, 2015). En términos operativos, se distingue entre programas sociales y no sociales de acuerdo con la clasificación funcional del Gasto, metodología de finanzas públicas que categoriza el propósito de las transacciones del Estado (DIPRES, 2021), de manera que los programas sociales corresponden a programas implementados por instituciones clasificadas en el ámbito social del gasto y los programas no sociales como los programas implementados por instituciones que no se encuentran clasificadas en el ámbito social (DIPRES, 2017).

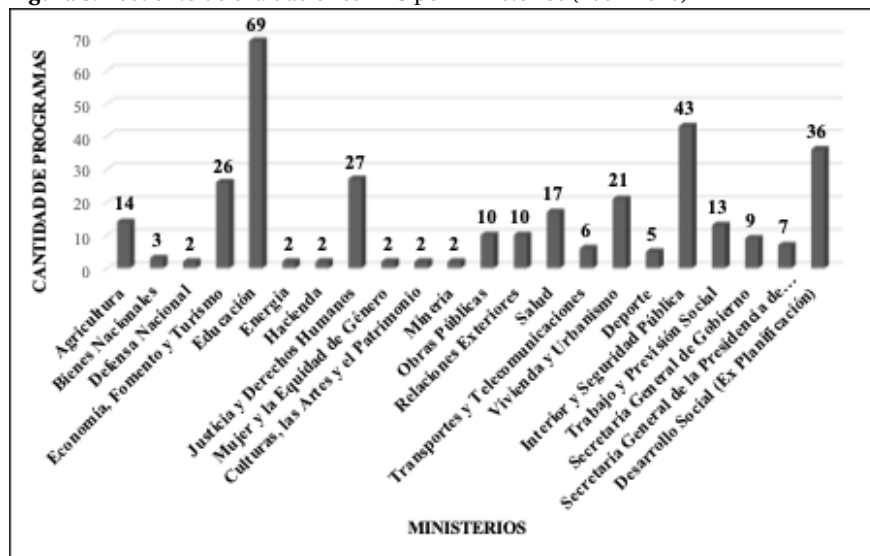
Entre los años 2001-2020 la línea EPG evaluó 328 programas sociales y no sociales. Respecto de la distribución por año de dichas evaluaciones, los datos muestran que la cantidad de programas evaluados fluctúan entre los 10 y los 23, con un promedio de 17 evaluaciones por año.

Como se muestra en la Figura 2, los años que presentaron la mayor cantidad de programas evaluados son 2016, 2017 y 2018 (con 22, 23 y 21 evaluaciones), seguidos por los años 2001, 2009 y 2012 (con 19 evaluaciones). Mientras que los años con menor cantidad de programas evaluados son 2003, 2004, 2006, 2011 y 2020 (con 13 evaluaciones, a excepción del año 2011 que presenta solo 10 evaluaciones). Respecto de los años 2011 y 2020, es necesario destacar que la baja cantidad de programas evaluados podría explicarse, en parte, debido a factores contextuales que obligaron al Estado chileno a cambiar sus prioridades presupuestarias. Esto porque en el año 2010, Chile fue sacudido por un terremoto de grandes proporciones, y en el año 2020 se vio enfrentado a las consecuencias políticas, económicas y sociales derivadas del denominado "estallido social" y la posterior pandemia del COVID-19.

En relación con la distribución de la cantidad de programas evaluados bajo la línea EPG según ministerios, los datos muestran que entre los años 2001-2020

los programas sociales y no sociales evaluados fluctúan entre 2 y 69. Como se muestra en la Figura 3, los ministerios que presentaron la mayor cantidad de programas evaluados en el periodo señalado son los de Educación, del Interior y Seguridad Pública, y de Desarrollo Social y Familia (con 69, 43 y 36 programas evaluados). Mientras que los que presentaron la menor cantidad de programas evaluados son los ministerios de Minería, de las Culturas, las Artes y el Patrimonio, de la Mujer y la Equidad de Género, de Hacienda, de Energía, y de Defensa Nacional (con 2 programas evaluados cada uno). Respecto de los ministerios con la menor cantidad de programas evaluados cabe señalar que el de las Culturas, las Artes y el Patrimonio, el de la Mujer y Equidad de Género, y el de Energía, fueron creados recientemente (2018, 2016 y 2010 respectivamente), por lo que podría explicarse en parte la baja cantidad de programas. No obstante, resulta preocupante que el ministerio de la Mujer y Equidad de Género presente tan baja cantidad de programas evaluados, puesto que en los últimos años Chile ha sido testigo de un aumento sostenido de la denuncia de violencia en contra de las mujeres y las disidencias sexuales, así como también un aumento de las denuncias de abusos laborales y sexuales según consigna la Dirección del Trabajo (2019).

Respecto del tipo de programas (i.e., sociales, no sociales) evaluados bajo la línea EPG en el periodo 2001-2020, los datos muestran un mayor número de evaluaciones de programas sociales (189 evaluaciones) respecto de los programas no sociales (139 evaluaciones). Como muestra la Figura 4, durante este periodo la mayor cantidad de evaluaciones de programas sociales fueron realizadas en los años 2009, 2014, 2016 y 2017 (con 13 evaluaciones en 2009 y 2014, y con 18 y 19 evaluaciones en 2016 y 2017), mientras que la menor cantidad se realizó en los años 2004, 2005, 2006 y 2007 (con 5, 4, 3 y 6 evaluaciones respectivamente). A su vez,

Figura 3: Recuento de evaluaciones EPG por ministerios (2001-2020)

Fuente: Elaboración propia en base a datos de la DIPRES.

la mayor cantidad de evaluaciones de programas no sociales fueron realizadas los años 2005, 2006, 2010 y 2012 (con 12, 10 y 11 respectivamente), mientras que la menor cantidad se presentó en los años 2011 y 2020 (con 1 y 3 evaluaciones).

2. Metodología

El diseño de la investigación realizada corresponde a un estudio cuantitativo, observacional o no experimental, pues pretende dar cuenta de la una realidad dada y sin la manipulación de sus condiciones de existencia, y de alcance descriptivo. La técnica de investigación utilizada es el Análisis de Contenido Cuantitativo (ACC). El ACC implica la producción de información a partir de la codificación de la información proveniente de fuentes primarias o secundarias, con el fin de cuantificar el contenido manifiesto o latente de ellos (Cea, 1999). Se destaca por el carácter sistemático de sus procedimientos y por su orientación a la objetividad, al requerir del acuerdo de más de una persona. Así mismo, el ACC contempla el tratamiento estadístico y la interpretación de los datos producidos, lo que corresponde a la descripción univariante de los ámbitos de los programas presentados como resultados en este artículo.

Las fuentes de información la constituyeron los resúmenes ejecutivos de las evaluaciones, llevadas a cabo por el departamento de Sistema de Evaluación y Control de Gestión de la DIPRES y disponibles en línea.

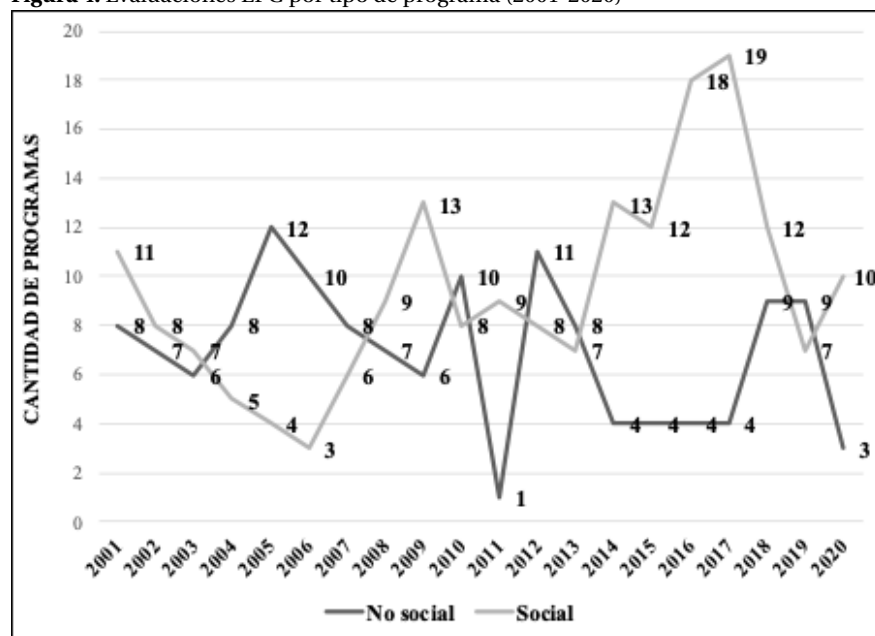
La muestra estuvo compuesta por 146 EPG seleccionadas a partir de 3 criterios de inclusión que se sintetizan en el diagrama PRISMA de Moher et al. (2009) que se muestra en la Figura 5. El primer criterio de inclusión consistió en seleccionar evaluaciones con resumen ejecutivo. El segundo criterio de inclusión consideró la selección de las evaluaciones de programas sociales. El tercer criterio de inclusión fue la selección

de las evaluaciones clasificadas como regulares o deficientes.

La investigación se centró en el análisis de las conclusiones de desempeño global y recomendaciones de mejora presentes en los resúmenes ejecutivos de las evaluaciones, indicados para distintos ámbitos (dimensiones y/o subdimensiones) de los programas sociales en la evaluación.

El equipo de investigación llevó a cabo la codificación de falencias presentes en los programas, asignando código 0 (ausencia de falencias) y 1 (presencia de falencias) a cada subdimensión analizada. Estas subdimensiones fueron agrupadas en cuatro grandes grupos o dimensiones generales: Diseño, Implementación, Gestión y Resultados. En total fueron analizadas 18 subdimensiones, éstas son presentadas en la Tabla 2.

La fiabilidad del análisis y los procedimientos de medida empleados, fueron puestos a prueba a través de un pretest que consideró el cálculo de un índice Kappa de Fleiss, siguiendo el método o medida de concordancia entre observadores (López de Ullibarri y Pita-Fernández, 1999), en este caso entre el equipo de investigación. El índice toma valores entre 0 y 1, donde 0 indica ausencia de concordancia y 1 completa concordancia, considerándose insuficientes valores <.40, aceptables valores >.40, y excelentes >.75 (Banerjee et al., 1999). Se utilizaron 19 evaluaciones que equivalen al 13% de la muestra, lo cual está dentro del porcentaje recomendado (i.e., entre 10% y 20%) para este tipo de análisis (Neuendorf, 2011). Los resultados indican que 9 de las 18 subdimensiones presentaron un índice aceptable, 6 presentaron un índice excelente, mientras que 3 presentaron valores insuficientes. Este último es el caso de las subdimensiones Falta de información a nivel de implementación, Evaluaciones anteriores DIPRES y Falta de información a nivel de resultados. En los casos de las últimas dos subdimensiones

Figura 4: Evaluaciones EPG por tipo de programa (2001-2020)

Fuente: Elaboración propia en base a datos de la DIPRES.

se ha considerado que el índice se comporta de forma incoherente, ya que se presenta un gran acuerdo entre los observadores (un 95% y 90% respectivamente). De acuerdo con Falotico y Quatto (2015) en situaciones de fuerte acuerdo entre calificadores, el índice Kappa podría asumir valores más bajos de lo esperado. Respecto a la subdimensión Falta de información a nivel de implementación, dado que presentó un bajo índice Kappa de Fleiss y un bajo acuerdo (47%), se decidió no considerarla en los resultados.

3. Resultados

3.1. Ámbitos críticos de los programas sociales a nivel global: resultados, diseño y coordinación institucional

El análisis de los ámbitos críticos a nivel global, –que incluye al conjunto de programas sociales en el período de estudio–, demuestra predominancia de los problemas centrados en la falta de información a nivel de resultados, y en los problemas relacionados al diseño de las intervenciones (ver Tabla 3).

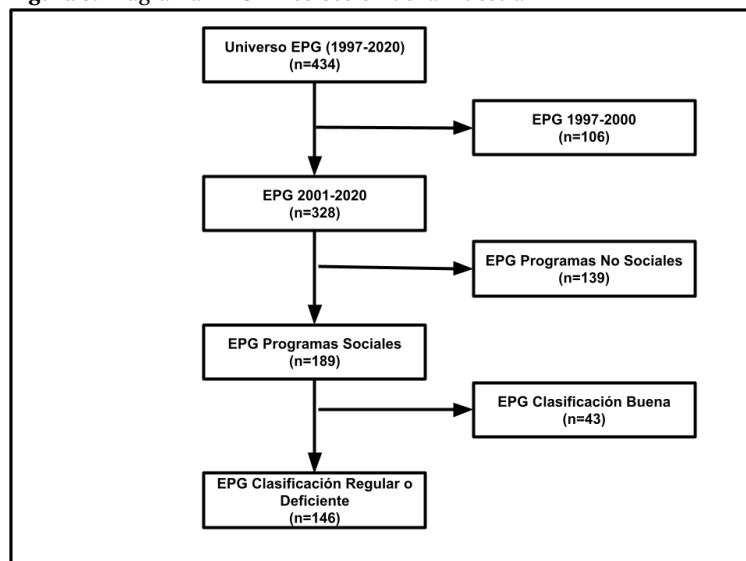
En este sentido, más de 9 de cada 10 programas evaluados presentaron evaluaciones que hicieron mención a la falta de información en aspectos como los resultados intermedios y finales, la satisfacción de los usuarios, cobertura, entre otros. Esta situación aparece asociada a su vez a otros elementos críticos a nivel de resultados, que tienen que ver con la falta de seguimiento y evaluación en los programas, ausencia de instrumentos de registro, sistemas de información y bases de datos, así como carencia de indicadores y metas. Este hallazgo permite dar cuenta de las dificultades que tienen los equipos que realizan las EPG, dado que

deben emitir juicios evaluativos con información muy limitada acerca del cumplimiento de los objetivos de los programas. A partir de lo anterior, es posible poner en cuestionamiento que la clasificación de desempeño de estos programas sea realmente deficiente, o bien recibir esta clasificación al no contar con la información necesaria para evaluar los resultados de los programas.

Por otra parte, entre 6 y 7 de cada 10 programas evaluados señalan problemas a nivel de diseño. En este caso, elementos críticos que parecen estar directamente relacionados con esta situación los conforman los problemas a nivel de marco lógico, asociados a deficiencias tanto en el diseño de componentes y actividades como la coherencia con los objetivos del programa (lógica vertical de marco lógico), deficiencias en el diseño y coherencia de metas, indicadores, fuentes de verificación y/o supuestos para cada actividad (lógica horizontal de marco lógico) y problemas de focalización. Este hallazgo refuerza en parte la falta de información mencionada en el párrafo anterior, ya que algunas de las falencias de diseño aluden a problemas en la definición de estrategias de producción de información que sirvan como fuente de verificación para las metas de los programas. Además, se evidencia que una de las debilidades más generalizada es la falta de estrategias para definir la población potencial y objetivo de la intervención, lo que repercute directamente en los juicios que se puedan hacer quienes evalúan respecto a la cobertura de los programas.

Por último, las falencias en la coordinación institucional como parte del ámbito de gestión de los programas sociales también presentan una alta presencia en el análisis global (58,9%). Evaluaciones realizadas a programas dependientes de distintos ministerios, como se verá, muestran esta problemática.

Figura 5: Diagrama PRISMA selección de la muestra



Fuente: Elaboración propia.

3.2. Ámbitos críticos de los programas sociales a nivel sectorial

Los resultados de las evaluaciones de los programas sociales a nivel sectorial (en este caso, ministerial) muestran algunas diferencias respecto al análisis global en términos de los aspectos críticos. No obstante, en general se mantienen las tendencias globales.

En el caso del Ministerio de Educación ($n = 48$) se evidencia que prácticamente la totalidad de los casos (97,9%) presenta problemas de falta de información a nivel de resultados. Asociado a ello se encuentra la falta de instrumentos de registro (77,1%), carencia de seguimiento y evaluación (56,3%), y ausencia de indicadores y metas (56,3%). Por otro lado, también se encuentran debilidades a nivel de diseño de las intervenciones. Asociados al diseño encontramos especialmente problemas en la lógica horizontal de marco lógico (79,2%), lógica vertical de marco lógico (66,7%) y problemas en la focalización (56,3%). Los problemas de focalización se encuentran asociados a definiciones que igualan la población potencial con la población objetivo del programa lo que señala la ausencia de criterios de focalización, como se señala en la evaluación del “Programa de Apoyo a la Retención Escolar” (Frías et al., 2018). En este sector además encontramos problemas asociados a falencias de coordinación institucional (54,2%).

En el caso de los programas implementados por el Ministerio de Desarrollo Social y Familia (y ex Planificación) ($n = 29$) se mantiene la tendencia de problemas relevantes relacionados a la falta de información acerca de resultados (93,1%), asociados principalmente a carencia de instrumentos de registro y bases de datos (65,5%), ausencia de indicadores y metas (62,1%) y falta de seguimiento y evaluación (55,2%). Por otra parte, se presentan relevantes deficiencias en diseño, en este caso asociado a problemas en la lógica verti-

cal del marco lógico (75,9%), seguido de deficiencias en la lógica horizontal de marco lógico (69%) y problemas de focalización (62,1%), aunque también aparecen como importantes los problemas a nivel de coordinación institucional (54,2%). En cuanto al último punto crítico, suele identificarse la falta de complementariedad entre programas, por ejemplo, en la evaluación del programa “4 a 7, Mujer trabaja tranquila” se señala que “(...) el programa está desarticulado de los otros programas institucionales, lo que reduce sus potenciales efectos, dificulta su ejecución y le resta coherencia en torno a los desafíos que se plantea” (Frías et al., 2015: 13).

En el Ministerio de Vivienda ($n = 15$), las evaluaciones también siguen la tendencia global, pues en un 93,3% de las evaluaciones se encontraron problemas de falta de información a nivel de resultados, asociados a la falta de seguimiento y evaluación (73,3%) y ausencia de indicadores y metas (60%). Los problemas más comunes en el ámbito de diseño están relacionados con la ausencia de un diagnóstico sistemático de la situación inicial que originó la existencia del programa (73,3%), lo que repercute en problemas en la coherencia de las matrices de marco lógico, tanto en su lógica horizontal (86,7%) como en su lógica vertical (73,3%), y en falencias de focalización (73,3%). En este ministerio también aparecen como relevantes los problemas de ejecución (60%), principalmente por no cumplimiento de los plazos de ejecución, y los problemas en la definición de gastos o de gestión del presupuesto (53,3%).

En el caso de los programas del Ministerio del Interior y Seguridad Pública ($n = 18$), destacan los problemas de falta de información a nivel de resultados (94,4%), asociado a la falta de seguimiento y evaluación (55,6%) y ausencia de indicadores y metas (50%). En diseño, los principales problemas se asocian en la lógica vertical del marco lógico (77,8%), lógica ho-

Tabla 2: Dimensiones y subdimensiones de evaluación

Dimensiones	Subdimensiones	Definición
Diseño	Focalización	Falencias en la definición de la población objetivo y/o potencial del programa.
	Ausencia de Marco Lógico	Ausencia de Matriz de Marco Lógico, necesidad de reconstrucción
	Lógica horizontal de Marco Lógico	Falencias en el diseño y coherencia de metas, indicadores, fuentes de verificación y/o supuestos.
	Lógica vertical de Marco Lógico	Falencias en el diseño y coherencia de componentes y actividades en relación con los objetivos y propósito del programa.
	Marco normativo	Limitaciones asociadas a leyes, decretos o reglamentos jurídicos-legales
	Falta de información a nivel de diseño	Ausencia de diagnóstico del problema, evaluación ex-ante o levantamiento de línea base o diagnóstico
Implementación	Recursos humanos	Falencias asociadas a la calidad o perfil de los profesionales, cantidad de profesionales, cargas laborales, tipo de contrato, etc.
	Ejecución	Falencias a nivel de instituciones ejecutoras, a nivel de la intervención o en su estructura organizacional.
	Seguimiento	Falencias en el seguimiento y monitoreo de la implementación del programa.
	Falta de información a nivel de implementación	Falencias asociadas a la falta de información asociada a aspectos de la implementación
Gestión	Coordinación institucional	Falencias a nivel de coordinación y comunicación entre instituciones sectoriales, ministeriales, gubernamentales; disponibilidad de información o seguimiento de la oferta pública.
	Definición de gastos/presupuesto	Falencias en la definición de los gastos y presupuesto de los programas.
	Falta de información a nivel de gestión	Falencias asociadas a la ausencia de registros de información o sistematización de los gastos y el presupuesto del programa.
Resultados	Falta de seguimiento y evaluación	Ausencia de seguimiento y/o evaluación (ex-dure, ex-post) del programa
	Falta de instrumentos de registro/sistemas de información, bases de datos	Falencias asociadas a la ausencia de instrumentos de registro, sistemas de información y/o bases de datos.
	Ausencia de Indicadores y metas	Ausencia o falencias en la definición de indicadores y metas de los componentes del programa.
	Evaluaciones anteriores DIPRES	No consideración de sugerencias realizadas en evaluaciones anteriores del mismo programa.
	Falta de información a nivel de resultados	Ausencia de información y/o falencias en la verificación de los resultados producidos por el programa.

Fuente: Elaboración propia.

rizonal de marco lógico (66,7%) y de focalización (50%). En este ministerio el 50% de los programas presentan falencias asociadas a la falta de información a nivel de gestión, es decir, carecían de información sistematizada de gastos.

En el caso del Ministerio de Salud ($n = 11$) todos los programas presentan como falencia la falta de información a nivel de resultados, especialmente asociados a la ausencia de registros e instrumentos (81,8%) y carencia de seguimiento y evaluación (54,5%). Sigue la tendencia global en cuanto a los problemas en el diseño, en la lógica vertical del marco lógico (72,7%) y lógica horizontal de marco lógico (54,5%). Por otra parte, se presentan como relevantes los problemas de coordinación institucional (72,7%). Al igual que en el Ministerio del Interior y Seguridad Pública, se presenta un 54,5% de programas con falta de información a nivel de gestión. En este ministerio se identifica que existe información, pero de manera no sistemática. Pa-

ra ejemplificar lo anterior, en el caso de la evaluación del programa “Vida Sana” se señala que el origen de los problemas de información se encuentra “(...) en el registro que se realiza en terreno durante la atención de los beneficiarios, ya que las planillas Excel carecen de campos precodificados, de modo de disminuir los errores de registro” (Daigre et al., 2016: 5).

En el caso del Ministerio de Justicia y Derechos Humanos ($n = 10$), la falta de información a nivel de resultados (90%) aparece asociada a la falta de seguimiento y evaluación (80%) y carencia de instrumentos de registro (80%). A diferencia del resto de los sectores y la tendencia global, en este ministerio aparecen más claramente los problemas asociados a la carencia de coordinación institucional (90%) y la falta de información a nivel de gestión (80%).

Tabla 3: Resumen falencias de los programas evaluados por ministerio y global (%)

Dimensión	Subdimensiones	Educación	Desarrollo Social	Vivienda	Interior y seguridad	Salud	Justicia	Trabajo	Deporte*	Mujer*	Agricultura*	Global
Diseño	Focalización	56,3	62,1	73,3	50,0	45,5	80,0	100,0	100,0	100,0	0,0	63,0
	Ausencia de Matriz ML	0,0	3,4	6,7	0,0	18,2	0,0	12,5	0,0	0,0	0,0	3,4
	Lógica horizontal ML	79,2	69,0	86,7	66,7	54,5	10,0	37,5	40,0	100,0	100,0	66,4
	Lógica vertical ML	66,7	75,9	73,3	77,8	72,7	40,0	62,5	100,0	100,0	100,0	70,5
	Marco normativo	18,8	6,9	20,0	5,6	9,1	70,0	25,0	60,0	0,0	0,0	19,2
	Falta de información a nivel de diseño	27,1	31,0	73,3	33,3	27,3	10,0	37,5	0,0	0,0	0,0	31,5
Implementación	Recursos humanos	35,4	41,4	20,0	44,4	45,5	100,0	37,5	40,0	0,0	0,0	41,1
	Ejecución	39,6	34,5	60,0	38,9	45,5	30,0	50,0	40,0	0,0	100,0	41,1
	Seguimiento	39,6	10,3	33,3	44,4	18,2	0,0	12,5	0,0	100,0	0,0	26,7
Gestión	Coordinación institucional	54,2	69,0	20,0	50,0	72,7	90,0	50,0	100,0	100,0	100,0	58,9
	Definición de gastos/presupuesto	29,2	27,6	53,3	5,6	0,0	0,0	25,0	100,0	0,0	100,0	26,7
	Falta de información a nivel de gestión	41,7	24,1	26,7	50,0	54,5	80,0	12,5	40,0	0,0	100,0	39,7
Resultados	Falta de seguimiento y evaluación	56,3	55,2	73,3	55,6	54,5	80,0	50,0	100,0	0,0	0,0	59,6
	Falta de instrumentos de registro/sistemas de información, bases de datos	77,1	65,5	46,7	44,4	81,8	80,0	37,5	60,0	100,0	100,0	65,8
	Ausencia de Indicadores y metas	56,3	62,1	60,0	50,0	45,5	30,0	62,5	100,0	100,0	100,0	56,8
	Evaluaciones anteriores DI-PRES	2,1	10,3	0,0	5,6	9,1	0,0	0,0	0,0	0,0	0,0	4,1
	Falta de información a nivel de resultados	97,9	93,1	93,3	94,4	100,0	90,0	100,0	100,0	100,0	100,0	95,9

Nota: *Ministerios con 5 o menos programas seleccionados.

Fuente: Elaboración propia.

En cuanto a diseño aparecen falencias asociadas a problemas de focalización (80%) y a limitaciones del marco normativo-legal (70%), que generan tensiones entre los enfoques de protección a la infancia y de prevención del delito, como se señala en la evaluación del programa “Vida Nueva” (Monreal et al., 2012). Finalmente, resulta relevante que todos los programas presentan falencias en cuanto a falta de recursos humanos o con mayor cualificación.

En el Ministerio del Trabajo ($n = 8$) se sigue la tendencia global, todos los programas tuvieron problemas de falta de información a nivel de resultados, asociados a ausencia de indicadores y metas (62,5%) y a falta de seguimiento y evaluación (50%). En la dimensión de diseño, todos los programas presentan problemas de focalización y el 62,5% de los programas presenta deficiencias en la lógica vertical de marco lógico. Al igual que en los ministerios de Vivienda y del Interior las falencias están asociadas a la ausencia de información e indicadores, derivadas de la falta de diagnósticos y construcción de líneas base. En este ministerio también se presentan problemas en ejecución de los programas (50%) y de falta de coordinación institucional (50%).

En el caso del Ministerio del Deporte ($n = 5$) las 5 evaluaciones realizadas en el año 2017 fueron clasificadas con la categoría “Mal desempeño” por la DIPRES, este análisis muestra que todos los programas evaluados tienen problemas en diseño del marco lógico (tanto en la lógica horizontal como vertical), falta de información a nivel de resultados, pero además en la coordinación institucional y la definición de presupuestos. En este sentido, los problemas de eficiencia se pueden sintetizar en el juicio de la evaluación al Fondo Nacional para el Fomento del Deporte donde se señala que “(...) se identifica que algunas actividades/componentes están siendo cubiertas por más de una línea de financiamiento, situación que podría indicar una duplicidad e ineficiencia en el uso de los recursos” (Marcelo et al., 2017: 24). En este ministerio también se destacan tensiones entre el diseño de los programas con el marco normativo (60%) que manda el fomento del deporte recreativo y el apoyo al deporte de alto rendimiento en una misma intervención.

Por último, en el Ministerio de la Mujer y Equidad de Género y el Ministerio de Agricultura sólo 1 caso fue seleccionado para cada uno. En el caso del “Programa Prevención Integral de Violencias contra las Mujeres” los problemas también siguen la tendencia a nivel de diseño y resultados, pero además se identifican problemas a nivel de coordinación institucional y seguimiento. En el caso del “Programa de Agricultura”, además de problemas a nivel de diseño y falta de información a nivel de resultados, también se identificaron problemas en la coordinación institucional, presupuesto y la implementación.

4. Conclusiones

4.1. Ámbitos críticos de los programas sociales

Los resultados sugieren falencias prevalecientes a nivel global en los ámbitos de resultados, diseño y coordinación institucional, tendencias que en general

se mantienen a nivel sectorial. En términos de diseño, las falencias se encuentran relacionadas principalmente a problemas a nivel de marco lógico (tanto en el diseño y coherencia de los componentes y actividades con los objetivos del programa, como en el diseño y coherencia de metas, indicadores, fuentes de verificación y supuestos) y a nivel de focalización.

En relación con el ámbito de resultados, los problemas se asocian a la falta de registros, sistemas de información y bases de datos, carencia de seguimiento y evaluación constante, y ausencia de indicadores y metas. Ello en concordancia con otros estudios citados. De acuerdo con el análisis realizado, y en relación con el ámbito de diseño, es recomendable que la formulación de los programas pueda detenerse con mayor interés en los aspectos de focalización, definición de indicadores y levantamiento de líneas base. En el caso específico de los programas asociados al Ministerio de Vivienda destaca además la falta de información a nivel de diseño, como una situación crítica en este sector.

Con relación a resultados, en tanto, se sugiere incorporar la modernización en la gestión de datos, incluyendo mejoras tanto en el ámbito de producción como de sistematización de la información. Cabe destacar la importancia que tiene entre los programas asociados a los Ministerios de Educación, Desarrollo Social y Familia y Ministerio de Vivienda y Urbanismo, la ausencia de indicadores y metas, como una característica para poner atención en estos tres sectores de tanta relevancia.

Transversalmente, aparece el problema de la coordinación institucional entre los servicios que implementan los programas evaluados, lo que se asocia a la revisión de la oferta pública ante programas muy similares, que atienden a una misma población objetivo o, por otra parte, en relación con la necesidad de una labor conjunta de distintos servicios interministeriales para aumentar la capacidad de abordar integralmente los problemas sociales atendidos.

Por último, las falencias en la dimensión de implementación de los programas son menos frecuentes (menos del 50%) a nivel global y a nivel sectorial, con excepción de los programas del Ministerio de Justicia y Derechos Humanos, donde los problemas aparecen asociados a la falta de estandarización de los procedimientos en las residencias del Servicio Nacional de Menores (SENAME). En esa línea, y como se ha señalado en estudios anteriores (Peroni y Páez, 2019), muchas veces las recomendaciones presentadas en las evaluaciones no pueden ser tomadas en cuenta, porque escapan a las posibilidades de los ejecutores, o porque tienen que ver con problemas más estructurales como el presupuesto, y aspectos de carácter normativo que condicionan a los programas vía ley o decretos.

4.2. Ámbitos críticos en los programas como posibles obstáculos para las evaluaciones

La identificación del diseño, los resultados y la coordinación institucional como ámbitos críticos de los programas permite dar cuenta de algunos obstáculos en los procesos de evaluación. Así como fue mencio-

nado en la introducción, diversas evaluaciones al Sistema de Evaluación chileno han destacado considerables logros en su funcionamiento (e.g., su permanencia en el tiempo, su sistema de elección de jueces a través de concurso público y la disposición de los informes de evaluación de forma pública), así como también han planteado aspectos que conforman desafíos para el mejoramiento del proceso evaluativo, como la cooperación y el compromiso interinstitucional a lo largo del ciclo de evaluación (desde la selección de los programas a evaluar hasta la valoración y el uso de sus resultados); y, por otra parte, desafíos técnicos relativos a la producción de líneas base y evaluaciones de mayor complejidad.

De acuerdo con el análisis realizado, los equipos de evaluación deben lidiar con la falta de información a nivel de resultados, porque carecen de información sistematizada sobre el desarrollo de la intervención, ya sea por falta de seguimiento, ausencia de sistemas de registro o medios de verificación de los indicadores y metas del programa. Como muestran los análisis a nivel global y a nivel sectorial, estas falencias están estrechamente relacionadas con el diseño de los programas, que no cuentan con diagnósticos sistemáticos de los problemas que se pretenden intervenir, y muestran dificultades o ausencia de criterios de focalización, imposibilitando la construcción de líneas base con las cuales comparar los resultados de los programas. La coordinación institucional también juega un rol en los problemas asociados a la falta de complementariedad de información interministerial o la falta de códigos comunes internos, como es el caso de la información con la que cuentan los programas del Ministerio de Salud.

Se considera que la constatación de estas falencias podría resultar en la formulación de juicios evaluativos parciales o no concluyentes al respecto de los logros y resultados de los programas, pudiendo explicar las altas cifras de la baja evaluación de desempeño de los programas. Es por esto que resulta necesario cuestionar la perspectiva centrada en el desempeño que prevalece desde 2011, cuando la DIPRES cambia su clasificación centrada en el aprendizaje, transitando a un enfoque centrado en el costo-eficiencia, pues los problemas metodológicos expuestos permiten dudar de la clasificación, que reciben los programas, basada en el desempeño.

Referencias

- Banco Mundial (2005). *Chile: estudio de evaluación de impacto del Programa de Evaluación de Programas, Evaluaciones de Impacto y Evaluaciones de los Programas del Gobierno. Resumen Ejecutivo*. Unidad de Reducción de la Pobreza y Gestión Económica. En línea: [enlace](#).
- Banerjee, M., Capozzoli, M., McSweeney, L., y Sinha, D. (1999). Beyond kappa: A review of interrater agreement measures. *Canadian Journal of Statistics*, 27(1):3–23. DOI: [10.2307/3315487](#).
- Cea, M. A. (1999). *Metodología cuantitativa: estrategias y técnicas en investigación social*. Madrid: Síntesis Sociológica.
- Daigre, M., Ferrer-Lues, M., y Heynig, E. (2016). *Vida Sana - Intervenciones en Factores de Riesgo de Enfermedades Crónicas. Resumen Ejecutivo Evaluación de Programas Gubernamentales*. En línea: [enlace](#).
- DIPRES (2008). *Informe de Finanzas Públicas 2009*. Santiago de Chile: Ministerio de Hacienda. En línea: [enlace](#).
- DIPRES (2012). *Informe de Finanzas Públicas 2013*. Santiago de Chile: Ministerio de Hacienda. En línea: [enlace](#).
- DIPRES (2015). *Evaluación Ex-Post: Conceptos y Metodologías*. Santiago de Chile: División de Control de Gestión. En línea: [enlace](#).
- DIPRES (2017). *Informe de Finanzas Públicas 2018*. Santiago de Chile: Ministerio de Hacienda. En línea: [enlace](#).
- DIPRES (2020). *Resultados evaluaciones 2020*. Santiago de Chile: Ministerio de Hacienda. En línea: [enlace](#).
- DIPRES (2021). *Estadísticas de las Finanzas Públicas 2011 - 2020*. Santiago de Chile: Ministerio de Hacienda. En línea: [enlace](#).
- Dirección del Trabajo (2019). *Novena Encuesta Laboral (ENCLA)*. Santiago de Chile: Departamento de Estudios. En línea: [enlace](#).
- Falotico, R. y Quatto, P. (2015). Fleiss' kappa statistic without paradoxes. *Quality & Quantity*, 49(2):463–470. DOI: [10.1007/s11135-014-0003-1](#).
- Frías, C., Bello, R., y González, S. (2015). *4 a 7, Mujer trabaja tranquila. Resumen Ejecutivo Evaluación de Programas Gubernamentales*. En línea: [enlace](#).
- Frías, C., Díaz, D., Maripangui, C., y Ramaciotti, L. (2018). *Programa de Apoyo a la Retención Escolar "Enriquecimiento de la experiencia educativa". Resumen Ejecutivo Evaluación de Programas Gubernamentales*. En línea: [enlace](#).
- Haefner, C. y Villegas, X. (2019). Evaluación de programas públicos. Análisis del desempeño global de programas gubernamentales en Chile. *Revista de Estudios Políticos y Estratégicos*, 7(2):14–38. En línea: [enlace](#).
- Irrarázaval, I., Larrañaga, O., Rodríguez, J., Silva, E., y Valdés, R. (2020). *Propuestas para una mejor calidad del gasto y las políticas públicas en Chile*. Centro de Políticas Públicas. Universidad Católica.
- Libertad y Desarrollo (2020). *Una mirada a fondo a la evaluación de programas*. Serie Temas Públicos. En línea: [enlace](#).
- López de Ullibarri, I. y Pita-Fernández, S. (1999). Medidas de concordancia: el índice de Kappa. *Cadernos de Atención Primaria*, 6:169–171. En línea: [enlace](#).

- Marcelo, H., Bello, R., Frenk, P., y Signorini, V. (2017). *Programas Deportivos: Deporte y Participación Social, Escuelas Deportivas, FONDEPORTE, Sistema Nacional de Competencias Deportivas y Liderazgo Deportivo. Resumen Ejecutivo Evaluación de Programas Gubernamentales*. En línea: [enlace](#).
- Moher, D., Liberati, A., Tetzlaff, J., y Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, 6(7):e1000097. DOI: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097).
- Monreal, M., Neira, C., y Olavarría, M. (2012). *Programa Vida Nueva. Resumen Ejecutivo Evaluación de Programas Gubernamentales*. En línea: [enlace](#).
- Neuendorf, K. A. (2011). Content Analysis—A Methodological Primer for Gender Research. *Sex Roles*, 64(3-4):276–289. DOI: [10.1007/s11199-010-9893-0](https://doi.org/10.1007/s11199-010-9893-0).
- Olavarría, C. y Peroni, A. (2012). La evaluación en el Estado chileno, avances y desafíos. *Políticas Públicas*, 4(2). En línea: [enlace](#).
- Olavarría, M. (2012). La evaluación de programas en Chile: análisis de una muestra de programas públicos evaluados. *Revista del CLAD Reforma y Democracia*, 54:143–178. En línea: [enlace](#).
- Peroni, A. (2014). *Diseño Integral de una Intervención Social (programas y proyectos)*. Santiago de Chile: Universidad de Chile. En línea: [enlace](#).
- Peroni, A. y Páez, K. (2019). *La evaluación en medio del presupuesto y la institución. XIII Jornadas de Sociología*. Buenos Aires: Universidad de Buenos Aires. En línea: [enlace](#).
- Ríos, S. (2007). *Fortalecimiento de los Sistemas de Monitoreo y Evaluación en América Latina, Diagnóstico de los Sistemas de monitoreo y Evaluación en Chile*. Santiago de Chile: CLAD-BID.
- Sepúlveda, N. (2019). *¿Quién se preocupa por el Gasto Fiscal?* Fundación Observatorio Fiscal. En línea: [enlace](#).