

CUATRO HIPÓTESIS SOBRE LA INSTITUCIONALIZACIÓN DE LA EVALUACIÓN EDUCATIVA EN AMÉRICA LATINA: EL URUGUAY EN MIRADA COMPARADA

FOUR HYPOTHESES ON THE INSTITUTIONALIZATION OF EDUCATIVE
EVALUATION IN LATIN-AMERICA: URUGUAY IN COMPARATIVE
PERSPECTIVE

Tabaré Fernández

Revista Iberoamericana de Evaluación Educativa 2010 - Volumen 3, Número 3

<http://www.rinace.net/riee/numeros/vol3-num3/art5.pdf>

Fecha de recepción: 31 agosto de 2010
Fecha de dictaminación: 24 octubre de 2010
Fecha de aceptación: 03 de noviembre de 2010

¿ Por qué se observa una heterogeneidad tan importante en la institucionalización de la evaluación de aprendizajes en la educación básica y media de América Latina? Actualmente, sólo existe un reducido grupo de países donde *regularmente* se ha evaluado el aprendizaje *en varios grados*, con metodologías que permiten *comparar* los resultados a lo largo del tiempo y *el público accede* a los informes. La norma general es que alguno o varios de estos atributos han estado ausentes. Pasados casi treinta años y contando con una significativa cantidad de evaluaciones en la región, son varios los países en los que hubo programas ya no tienen, directamente no han podido establecer ninguno.

Uruguay es un caso destacado entre éstos últimos. Inició en 1996 las evaluaciones de aprendizaje periódicas, estableciendo un programa de evaluación en 6tos. Primaria con amplia difusión, contextualización y análisis de los resultados al punto que fue citado como un modelo en la región (Benveniste, 2000 y 2002; Ravela *et al.*, 2000). En el año 2000, se estimaba que el 44.8% de los maestros de Primaria habían leído al menos uno de los informes de evaluación (Fernández *et al.* 2000: cuadro IV.4), guarismo que se mantenía en el 45.8% para el año 2003 (Fernández and Midaglia, 2005: cuadro 4.1), con un menor impacto sobre la planificación de la enseñanza de aula y una valoración positiva de los instrumentos (Aristimuño and Kaztman, 2005: cuadro 7). Una década más tarde, la Administración Nacional de Educación Pública (ANEP)¹ no ha logrado equiparar las últimas pruebas aplicadas en 6tos. Primaria y no ha logrado establecer ningún programa para evaluar la educación media, siendo PISA el único mecanismo para monitorear la calidad en este nivel. A esto cabe añadir episodios calificables de censura de resultados, el primero ocurrido en 2003 y el último relacionado con las evaluaciones que la Secundaria Técnica Pública viene aplicando a sus estudiantes desde 2007. La Ley General de Educación (n°18437) promulgada en diciembre de 2008, procuró abrir una nueva etapa en este ámbito de las políticas educativas creando el Instituto Nacional de Evaluación Educativa (INEE) como entidad para-estatal, independiente de la ANEP, pero hasta el presente (2010), el Poder Ejecutivo aún no ha dado pasos concretos para establecer el INEE.

En otros trabajos, he explorado aspectos históricos y organizacionales *singulares* que refieren a este proceso (Fernández *et al.*, 2000; Fernández and Midaglia, 2003 y 2005; Fernández, 2008 y 2009). Aquí deseo explorar hipótesis explicativas que relacionan los niveles de institucionalización de la evaluación en los países de la región, con las reformas de segunda generación (en particular, la descentralización), la orientaciones ideológicas de los gobiernos (el peso del mercado y del ciudadano), utilidad técnico-pedagógica la información y el diseño institucional del gobierno de las agencias de evaluación. Este trabajo me permitirá discutir para el caso específico, el problema politológico más general de la formación de instituciones en contextos de los conflictos redistributivos experimentados por la educación básica y media desde los años noventa; conflictos de los que Uruguay por cierto no ha escapado.

1. LAS HIPÓTESIS

1.1. Hipótesis sobre la institucionalización

¿Qué factores pueden explicar la débil institucionalización que tendría hoy en día la evaluación educativa

¹ En Uruguay, la máxima autoridad educativa no es el Ministerio sino este cuerpo colegiado de competencia nacional y de organización hipercentralizada, que administra los niveles inicial, primaria, secundaria general, secundaria técnica y formación docente inicial.

en Uruguay, a pesar de sus vigorosos inicios? En la bibliografía sobre la evaluación educativa de América Latina se pueden rastrear hasta cuatro grandes teorías que permiten construir explicaciones rivales, aunque no contrapuestas, sobre la formación de instituciones de evaluación. Si bien no son necesariamente contradictorias, cada una pretende para sí el papel de explicación principal.

La teoría más recurrida, al menos en los años noventa, asociaba la existencia y pervivencia de las evaluaciones a la puesta en marcha de reformas de segunda generación en el sistema educativo (DiGropello, 1999, 2004; Winkler and Gershberg, 2000), y a las condiciones puestas por recibir préstamos de parte de los organismos multilaterales tales como el Banco Mundial y el Banco Interamericano (Narodowski, Nores and Andrada, 2002; Bertolini, 2001). En particular, la descentralización de la prestación de los servicios a unidades subnacionales, conllevó como contrapartida, nuevos roles para “ministerios de educación sin escuelas”, fundamentalmente relacionados al control de desempeño de escuelas a las que se le atribuían crecientes competencias (Fernández, 2004).

Los procedimientos estándares de generación de información estadística y pedagógica dentro de los sistemas, más allá de cuál fuera el grado de desarrollo alcanzado, resultaron insuficientes para que el ministerio evaluara la *calidad* de la educación; fue necesario un *sistema de evaluación* y así quedó consagrado en las leyes de educación de ese período. Las evaluaciones externas, legitimadas ampliamente en el diseño de políticas educativas en Estados Unidos, resultaron ser el mecanismo ideal (Narodowski, Nores and Andrada, 2002). En consecuencia de este argumento *sistémico* o *estructural*, debiera observarse que las evaluaciones están más sólidamente institucionalizadas en aquellos países donde más se avanzó en reformas de segunda generación, sea por la vía de la descentralización o la privatización (hipótesis 1). El indicador más general será si el ministerio es el principal prestador de servicios de educación básica.

La segunda teoría vincula la fortaleza de la institucionalización de la evaluación a la *orientación liberal* de un gobierno. En un mercado signado por una fuerte asimetría entre prestadores y clientes, dichas políticas asignarían un papel crucial a la información en la toma de decisiones de las familias. Claramente, esto modificaría los comportamientos agregados al declinamiento de la calidad de la educación, y demandaría la habilitación de mecanismos de salida o de voz (Hirshman, 1970). Esta visión sobre el funcionamiento de la prestación educativa genera un conflicto dado que propone *empoderar* a los actores externos a la escuela (incluido aquí la jerarquía ministerial) y en general *a los ciudadanos*, frente a los maestros y directivos de las escuelas. Aquellos serían los actores más interesados en conservar las evaluaciones porque la institucionalidad educativa disminuye a su favor las asimetrías de información. Dentro de esta visión *liberal*, la difusión de resultados tanto a nivel *general* como *a nivel de cada escuela*, es una propiedad crítica de cualquier evaluación. De aquí el impulso por establecer evaluaciones con dos características principales: periodicidad (anual por ejemplo) y cobertura censal (para poder entregar información de cada escuela).

Este argumento permitiría sustentar dos hipótesis sutilmente distintas. La primera hipótesis asociaría la estabilidad de partidos liberales con una mayor institucionalización de la evaluación, y la alternancia “derecha /izquierda” con la supresión los programas de evaluación (hipótesis 2b). Aunque así planteada puede parecer muy restringida, tiene una clara conexión con una de las teorías generales que explican la formación de instituciones a través de la *estabilidad* en el tiempo de las reglas, de la inercia, y de los costos implicados en el cambio de reglas que ya han estabilizado expectativas sociales (North, 1995). La *estabilidad* ha sido un valor particularmente importante para el liberalismo político en la región dado que

en su visión, esto permite a todos los actores (públicos y privados) ajustar expectativas recíprocas, costos y beneficios, partiendo de la base de los resultados anteriores y sin emplear tiempo y recursos adicionales para acceder a información adicional. La segunda hipótesis es más sutil y considera exclusivamente la orientación de la política de difusión de resultados que desarrolló la agencia de evaluación frente a cada proyecto. En el límite, si hubo libre acceso o si por el contrario, se introdujo algún tipo de censura a la comunicación de resultados al público (en adelante, hipótesis 2b).

En este trabajo se testearán ambas hipótesis, definiendo cuatro indicadores: i) el número de evaluaciones realizadas en cada país; ii) el total de años cubiertos entre la primera y la última evaluación; iii) y si actualmente hay evaluaciones ("vigencia"); y iv) si hubo evaluaciones en las que no se difundieron resultados al público².

La tercera teoría define la formación de las instituciones de evaluación con base en su *utilidad* (pedagógica) para actores educativos (maximizadores): maestros, inspectores, autoridades. Supondré que toda evaluación bien diseñada es un instrumento óptimo para concretar los intereses pedagógicos que motivan a los actores³. El argumento deposita especial atención a las distintas dimensiones de la hechura *técnica* de la política y al seguimiento de *buenas prácticas* en materia de evaluación: clara definición de los conocimientos a evaluar para cada área y grado ("tabla de especificaciones"); sólida validez de constructo; coherente aplicación de una teoría de la medición (clásica o TRI); diseño muestral o censal tal que les permite realizar inferencias válidas para la población estudiantil delimitada; equiparación de las pruebas para comparar en el tiempo; contextualización de resultados con indicadores sociales; análisis multivariados de factores asociados; comunicación a tiempo de resultados a las escuelas; redacción de informes amistosos y elocuentes, acompañados de recomendaciones didácticas (Ravela, 2001; Ravela *et al.*, 2008). Ahora bien, si todas las evaluaciones tienen altos costos, las buenas lo son aún más (tanto para los maestros como para los ministerios). Pero un buen diseño, que obtenga un consenso sustantivo de los actores en materia pedagógica, produciría beneficios que excederían ampliamente a los costos públicos y privados. Estos extremos son más complejos de evaluar y aquí no puede ser discutido en toda su extensión, aunque hay estudios de caso que profundizaron en estos aspectos (Pinkasz *et al.*, 2005; Fernández and Midaglia, 2005; Aristimuño and Kaztman, 2005; Zorrilla, 2003; Narodowski, Nores and Andrada, 2002). En términos comparados, contrastar esta teoría requeriría de máxima, evaluar directamente las evaluaciones realizadas, un trabajo de extensión que sobrepasa los recursos disponibles. En su lugar, aquí adoptaré una hipótesis mínima que reportará como indicador dicotómico si existen cambios declarados por las propias agencias en la metodología empleada para hacer evaluación. Esto permitirá sondear una de las dimensiones consideradas más atrás: la *comparabilidad en el tiempo de los resultados generada por una metodología consistente* (hipótesis 3).

1.2. Agente principal en la evaluación educativa

La bibliografía neo-institucionalista sobre educación ha resaltado constantemente que la educación es un caso típico de asimetría de información en la que, el cliente principal de la relación educativa (el alumno y sus familias), carece de información válida sobre el resultado final del proceso educativo: ¿qué tipo de

² La duración de este período debiera definirse laxamente, puesto que depende en primer lugar de las capacidades técnicas existentes en las agencias de evaluación, pero también, depende de la presión oficial por conocer u ocultar resultados. En este balance, el período de un año luego de aplicada una prueba es un tiempo más que razonable para definir censura.

³ Refiero tanto a las externas como internas. El supuesto es fuerte pero encuentra total consenso en todas las teorías pedagógicas contemporáneas.

aprendizajes ha hecho el alumno y qué nivel de competencia ha alcanzado? (Chubb and Moe, 1990; Gauri and Wawda, 2003)⁴.

Las calificaciones de las maestras, si bien tienen valor jurídico (implican por ejemplo, la promoción o reprobación de un grado), carecen de validez informativa clara: ¿una calificación de 6 en esta escuela del sur es comparable a la asignada en una escuela al norte del país?. En la Educación Media Superior por lo general esta falta de validez suele afectar las trayectorias académicas en forma significativa cuando los estudiantes, que emigran de una localidad a una capital, encuentran que sus desempeños han sido sobre-valorados y que sus posibilidades de progresar son mucho menores que las imaginadas.

Este argumento clásico también se puede extender y se puede agregar al conjunto de la ciudadanía para hacer referencia a la asimetría de información que existe entre la autoridad representativa (por ejemplo, el Parlamento) y el sistema educativo. En varios países, los legisladores que deciden anualmente sobre montos presupuestales para la educación, tienen información escasa: a lo sumo sobre la tasa de reprobación y la tasa de abandono de cursos. También es razonable suponer que en la misma situación de asimetría se encuentran otras autoridades del Poder Ejecutivo, tales el Presidente y ministerios de otras ramas no educativas, incluso cuando tienen directo interés en los resultados educativos o realizan transferencias importantes para financiar proyectos en la educación.

Los estudios sobre uso de la información para la formulación de políticas educativas hechos en la región (Cueto, 2005), muestran que los ministerios o secretarías de educación también disponen una reducida y endeble información válida sobre el aprendizaje de sus alumnos. Tampoco cuentan en muchos casos, con información sobre variables clave de los procesos: tal como asistencia docente, la implementación de proyectos educativos institucionales o la realización de coordinaciones docentes. Se puede inferir por tanto, que esta relación entre los docentes y las autoridades se desarrolla en una asimetría de información tal, que hace imposible controlar las condiciones de desempeño de las escuelas.

La introducción de una evaluación externa, estandarizada, periódica, con base en un diseño metodológico documentado, con una teoría de la medición y un muestreo rigurosos, acompañada de cuestionarios de contexto, asegura indicadores válidos y confiables. Las escuelas dejan de ser "cajas negras". Es claro entonces, que la evaluación tiene potencialidades para alterar la asimetría de información de la relación educativa (al menos) en favor de los tres principales identificados: familia, parlamento y las propias autoridades educativas⁵. Algunas de las consecuencias de política que acarrea esa nueva posición informada del principal pueden ser perjudiciales para los docentes: atar salarios a desempeños (Carrera Magisterial de México o Chile); asignar partidas presupuestales a metas de aprendizaje (ENIA en Uruguay); focalizar programas socio-educativos en escuelas de bajos desempeños (el P900 de la Concertación en Chile entre 1990 y 1994); etc.

Ahora bien, a esta altura del desarrollo de mi argumento se hace necesario introducir el segundo componente del marco teórico: la teoría de la agencia. Junto con examinar cómo la evaluación modifica

⁴ La bibliografía también destaca otras características de esta relación: i) la dificultad de medir estos resultados educativos; ii) la dificultad de programar las tareas propias del proceso de aprendizaje; iii) la falta de una "tecnología" (blanda) estándar para el desarrollo de las tareas. Estos aspectos generan altas incertidumbres sobre el desarrollo y resultado de la tarea.

⁵ Los neo-institucionalistas no suelen tratar el tema sobre el aislamiento didáctico y pedagógico en que suelen estar los docentes, más fiscalizados que asesorados, y con reducidas oportunidades de comparar *formativamente* sus prácticas y sus resultados con otros colegas. Este aspecto generalmente ignorado, ha sido largamente trabajado en el contexto del Grupo de Trabajo sobre Estándares y Evaluación; véase por ejemplo, Ravela *et al* (2008) y (2009).

la relación educativa, es necesario examinar sobretodo, qué relación se establece entre la agencia de evaluación y la autoridad ministerial; el primer principal de esta relación evaluativa. En consecuencia, propondré que el estudio de las evaluaciones educativas desde el punto de vista institucional, debiera concentrarse en el examen de las modificaciones establecidas en dos aspectos importantes conectados con la modificación de la asimetría principal/agente: i) la producción de información *periódica* sobre los aprendizajes, y ii) el acceso a la información de parte de la pluralidad de principales definidos más arriba⁶.

Además de los supuestos clásicos que la teoría hace sobre el agente (Eisenhardt, 1989), supondré que: i) las agencias de evaluación están interesadas en hacer más y mejores evaluaciones, ya que estos dos son elementos que les aportan prestigio en el campo internacional de los evaluadores; y ii) su aversión al riesgo se reduce a evitar fracasos en el desarrollo del operativo de la evaluación (por ejemplo, por falta de apoyos políticos o por bloqueos sindicales). Sin detallar la importancia diferencial que puedan tener estos dos objetos de comportamiento a regular, sostendré que la institucionalidad de la evaluación educativa es débil cuando el examen de los casos nacionales concluye señalado la *carencia* de alguno de aquellos elementos.

La hipótesis 4 que propongo reconoce la eficacia causal que sobre la formación institucional tienen las reformas de segunda generación (hipótesis 1), el tiempo y de las ideologías, (hipótesis 2a), la importancia difusión pública de resultados (hipótesis 2b) y la consistencia técnico-pedagógica y metodológica (hipótesis 3). Sin embargo, considera que este tipo de análisis desconoce la naturaleza misma de toda institución y de los conflictos de intereses específicos que existen entre evaluadores y ministerios. De aquí que sostengo que *la mayor o menor institucionalización de la evaluación es consecuencia del diseño institucional que regula, prescribe, habilita, las relaciones entre la agencia de evaluación y el primer principal de la relación evaluativa*. Los obstáculos mayores radican en que el primer principal de la relación evaluativa, el ministerio, no siempre comparte la opinión de que es necesario informar o que debe seguirse evaluando. Se trata de un actor que también tiene aversión al riesgo, en este caso, a la exposición pública de las ineficiencias y de la mala calidad⁷. Razón por lo cual, restringe o prohíbe la difusión, modifican la ubicación de las agencias en la estructura ministerial, reducen los salarios, desjerarquizan presupuestalmente la evaluación o restringen la contratación de personal calificado.

Frente a estas situaciones, las agencias encuentran reducidos medios para proseguir sus intereses. Esto sería típico en dos casos: i) si no existen mecanismos de intervención de los otros actores (por ausencia de otros principales, como por ejemplo, otros órganos del Poder Ejecutivo, del Parlamento o de las familias en el gobierno de la agencia evaluadora); y ii) cuando el ministerio es *también* el primer principal de la relación educativa, o sea, el prestador de los servicios educativos⁸.

Si la hipótesis del origen del déficit institucional es correcta, debiera ser posible observar que los países donde *no hay evaluación periódica ni difusión pública de resultados*, se caracterizan también por tener

⁶ En nuestro estudio comparado sobre México y Uruguay (Fernández and Midaglia, 2003 y 2005), el objeto fue la modificación de la relación educativa. Sin embargo, a poco de comenzar emergió con gran preeminencia en ambos países, el problema de la relación evaluativa entre las agencias y el principal. Basta recordar que en ambos países, la autoridad prohibió informar sobre al menos un proyecto de evaluación.

⁷ Las mediciones periódicas pueden reportar que los avances educativos durante una gestión no han sido espectaculares y esto conllevar a una pérdida de imagen del Ministro o de todo el Gabinete frente al electorado.

⁸ V.g. la prestación de los servicios no está descentralizada a las entidades, como en México o Argentina, o a los municipios con en el caso de Chile o Brasil.

una agencia cuyo principal es a la vez prestador del servicio (hipótesis 1), y sólo cuentan con *un único* principal en la relación evaluativa (el ministerio).

Para concluir este apartado, quiero resaltar que este enfoque no es novedoso. Ferrer (2006: 14) sostuvo en su balance de mediados de la década que: *"En cualquier caso, lo que resulta imprescindible es definir con claridad el grado de autonomía funcional que tendría la unidad evaluadora. El marco institucional puede tener fuertes implicancias para la credibilidad y legitimidad de sus acciones, para la sostenibilidad técnica-operativa y financiera del sistema, o para las posibilidades de difusión, uso e impacto de los resultados de las evaluaciones."* Precisamente, este componente de la institucionalidad de la evaluación, su gobierno, es a mi juicio, la clave que termina por calificar, mejor, explicar su fortaleza o debilidad; este artículo busca aportar evidencia empírica en esta dirección, la cual hasta el presente es inexistente.

2. EL ESTUDIO

El análisis empírico se realiza sobre la base de un relevamiento de información realizado originalmente en 2006, y actualizado al presente año (2010)⁹. El período histórico de observación estuvo comprendido entre 1980 y el año 2010; un total de treinta años. Tal ventana de tiempo tiene justificación casi exclusivamente para el registro del comienzo de las evaluaciones externas y periódicas en un país: Chile. En varios documentos, también se ha dejado constancia que México comenzó las evaluaciones en la década de 1970; sin embargo, no se ha podido registrar ninguna evaluación entre ese año y 1991. Para el resto de los países, la historia de la evaluación de aprendizajes comienza en los años noventa.

La unidad de registro de la información fue definida en la "evaluación". Se considera que un proyecto es una única evaluación cuando se realiza en un mismo ciclo escolar, durante un período determinado de meses en el año, con un mismo propósito general (débil o fuerte, en el sentido de Ravela, 2001), bajo un mismo diseño muestral, con un mismo marco conceptual, una misma teoría de la medición (clásica o TRI), aunque estuvieran dirigidos a varios grados, ciclos o niveles del sistema educativo. Bajo esta definición, un proyecto generalmente implica un operativo en el terreno, aunque puede darse el caso de que a través de una misma logística se apliquen dos proyectos, como ha sido el caso de Prova Brasil y SAEB.

Es clásica ya la distinción entre evaluaciones con alto impacto y con bajo impacto, formulada por Ravela (2001). Visto el marco teórico y los objetivos del análisis, tomé la decisión de sólo considerar evaluaciones externas que no tuvieran consecuencias directas sobre la trayectoria académicas formales de los estudiantes. En consecuencia, excluí aquellas cuya finalidad principal *fuera* la acreditación de un grado o de un nivel, tal como es el caso de la Prueba de Suficiencia Universitaria en Chile, los Exámenes de Estado en Colombia, las Pruebas de Conocimiento de Costa Rica, las pruebas PAES de El Salvador y las Pruebas Nacionales de la República Dominicana.

Con base en el marco analítico definido por dimensiones extraídas de un enfoque institucionalista combinado con la teoría agente/principal, se establecieron las siguientes dimensiones de análisis: i) definición institucional (nombre, agencia, dependencia, normativa reguladora); ii) levantamientos (años, grados, asignaturas); iii) diseño metodológico (enfoque de la prueba, tipo de reactivos, teoría de la

⁹ Agradezco la colaboración de la Lic. Inés Malán en la confección de la base de datos de 2006.

medición, equiparación, muestra); y iv) uso y difusión de resultados (destinatarios, tipos de informe, acceso a microdatos). Cada evaluación fue catalogada en estos aspectos, al menos, hasta donde fuera posible hallar información.

Cuatro fueron las fuentes principales de información, todas ellas secundarias, que fueron empleadas: i) los reportes publicados (cuando los hubo); ii) los estudios sobre los sistemas de evaluación; iii) las páginas web de PREAL, la OEI, los ministerios y las agencias de evaluación¹⁰; y iv) el Foro de Evaluación que abriera Ferrer en la página de PREAL con base en su libro *“Los sistemas de evaluación educativa en América Latina”*. La última actualización de sus fichas es para el mes de abril de 2009.

2.1. Casos seleccionados

Se construyó una base de datos para los países de la región latinoamericana¹¹, conteniendo información descriptiva sobre cada una de las evaluaciones nacionales estandarizadas realizadas. De los 19 países de la región, sobre dos no fue posible acceder a información secundaria precisa sobre las evaluaciones (Venezuela y Cuba). Otros dos, Costa Rica y República Dominicana, sólo realizaron evaluaciones con la finalidad de acreditación, hasta donde se tiene conocimiento, por lo que también fueron excluidos de los análisis. Un punto importante para el relevamiento fue establecer la vigencia de las evaluaciones. Esto se determinó genéricamente considerando si se habían realizado evaluaciones los dos últimos años, o si al menos, se habían producido informes de evaluaciones realizadas antes de 2009. Aquí resaltan tres casos: Honduras, Bolivia y Paraguay. En el primero, se identificó que la última evaluación fue en 2008 pero no hay información sobre su procesamiento. Es de recordar que el país vivió un golpe de estado en junio de 2009 y pervive aún el enfrentamiento civil desatado. Bolivia por su parte, aplicó por última vez en 2000, y Ferrer (2006) da cuenta que el sistema de evaluación conocido como SIMECAL fue cerrado en 2004, sin que haya sido re-emplazado por otro. Paraguay por su parte, realizó en 2006 la última aplicación conocida de su programa denominado SNEPE, sobre el cual se han publicado informes hasta diciembre de 2007. No he registrado actividad del SNEPE desde ese entonces.

Nicaragua fue incluida aunque presentó la siguiente ambigüedad: por un lado, no aplica evaluaciones basadas en el denominado SNE desde el año 2002; por otro, en 2008 se registró una evaluación diagnóstica hecha por un consorcio entre el Ministerio y la Universidad Nacional Autónoma aplicado al final de la Educación Media sobre el cual hay información parcial de resultados difundidos en 2009.

En síntesis, el análisis cubre el período de 30 años para Argentina, Brasil, Colombia, Chile, Ecuador, El Salvador, Guatemala, México, Nicaragua, Panamá, Perú y Uruguay.

2.2. Niveles de análisis e indicadores

El objeto mismo de la investigación requiere estudiar la evaluación educativa en términos institucionales para lo cual se requiere *inferir* si más allá de la realización de operativos o proyectos de evaluación, se han establecido haces de normas capaces de regularizar el comportamiento de los actores y redistribuir la información. En consecuencia, fue necesario establecer criterios para distinguir qué evaluación fue un emprendimiento excepcional, puntual, sin continuidad, de aquellas otras evaluaciones que formaban

¹⁰ http://www.preal.org/Grupo.asp?Id_Grupo=3&Id_Seccion=29&Id_Seccion2=139

<http://www.oei.es/quipu/evaluacion.htm>

¹¹ La región “América Latina” aquí refiere los siguientes países: Argentina, Bolivia, Brasil, Colombia, Costa Rica, Cuba, Chile, República Dominicana, Ecuador, El Salvador, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Perú, Uruguay y Venezuela.

parte de una política de largo plazo. El concepto de "programa" fue utilizado para marcar dos grados de institucionalización.

El "nombre" oficial de la evaluación fue el primer criterio. La pretensión de continuidad está generalmente asociada al uso recurrente de los términos "sistema", "programa" u "operativo". Es el caso, por ejemplo, del Sistema de Medición de la Calidad (SIMCE) de Chile y el Programa de Estándares Nacionales (EN) de México. En cambio, el término está ausente en los proyectos hechos sin pretensión alguna de continuidad, como fue el caso de la Evaluación en el Primer Nivel de Escolarización hecha por Uruguay en 2001 (ANEP, 2002). Sin embargo, el nombre como indicador tiene limitaciones. Un cambio de nombre en el tiempo no implica necesariamente un cambio de programa, sino más bien un cambio en la agencia de evaluación. Es el caso del término "Pruebas Nacionales" para los ciclos 2002-2003 y 2003-2004, que adoptara el Instituto Nacional de Evaluación de la Educación de México (INEE) para designar los levantamientos del Programa Estándares Nacionales antes realizados por la Dirección General de Evaluación (DGE) de la Secretaría de Educación Pública.

A su vez, la falta de nombre no indica necesariamente ausencia de programa. Algunas evaluaciones fueron tituladas simplemente "pruebas" o "evaluaciones", aunque luego generaron dos, tres y hasta más proyectos. El segundo criterio que apliqué introduce la idea *metodológica* de la replicación: sin perjuicio del nombre, habrá programa toda vez que dos o más evaluaciones sucesivas se hayan dirigido a los mismos grados y hayan utilizado una metodología similar ("tabla de especificaciones" y teorías de la medición). Uruguay cae dentro de esta categoría ya que sus evaluaciones en 6tos. Primaria de 1996, 1999 y 2002 conformaron un programa en este sentido. En Perú, las evaluaciones de 2001, 2004 y 2008 informan de una continuidad clara, auto referida en las publicaciones, pero sin apelar al término Programa. La noción utilizada fue bastante laxa: alcanzó con que dos evaluaciones tuvieran *explícitamente* esta lógica para *inferir que hubo(hay) programa* y por tanto, indicar un grado de institucionalización *incipiente* de la evaluación educativa.

Tal como lo establece la hipótesis 2A, es razonable sostener que un período más prolongado de existencia de un mismo programa indicaría mayor aceptación, conformidad, consentimiento y utilización de las reglas de juego. En consecuencia, adopté el criterio de que si hubo 3 o más operativos que siguen en la lógica de la replicación, sería un indicador de clara institucionalización de la evaluación. El criterio obedece a que en la región los mandatos presidenciales duran 4 años (Chile desde 2006, Argentina, Guatemala, Paraguay, Colombia, Brasil, Honduras, Perú), 5 años (Uruguay) o 6 años (México). El diseño de una evaluación, una vez decidida, demanda un año adicional previo al ciclo escolar de su aplicación. Tres años de levantamientos (si fueran consecutivos) y uno primero de preparación totalizan cuatro años calendario, que lindan con esos mandatos.

Ahora bien, el uso del mismo nombre puede ocultar discontinuidades importantes en uno o varios aspectos *metodológicos en la elaboración del diseño de la política* y por tanto afectar también la formación institucional (hipótesis 3). En los hechos, tanto el SIMCE en Chile como el ONE en Argentina, la PAES de El Salvador, el SAEB de Brasil, conservan programas de largo aliento pero en ambos casos han realizado al menos un cambio en la metodología que generaron cambios importantes en las reglas de juego para los actores. Estos antecedentes demandaron un cuarto criterio, a saber, examinar la existencia de una metodología *consistente* de equiparación de pruebas a lo largo del tiempo; un tema teóricamente complejo que demanda un elevado nivel de formación pedagógico-metodológico-estadístico de los técnicos que trabajan en las agencias de evaluación. Sin embargo, este registro requeriría realizar evaluaciones de las evaluaciones, una cuestión que por el momento está fuera de las posibilidades de

este trabajo. En consecuencia, el indicador utilizado es muy rudimentario e informa que un programa es metodológicamente consistente si expresamente se declara que se han conservado, de un proyecto a otro, las mismas operacionalizaciones de los objetivos o estándares educativos ("tabla de especificaciones") y la misma teoría de la medición¹². Ambos son elementos imprescindibles para poder realizar en forma válida y confiable una equiparación de las pruebas y asegurar la comparabilidad de los resultados en el tiempo. Es de destacar que el relevamiento de este indicador ha sido difícil, debido a la falta de información metodológica pública precisa que existe en muchos de los proyectos de evaluación; sólo excepcionalmente se indica que los resultados *no* son comparables en el tiempo por fallas en la equiparación de pruebas. Tal vez, esto sea, probablemente, el resultado de un comportamiento de aversión al riesgo por parte de evaluadores que podrían perder prestigio si no logran resolver en este problema metodológico tan delicado. En consecuencia, es posible que donde se identifique un programa institucionalizado en realidad haya dos, considerando la metodología. Así, la base de datos permite distinguir cuatro categorías de evaluaciones, ordenadas de acuerdo a los objetivos del trabajo, donde la consistencia metodológica en el tiempo es indicador de su robustez científica, en tanto que la pervivencia por tres o más operativos será indicador de institucionalización.

- proyectos realizados por única vez
- programas de evaluación (2 aplicaciones con los mismos objetivos)
- programas institucionalizados (3 y más evaluaciones)
- programas institucionalizados y consistentes metodológicamente

Se comprende inmediatamente que la noción de programa que aquí se utiliza es bastante amplia. En términos ideales, buena parte de la bibliografía regional ha expresado la recomendación de que los países adopten un programa de evaluación, entendido como una institución educativa, científicamente rigurosa y burocrático establecida, cuyo objetivo es la producción periódica de información sobre los niveles y contextos de aprendizajes de los alumnos de la educación básica. En consecuencia, es factible que el número de programas históricamente desarrollados en la región haya sido sobreestimado por la metodología aquí empleada. El lector deberá tener presente esta limitación a la hora de valorar las inferencias y conclusiones plateadas.

3. PRINCIPALES HALLAZGOS

Este capítulo expone los hallazgos del relevamiento realizado en 15 países de la región. Ha sido ordenado en cuatro grandes apartados: evaluaciones nacionales, evaluaciones internacionales, distribución de la información y agencias. La motivación que guía la exposición es evaluar la razonabilidad de las tres hipótesis que se introdujeron más atrás para explicar el estado actual de la evaluación en la región; brevemente: la metodológica, la política y la institucional.

3.1. Evaluaciones y programas nacionales

El cuadro 1 presenta un primer conjunto de indicadores que caracterizan la evaluación en América Latina según el número de operativos realizados y programas erigidos. En la última fila del cuadro que contiene

¹² No es necesario que el diseño muestral se mantenga de un proyecto a otro.

resúmenes, se puede apreciar que en estos 17 países y en los últimos 30 años, se estima¹³ que se realizaron 186 evaluaciones de aprendizaje; un promedio de 12 evaluaciones por país.

Habría tres grupos de países con diferente nivel de desarrollo en este campo. México es el país más evaluaciones ha realizado (43). Junto con otros cuatro países, Chile, Brasil, El Salvador y Argentina, concentran 110 evaluaciones que representa el 60%. Si se recorre la columna [1] del cuadro, se aprecia que si bien todos los países de la región han realizado evaluaciones, Nicaragua tiene la menor experiencia acumulada (3 evaluaciones), Honduras (5), Panamá (5) y Ecuador (7). Estos países a su vez, implementaron sólo en un sentido limitado las reformas de segunda generación.

Tal como se aprecia, la distribución por país parece acompañar la distribución esperada según la hipótesis 1. Chile municipalizó la educación pública y privatizó un tercio de sus escuelas, generando un sistema de vouchers en 1982. México y Argentina descentralizaron a las entidades federativas la educación básica entre 1992 y 1993. Brasil municipalizó la educación en 1988. Tres países no concuerdan con la teoría. El Salvador, Perú y Uruguay tienen una importante cantidad de evaluaciones pero que durante los años 90, no generaron una reforma estructural de tipo descentralizadora¹⁴. A su vez, Bolivia (4) ensayó programas de segunda generación importantes, sobre todo durante los gobiernos de Sánchez de Lozada. Sin embargo, la evaluación educativa no tuvo gran desarrollo en ese país en el que prosperó el gobierno unitario nacional.

3.1.1. Institucionalización incipiente y programas institucionalizados

Se aprecia que aquellas 186 evaluaciones se han realizado en su gran mayoría (179) conforman proyectos replicados al menos en una ocasión, esto es como programas, configurándose un grado de institucionalización incipiente de interés. En el otro extremo, la columna [2] del cuadro informa el número de evaluaciones *no replicadas* que hubo en la región y destacan dos países, Nicaragua y Guatemala. Este dato correlaciona bien con la mera contabilidad. Pocas evaluaciones abonan una precaria institucionalización. Sin embargo, Uruguay constituye un caso desviado, ya que de sus 17 evaluaciones, 5 fueron no replicadas: (i) la evaluación en 3ero. Primaria de 1998; (ii) la evaluación en final del Ciclo Básico de Educación Media¹⁵; (iii) la evaluación en el primer ciclo de escolarización que comprendió nivel 5 años de Inicial y 1ero y 2do Primaria en 2001; (iv) la evaluación en el último año de la Educación Media Superior en 2003; y (v) la evaluación on-line en Primaria en 2009.

Habrían existido o existen actualmente, 39 programas nacionales de evaluación externa de aprendizajes en la educación básica. Esto parecería confirmar la hipótesis general de formación de las instituciones por la estabilidad (pervivencia) de las reglas de juego. Sin embargo, una más refinada discusión de la hipótesis 2a, requiere distinguir entre programas incipientes e institucionalizados. De los 39, 8 programas sólo se aplicaron 2 veces y pertenecen a Ecuador, El Salvador, Nicaragua, Panamá, Perú y Uruguay. Estos programas registrados responden a tres períodos y situaciones distintas. Por un lado, están APRENDE de Nicaragua y la Evaluación Nacional en 6tos Primaria de Uruguay: se he hicieron en los noventa y se descontinuaron luego. Por otro lado, están el SINECA de Panamá, CRECER de Perú y SINEA (2001 y 2003) de

¹³ No puede eliminarse la probabilidad de error en las medidas hechas en cada país. Debe recordarse que este estudio realizó su relevamiento sobre la base de fuentes secundarias públicas.

¹⁴ Es discutible la ubicación de El Salvador en este esquema. El sistema, en su concepción normativa, deja en manos del Ministerio de Educación el papel de administrador, supervisor y financiador del sistema educativo formal público. Sin embargo, el sistema EDUCO generó espacios amplios de cooperación público-privado; otros instrumentos también estuvieron en el camino de la flexibilización del control centralista del Ministerio.

¹⁵ En términos internacionales, noveno grado.

El Salvador. Son casos de programas realizados en los noventa y *sustituídos* por otros en los dos mil.

CUADRO 1. CARACTERIZACIÓN DE LOS PAÍSES SEGÚN LAS EVALUACIONES DE APRENDIZAJE EN EDUCACIÓN PRIMARIA Y MEDIA REALIZADAS ENTRE 1980 Y 2010

Países	[1] Nº evaluaciones nacionales	[2] Nº proyectos con única aplicación	[3] = [4] + [5] Nº total de programas	[4] Nº Programas con sólo 2 levantamientos	[5] Nº programas institucionalizados	[6] Nº programas institucionalizados y consistentes	[7] Programa(s) vigente(s) (2009-2010)	[8] Grados Evaluados por Programa(s) vigente(s)
Argentina	12	0	2	0	2	2	One	3º, 6º, 9º
Bolivia	4	0	1	0	1	0		
Brasil	14	0	4	0	4	3	Saeb, Prova, Enem	4º y 8º, 3ºM
Chile	26	0	3	0	3	3	Simce	4º, 8º y 2ºMedio
Colombia	8	0	1	0	1	1	Saber	5º, 9º y 11º
Ecuador	7	0	2	1	1	1	Ser (**)	4º, 7º, 10º y 3ºM
El Salvador	15	0	4	1	3	3	"paesita"	1º, 3º, 6º y 9º
Guatemala	10	1	2	0	2	0	Graduandos	1º, 3º, 6º y 3ºM
Honduras	5	0	1	0	1	1		
México	43	0	7	0	7	6	Excale Enlace	3º, 6º y 9º 4º, 5º, 6º, 7º, 8º y 9º
Nicaragua	3	1	1	1	0	0		
Panamá	5	0	2	1	1	sd	Sineca (**)	3º, 6º, 9º y 12º
Paraguay	8	0	1	0	1	sd		
Perú	9	0	3	1	2	1	Evaluación Nacional	Pre-escolar, 2º, 4º, 6º y 2º Medio
Uruguay	17	6	5	3	2	1	Evaluación Nacional (**)	6to Primaria (**)
Totales	186	8	39	8	31	22	14	

Fuente: elaboración propia

En Ecuador se observa la puesta en marcha de una nueva política de estado. El Sistema Nacional de Evaluación y Rendición Social de Cuentas (SER) emerge de la Consulta Popular sobre el Plan Decenal de Desarrollo Educativo realizada en 2006. En su sexta línea establecía el "mejoramiento de la calidad y equidad e implementación de un sistema nacional de evaluación y rendición social de cuentas del sistema educativo". El SER incluirá cuatro componentes: la evaluación del desempeño de los estudiantes, del desempeño de los docentes, de la gestión institucional y la evaluación de la aplicación del currículo (Ministerio de Educación, 2010). La primera evaluación se concretó en 2008 y la segunda en 2009, aunque al momento sólo hay publicados resultados para 2008.

Panamá implementó durante el gobierno de Torrijos el programa de Pruebas de Logro SINECA con dos aplicaciones, 2005 y 2008, produciendo informes generales y estudio muy preliminar de factores asociados. El nuevo gobierno que asumió el año pasado no ha informado si la próxima aplicación se realizará en 2011.

La columna [5] del cuadro 1 presenta datos para discutir la hipótesis de institucionalización por estabilidad del programa (3 y más aplicaciones). Este indicador muestra ya un primer grupo de países donde la evaluación no se ha institucionalizado o lo estuvo pero en los años noventa: Nicaragua y Bolivia respectivamente.

Dos comentarios específicos merece aquí Uruguay que muestra una débil institucionalización en estos indicadores. Ha tenido dos programas en 6tos años de Primaria debido un cambio sustantivo en el diseño del programa ocasionado principal aunque no exclusivamente por la adopción de la TRI en 2004. Las dos aplicaciones más recientes, 2005y 2009, reiniciaron la serie. También es de notar que entre los programas *institucionalizados*, dos tuvieron características limitadas: uno ya referido (6tos. Primaria 1996, 1999 y 2002); y el otro asociado a la implementación de la Reforma del Ciclo Básico (1996, 1997 y 1998). Sólo hay un programa institucionalizado *multigrado*: el que aplica la Educación Media Técnica a sus estudiantes.

3.1.2. Utilidad, actualidad y consistencia

En la columna [6] del cuadro se incluye el número de programas nacionales que a la vez califican como institucionalizados y consistentes, que serían los más robustos de todos. El número total es de 24. Bolivia, Guatemala, Panamá y Paraguay no habrían tenido este tipo de programas (al menos con la información parcial disponible); a esto se le añade Panamá, dado que no se encontró información sobre este aspecto. Es notable observar aquí que Uruguay si bien tuvo un programa con estas características, ya no lo tiene.

La vigencia de los programas aporta otro matiz al cuadro. Se aprecia en la columna [7] del cuadro que sólo 14 programas (de los 39) se encuentran vigentes en el período 2009-2010, que a su vez corresponden a 11 países. Uruguay, junto con Bolivia, Ecuador, Honduras, Nicaragua, Panamá y Paraguay, carecen *al día de hoy* de un programa de evaluación externa de aprendizajes que haya alcanzado el primer grado de institucionalización (3 y más aplicaciones).

Finalmente, la columna [8] de este cuadro 1 informa la extensión de estos programas vigentes en términos de grados y niveles evaluados. Rápidamente se aprecia otro rasgo distintivo de Uruguay: su programa (no institucionalizado) *sólo evalúa en 6tos. Primaria*. En todos los demás casos, el programa es multigrado y multinivel (Primaria y Media).

3.2. Evaluaciones internacionales

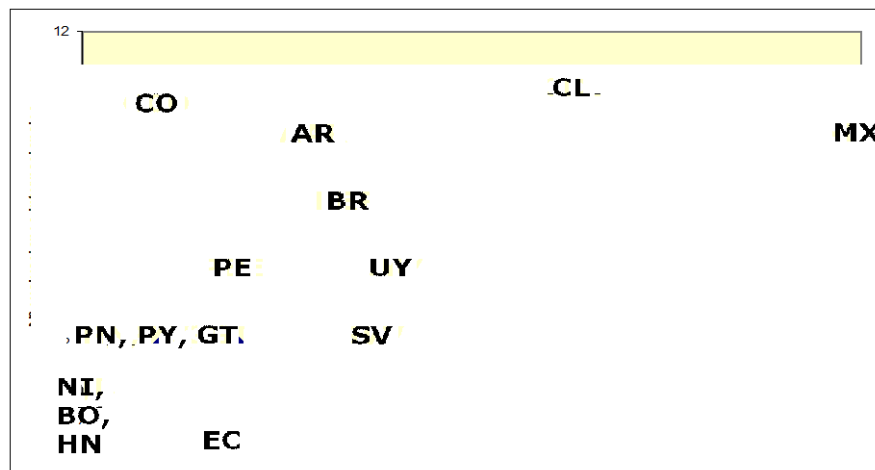
En el período observado se han realizado 13 proyectos de evaluación internacionales en los que participaron países de la región con un total de 65 operativos locales. Estos han sido llevados adelante por tres agencias: la IEA, la OECD y UNESCO. La más alta participación regional ha sido lograda por la UNESCO tanto en 1997 como en 2006, seguida por PISA ciclo 2009. (Véase en los anexos cuadro 1).

En términos de países, cinco destacan por su extensa participación: Argentina (5), Brasil (6), Colombia (9), Chile (10) y México (8). La lista prácticamente coincide con los países que concentran el mayor número de evaluaciones nacionales indicada más arriba, que tienen programas institucionalizados. La excepción aquí es El Salvador: ha participado en sólo en tres operativos (PERCE 1997, TIMSS 2007, SERCE 2006).

Esto se visualiza en el gráfico 1. Es interesante notar que, al correlacionar este indicador con el número de evaluaciones nacionales, también se descubre que también estos países presentan debilidades históricas en el plano interno (con la excepción de Ecuador), sea porque nunca tuvieron programas (Nicaragua), no los tienen vigentes (Bolivia, Honduras, Paraguay) o son de muy reciente surgimiento (Ecuador). Aquí claramente hay un cluster con cuatro países sin *programas institucionalizados vigentes* de evaluación (Bolivia, Ecuador, Honduras, Nicaragua y Paraguay), y dos países con programas institucionalizados pero con débil exposición al exterior (Panamá y Guatemala).

En el otro extremo, el gráfico 1 muestra un clúster de países larga y ampliamente evaluados, tanto en el plano nacional como en el internacional, con programas institucionalizados, consistentes y vigentes, que son Argentina, Brasil, Chile y México. Habría un cuarto grupo de países que parecerían tener un desarrollo desigual en alguna de las dos dimensiones: importante exposición internacional (Colombia) o un alto número de evaluaciones nacionales (El Salvador, Perú y Uruguay) con una participación internacional moderada y reciente (realizada en los años dos mil).

GRÁFICO1.



3.3. Población y muestra

Las constataciones anteriores permiten realizar algunas inferencias preliminares. La evaluación educativa se habría extendido en todos los países de América Latina, y habría alcanzado un primer grado de institucionalización excepto en dos países (Bolivia y Nicaragua). Esto no necesariamente conlleva tener programas robusto y menos aún, vigentes. Esta asociación parecería amparar sólo parcialmente la teoría de que la pervivencia en el tiempo es una condición muy importante para la institucionalización de los programas de la evaluación. Este apartado ensayará analizar la hipótesis 2 A, cuestión que permitirá una primera consideración de la teoría de los efectos redistributivos.

A continuación presentaré un primer aunque rudimentario indicador sobre si los programas modificaron la distribución de la información que disponen los actores externos al sistema educativo no calificados más allá de su ciudadanía. La pregunta es muy sencilla ¿faltó en alguna oportunidad un reporte público de resultados? Con base en las fuentes consultadas, la estimación preliminar se incluye en el cuadro 2.

Al menos en 6 países de la región se han registrado ausencia de información pública sobre los resultados

de las evaluaciones. Hay dos tipos de situaciones que son pertinentes destacar. Por un lado, los casos de El Salvador, México, Panamá y Paraguay, donde la ausencia de difusión alcanza a programas completos o a una parte importante de los levantamientos realizados. Las pruebas SABE de El Salvador no tuvieron reportes públicos; sólo a usuarios internos del sistema (escuelas y autoridades). El programa de evaluación aplicado en Panamá durante fines de los noventa Paraguay habría realizado evaluaciones entre 1996 y 2006, pero sólo se conocen públicamente resultados para el primer y el último levantamiento. México constituye el caso más extremo de todos, donde la política de la Secretaría de Educación, durante 6 años, fue no difundir evaluaciones de ninguno de los 4 programas nacionales que tenía implementados ni tampoco de la evaluación internacional de TIMSS 1994-1995.

CUADRO 2. CLASIFICACIÓN DE LOS PAÍSES SEGÚN SE REGISTREN CENSURAS A LA PUBLICIDAD DE RESULTADOS DE SUS EVALUACIONES

Países que restringieron más de una evaluación:	Países que no registran censuras:
<ul style="list-style-type: none"> ▪ El Salvador (Hasta 2001) ▪ México (hasta 2000) ▪ Panamá (hasta 2001) ▪ Paraguay ▪ Uruguay (2003 y 2007-2009) 	<ul style="list-style-type: none"> ▪ Bolivia (período SIMECAL, hasta 2000) ▪ Brasil ▪ Colombia ▪ Chile ▪ Perú ▪ Ecuador
Países que restringieron al menos una publicación o la demoraron:	Países sin información concluyente:
<ul style="list-style-type: none"> ▪ Argentina (ONE2003) 	<ul style="list-style-type: none"> ▪ Guatemala

En el caso de Argentina y de Uruguay, la falta de información se refiere a coyunturas muy específicas. La crisis política de los años 2001 a 2003 (renuncia del Presidente de la Rúa), llevó al cierre del Instituto de Desarrollo de la Calidad de la Educación (IDECE), generó una gran inestabilidad en la política educativa que llegó incluso a suprimir el Ministerio de Educación de la Nación durante la breve presidencia de Rodríguez Saa. En ese período, Argentina no publicó informe nacional para PISA 2000 ni para PIRLS 2001. En mayo de 2003 asume como Presidente Kirchner y como Ministro de Educación, Filmus, quienes continuaron la implementación del ONE. El informe de resultados vio demorada su publicación y tuvo un carácter muy general, no reportó por provincias ni hizo análisis más sofisticados, tal como se había venido haciendo con anterioridad. En Uruguay, la evaluación realizada en el último grado de Bachilleratos en octubre de 2003 tenía listo el informe para 2004, pero las autoridades de ese momento consideraron inoportuno su publicación¹⁶. El informe fue publicado una vez que las nuevas autoridades asumieron en 2005, aunque tuvo reducida difusión en la opinión pública y sus resultados no fueron utilizados en ningún marco (académico o político).

Cabe anotar aquí también que algunos programas pusieron a disposición de usuarios externos, las bases de datos de las evaluaciones. Es el caso del ONE, que en 2000 "colgó" las bases de 1997 a 1999, y posteriormente también la del año 2000. De hecho, Argentina fue el primer país, junto con Perú, en darle publicidad on-line a los microdatos, política que ha sido seguida por México y Brasil. Tal nivel de publicidad es más bien excepcional. Chile y Uruguay tienen políticas reguladas de acceso a los microdatos

¹⁶ Extrañamente, también censuraron otra publicación: una síntesis detallada de los logros alcanzados luego de 12 años de reformas en la educación entre 1992 y 2004.

para académicos¹⁷. Los restantes países, hasta donde se pudo llegar con este relevamiento, no existen protocolos públicos para el acceso a los microdatos por parte de académicos.

Entre los países centroamericanos se encuentra al menos un informe de difusión en El Salvador, Honduras y Guatemala, sea en formato papel o en medios electrónicos. En el primer caso, la publicidad sobrevino con las pruebas que el SINEA comenzara a aplicar en 2001 y está innegablemente asociada al uso de acreditación que tiene el PAES. Honduras ha realizado informes descriptivos, estudios de factores asociados y de tendencias en el tiempo para sus sucesivos levantamientos, y resultó hasta 2008 el país con mayor nivel de estudios de la sub-región. La prueba Graduandos de Guatemala tiene difusión en web desde 2007 y la evaluación en Primaria realizada en 2004 alcanzó amplia difusión, principalmente “empujada” por la USAID.

3.4. Agencias

La realización de la primera evaluación de aprendizajes requiere el reclutamiento de dos equipos: uno de especialistas en curriculum que definan y validen los contenidos incluidos en cada ítem de las pruebas, y otro equipo con conocimientos sólidos en muestreo, medición y análisis estadístico. También es necesario conformar una unidad de logística que se ocupe de los traslados y de la aplicación externa en cada escuela del país, siguiendo los estrictos protocolos que aseguran la validez y confiabilidad científica de las mediciones. Los estudios de caso hechos sobre este tema coinciden en señalar que son competencias profesionales escasas en casi todos los países de la región, con lo que se puede afirmar razonablemente que son muy altos los “costos hundidos” que conlleva este primer operativo. Más aún, el diseño organizacional de una agencia como unidad especializada (dentro o fuera de la administración pública), que realice sucesivas evaluaciones.

Por consiguiente, era de esperar que el relevamiento sobre el número de agencias de evaluación que han existido en gruesamente desde inicios de los noventa fuera muy bajo, próximo a 1 y que la región no hubiera constituido más de 15 o 16 agencias. Sin embargo, el panorama resulta bastante distinto: en las dos décadas pasadas se erigieron 30 agencias con un promedio de 2 por país. Esto es, una agencia por década. La moda de la distribución también es 2, habiendo tres países (Argentina, Ecuador y Guatemala) con tres agencias y 5 países que tuvieron sólo una hasta donde se ha podido relevar (Bolivia, Brasil, El Salvador y Paraguay). Uruguay se destaca en la región por ser el país que ha instituido en estos años mayor cantidad de agencias de evaluación: 5 en total.

La agencia más antigua de la región es el Instituto Nacional de Estudios y Pesquisas Anísio Teixeira (INEP) de Brasil fundado en 1937 aunque su diseño institucional actual data de 1995. Le sigue la Dirección General de Evaluación (DGE) de México creada en 1970, y el Instituto Colombiano de Fomento a la Educación Superior (ICFES) creado en los años ochenta. Las agencias más nuevas que se han creado están en Perú (Ley General de Educación N° 28044), el Instituto Peruano de Evaluación, Acreditación y Certificación de la Calidad de la Educación Básica (IPEBA) y el Instituto Nacional de Evaluación Educativa (INEE) de Uruguay (Ley de Educación n°18437), aunque este último aún está fuera de funcionamiento por razones políticas y no ha realizado ninguna evaluación (Fernández, 2009).

Por lo general, en su origen las agencias fueron gestadas en dos grandes modalidades: i) como componentes de evaluación en proyectos de mejoramiento de la educación básica financiados por

¹⁷ La excepción ha sido con los microdatos de PISA, ciclos 2003 y 2006, donde la DIEE liberó on-line la base incluyendo las variables “nacionales”.

organismos internacionales (el Banco Mundial, el Banco Interamericano, la Agencia para el Desarrollo de Estados Unidos), o ii) como consorcios inter-institucionales nacionales en los que participa al menos una universidad local. Dependiendo de los países, de esta primera agencia se pasa a una agencia incorporada a los ministerios de educación como unidades especializadas, en general, ubicadas dentro de las subsecretarías o direcciones de planificación. Este es el formato con mayor predominancia actualmente. La excepción a este formato se encuentra en Colombia con el ICES, en Brasil con el INEP y en México con el INEE, todos organismos con personería jurídica propia bajo la forma de entidades paraestatales o empresas públicas¹⁸.

En varios casos, las agencias cambian de posición en la jerarquía ministerial, aunque sin una transformación sustancial de sus funciones y dependientes. Por ejemplo, Argentina alojó a los ONE desde 1993 en la Dirección Nacional de Evaluación (DNE) dentro de la Subsecretaría de Promoción y Evaluación de la Calidad que a su vez pertenecía a la Secretaría de Programación del Ministerio; en algunos períodos, fue más bien la Subsecretaría la agencia. Durante la gestión de Lach en el Ministerio se impulsó la creación de una agencia autónoma del Ministerio que incorporara más sistemáticamente y con mayor peso la opinión de las provincias expresadas en el Consejo Federal de Cultura y Educación. El proyecto se concretó a comienzos del 2001 con la creación del Instituto para el Desarrollo de la Evaluación (IDECE). Sin embargo, la crisis política de diciembre de 2001 también arrastró al IDECE, que fue cerrado en 2002, volviendo las tareas de evaluación a ser desarrolladas por una Dirección dentro del Ministerio (actual DINIECE) (Nadorowski, Nores and Andrada, 2002: 20-24).

En la región, sólo hay un país, México, que tiene dos agencias de evaluación activas (INEE, DGE), realizando simultáneamente evaluaciones, ambos programas institucionalizados y públicos (EXCALE y ENLACE, respectivamente). Perú, a partir de 2008 creó el IPEBA pero hasta el presente no ha realizado evaluaciones y según información parcial, no las hará, manteniéndose la UMC como agencia evaluadora¹⁹.

En este contexto, destaca Uruguay donde existen actualmente tres agencias de evaluación: la Dirección de Investigación, Evaluación y Estadísticas (DIEE), la Dirección de Planeamiento de UTU, ambos dependientes de la Administración Nacional de Educación Pública (ANEP) y el creado por ley pero aún no instalado INEE.

3.5. Alternancias y evaluaciones

Cuatro países carecen de programas en la actualidad: Bolivia, Honduras, Nicaragua y Paraguay. Nicaragua nunca tuvo un programa. De hecho, ninguno de los cuatro países parecería tener hoy en día una agencia activa para la evaluación. En el otro extremo, México y Uruguay tienen dos agencias activas, cada una con evaluaciones propias, pero en una situación muy diferente como se ha revisado: el primero con 2 programas institucionalizados y el segundo con programas sin institucionalizar. Estos países, junto con Argentina y Venezuela presentan expedientes de cambios en los partidos de gobierno.

Los cuatro primeros países han tenido alternancia de partidos en el poder junto con cambios drásticos de orientación en las políticas generales de los nuevos gobiernos (Morales, Zelaya, Ortega y Lugo), siendo tal vez Honduras el caso más extremo debido a la altísima inestabilidad política en que se procesó este giro, con el consiguiente golpe de Estado, estado de violencia superviviente aún y nuevo giro político con la

¹⁸ El INEP fue erigido en 1937 dentro del Ministerio. En 1972 fue rediseñado alcanzado el estatuto de organismo autónomo, que mantuvo hasta 1985.

¹⁹ Comunicación personal de Guillermo Ferrer, julio 31 de 2010.

asunción de Lobo. Los cambios además, se han verificado en fechas próximas a la detención de los programas de evaluación. La misma asociación podría aplicársele a Venezuela, aunque Chávez asumió en 1999 y la última aplicación del SINEA de que se tiene noticia habría sido probablemente en 2003 (Ferrer, 2003).

Otros cambios de partido han tenido efectos de reafirmación, aunque no sin conflictos. La alternancia tuvo efectos drásticos en Argentina con la caída de De la Rúa y el interinato de Rodríguez Saá, pero a diferencia de los anteriores, Kirchner estableció la DINIECE y el primer ONE de la nueva era se realizó ya en 2003. Algo similar debiera decirse respecto al Perú con la caída de Fujimori y el gobierno de Toledo. Chile por su parte, cambió no solo de partido o de orientación, sino también de régimen político en 1990, conservando y profundizando el entonces incipiente programa SIMCE, iniciado en 1988.

El caso de México es más especial aún. El año 2000 marcó la alternancia en el gobierno entre el Partido Revolucionario Institucional, en el poder por casi 70 años, y el Partido de Acción Nacional. Si bien en términos ideológicos fuertes, no hubo un cambio profundo en las orientaciones de ambos gobiernos, sí se produjo una modificación fuerte referida a las condiciones institucionales de la evaluación: en 2002 el Presidente Fox creó por Decreto el INEE como organismo autónomo y le asignó la competencia de evaluar la educación básica (inicial, primaria y secundaria, general y técnica).

Sin embargo, otros tantos países han experimentado alternancias de partido con en el gobierno con orientaciones contrapuestas: Brasil, Ecuador, El Salvador, Guatemala y Uruguay. En todos ellos, asumió el gobierno un partido de izquierda sin que esto hubiera significado una suspensión de los programas de evaluación existentes. En El Salvador, el gobierno ha pasado desde la Alianza (derecha) al Frente Farabundo Martí, la ex-guerrilla, sin que esto suspendiera la periodicidad de los programas PAES; es más, continuó con la instrumentación de la PAESITA. En Uruguay, la izquierda en el poder dio una señal de afirmación parlamentaria sobre la necesidad de contar con evaluaciones externas en mejores condiciones institucionales, creando el INEE como persona de derecho público no estatal, autónoma de la autoridad educativa.

Es claro que la alternancia de partidos de gobierno con orientaciones disímiles tiene efecto sobre la política de evaluación, y más en particular, sobre la supervivencia de los programas o sobre condición institucional de las agencias. Sin embargo, algunos ejemplos muestran que salvo extremos ideológicos pronunciados (tal vez Bolivia), estos giros restringen pero no anularían totalmente la evaluación, en particular en aquellos casos en que ha alcanzado un importante periodicidad y volumen: Argentina, Chile, Brasil, El Salvador, México, Perú y Uruguay. Incluso se podría augurar que una vez superado el enfrentamiento civil generado por el golpe de estado en Honduras, se retomaran las actividades de evaluación por parte de la UMCE. En mi opinión, la alternancia tiene un efecto menor sobre la institucionalización que la censura a la publicidad de la información analizada más arriba.

3.6. Principales y agencias

El análisis comparativo de la evaluación externa de aprendizajes en la región muestra diferencias significativas en el diseño institucional de las agencias, en especial, relativas al gobierno y a la dependencia funcional. En todos los países, las agencias de evaluación están vinculadas al Poder Ejecutivo, por lo general, a través del Ministerio, Secretaría o Administración de la Educación. Sin embargo, y aquí está la clave a mi entender, los países se diferencian según si la evaluación es un asunto *exclusivamente* librado a aquellas autoridades, y si esas autoridades *además*, son a su vez, prestadoras de los servicios educativos (hipótesis 4).

Con base en la teoría agente/principal, se ha elaborado el siguiente cuadro que sintetiza la información sobre dos dimensiones. La primera, ubicada en las columnas del cuadro, distingue las agencias que tienen un único principal (en todos los casos el Ministerio, Secretaría o Administración de Educación), de aquellas que tienen más de un principal, generalmente observado en la integración de la junta directiva de la agencia con miembros externos al sistema educativo. La segunda dimensión refiere a si en la estructura del sistema de educación básica se ha separado la regulación de la prestación del servicio (por ejemplo, por la vía de la descentralización o la privatización). Si ambas funciones están concentradas directamente en el Ministerio de Educación, diré que el principal (o primer principal) es prestador de servicios; en caso contrario diré que es regulador. Dentro del cuadro y para cada agencia clasificada, se añaden entre paréntesis el número de evaluaciones que ha realizado la agencia, el número de programas institucionalizados que ha tenido o tiene, y la palabra "SI" indica hechos de censura y no "NO" indica ausencia de censuras.

El resultado es bien interesante. La celda superior izquierda clasificó a las agencias cuyo principal es único y además, prestador del servicio. Ahí están Bolivia, Ecuador, El Salvador, Guatemala, Nicaragua, Panamá, Paraguay y Uruguay, totalizando 11 agencias y 68 evaluaciones. Sin embargo, estos ocho países, siete carecen de un programa institucionalizado de evaluación en marcha (la excepción es El Salvador) y cinco están actualmente vigentes. Lo más notorio es que se identifican episodios de censura en la actuación en siete agencias.

La celda inferior izquierda del cuadro informa de las agencias en las que el único principal, el Ministerio, no es el prestador principal del servicio: Chile, México, Argentina (con dos agencias) y Brasil. Es interesante notar que estos son los países ya identificados con más alto desarrollo de la evaluación educativa. De acuerdo a la hipótesis 1, la evaluación educativa debiera haber alcanzado una alta institucionalización por razones fundamentalmente sistémicas. La evidencia apoya esta idea aunque sólo parcialmente, vista la hipótesis 2B. En la mitad de las agencias relevadas, existieron episodios de censura. Destaca el historial de la DGE de México, aplicando censuras durante toda la década del noventa, a pesar de que el país implementó la federalización. La fecha de la última censura en Argentina corresponde al operativo 2003 de la DINIECE y en realidad como se dijo antes, sólo habría afectado cierto tipo de análisis.

En ambas celdas de la segunda columna, se constata que no hay casos de censura de la información, si bien es cierto que ahí sólo cuentan con tres casos en total: el INEE de México, el ICES de Colombia y la UMCE de Honduras. Dos de estos países tienen un alto desarrollo de la evaluación educativa, en tanto Honduras se encuentra en la situación opuesta. Aquí no hay episodios de censura. Esto es, con independencia de si el primer principal es prestador de servicios, parecería que la clave está en la multiplicidad de principales.

Las tres agencias tienen marcadas diferencias en la integración de su gobierno. El INEE de México está regido por una Junta Directiva de 15 miembros, titulares de organismos gubernamentales o de organizaciones civiles cuyas funciones se relacionan con los objetivos del Instituto: la Secretaría de Educación Pública (que lo preside), el Consejo Nacional de Ciencia y Tecnología, la Secretaría de Hacienda y Crédito Público, el Centro de Investigación y Docencia Económica, el Instituto Mexicano del Petróleo, la Fundación para la Cultura del Maestro, el Sindicato Nacional de Trabajadores de la Educación, el Centro de Investigación y Estudios Avanzados (Politécnico Nacional), la Universidad Pedagógica Nacional, Observatorio Ciudadano de la Educación, Consejo Rector de Transparencia Mexicana, la Unión Nacional de Padres de Familia, la Federación Nacional de Asociaciones de Padres de

Familia y la Comisión de Educación del Sector Empresarial. Esta integración corporativa es típica de la vida política en México: obsérvese que no hay representantes de los partidos políticos ni delegados del Congreso. Sin perjuicio de ello, la integración le otorga legitimidad decisoria sobre evaluación educativa a una gran pluralidad de actores sociales y burocracias públicas. En este otro ámbito, tampoco existe demasiados espacios para las acciones unilaterales y en cambio las constricciones del consenso son también evidentes. Por más que la fijación de una política nacional de evaluación sea competencia compartida entre la SEP y el INEE, esta se debe ventilar, discutir y aprobar en este órgano colegiado con otros actores ajenos a la burocracia escolar que están interesados en particular en reducir las asimetrías de información. Podríamos decir que el "principal" está claramente identificado.

CUADRO 3. CLASIFICACIÓN DE LAS AGENCIAS DE EVALUACIÓN SEGÚN PROPIEDADES DE SU PRINCIPAL

	Principal único	Varios principales
Principal prestador	UMRE, UY (3,1, NO) DIEE, UY (6, 1, SI) MESyFOD (3,1, NO) PE-UTU, UY (4,1, SI) DIR.EVL., PA (6,1, SI) SUBS.PLAN., EC (6,1, SI) DIGEDUCA, GT (10,2, ?) SNEPE, PY, (8,1, SI) SIMECAL, BO (4,1, NO) SNED, NI (3, 0, SI) DEM, SAV (15, 3, SI)	UMCE, HN (5,1,NO)
Principal regulador	UEC, CL (26,3, NO) DGE, MX (35,5,SI) UMC, PE (9, 3, NO) DNE, AR (6,1,NO) IDECE, AR (1,1, NO) DINIECE, AR (3,1,SI) INEP, BR (14,4,NO)	ICFES, CO (8,1, NO) INEE, MX (6,1,NO)
<p>Nota: Se incluye entre paréntesis: el número de evaluaciones nacionales hechas, el número de programas institucionalizados y si hubo (SI) o no casos de censura (NO) de los informes Fuente: elaboración propia con base en relevamiento realizado en 2006 y 2010.</p>		

El ICFES por su parte está dirigido por tres miembros, nombrados dos por el Presidente de la República y un tercero por el Ministro de Educación. Sin dudas, esta composición le otorga una menor representatividad a la diversidad de intereses existentes, pero cumple con darle mayor peso a los actores externos al sistema educativo, que por definición no tienen intereses contradictorios en tanto prestadores del servicio.

La UMCE por su parte nació y mantuvo hasta su última época, una naturaleza híbrida: surgida como parte de un convenio entre la Secretaría de Educación Pública y la Universidad Pedagógica Nacional Francisco Morazán, conservó un estatuto autónomo propio de la vida universitaria, más que de la vida política; dicho esto en el contexto muy relativo de la vida política hondureña.

Vale la pena indicar que el establecido INEE de Uruguay clasificaría en esta celda, dado que si bien su primer principal es prestador del servicio, su comisión directiva estaría integrada por un miembro designado por el Ministerio de Educación (que lo presidiría); tres miembros designados por la ANEP, dos

por la Universidad de la República y uno por el sector privado.

4. CONSIDERACIONES FINALES

La evidencia relevada en 15 países de la región muestra que la evaluación educativa es un fenómeno de una magnitud muy importante tanto por la cantidad de levantamientos (nacionales e internacionales) realizados en los últimos 30 años, como por el número de programas nacionales de evaluación puestos en marcha. Sin embargo, existen diferencias notorias. Por un lado, es visible un grupo de países con un alto desarrollo en la materia (Argentina, Brasil, Chile, México y Colombia). Todos han realizado extensas reformas de segunda generación en sus sistemas educativos. Por otro lado, al menos cuatro países presentan debilidades importantes debido a la inexistencia actual de programas (Nicaragua, Bolivia), la inactividad de la agencia en los últimos dos años (Honduras, Paraguay) o el reducido número de evaluaciones hechas, todas en los últimos años (Panamá, Guatemala). En todos, las reformas de segunda generación han sido muy parciales o han fracasado. Esto avalaría la primera hipótesis propuesta. Sin embargo, entre estos dos grupos de países, se ubican Perú, El Salvador y Uruguay, los que aún teniendo niveles semejantes en cuanto al número de evaluaciones, difieren entre sí tanto en la extensión de sus reformas como en la institucionalización de la evaluación. Destaca Uruguay por carecer de un programa institucionalizado *vigente*.

El examen de la hipótesis de la orientación de los partidos de gobierno y su alternancia han permitido mostrar que este fenómeno político tiene una incidencia notable en la supervivencia tanto de programas como de agencias. Sin embargo, así como hay casos notables que confirmarían la hipótesis, también hay otros que la disconfirman. Las excepciones, precisamente, parecen más relacionadas con la propia fortaleza de las instituciones de evaluación.

El análisis final practicado clasificó a las agencias según su dependencia funcional, distinguiendo cuatro posibilidades por la combinación de principal regulador / principal prestador y principal único / múltiples principales. Los tres indicadores de institucionalidad muestran un alineamiento bastante notorio: aquellos países cuyas agencias están subordinadas a un único principal y éste además es prestador del servicio educativo, tendrían mayores probabilidades de presentar episodios de censura a la distribución de la información. Dicho de otra forma, entre todas las instituciones que regulan la evaluación, la fortaleza o debilidad parecería ser el resultado de las condiciones en que es gobernada la agencia de evaluación. A pesar de que el modelo de análisis no está controlando simultáneamente el conjunto de variables que indican las hipótesis desplegadas, se puede avanzar refinando la hipótesis 4. Si bien es razonable atribuir la institucionalización de la evaluación educativa a las estructuras descentralizadas de los sistemas, y que las orientaciones liberales de los gobiernos fomentan la redistribución de información, parecería que son condiciones suficientes. El más alto grado de institucionalización se alcanza en los países que cuentan, además, con agencias no subordinadas exclusivamente al Ministerio. La conjunción de intereses, originarios de las relaciones evaluativas y prestatarias en el principal de las agencias de evaluación, parecería proporcionar una precaria institucionalidad a la evaluación educativa en términos de programas, tal como aquí se lo ha entendido. Este es un resultado muy importante en la discusión de las políticas sobre la materia. Si bien la bibliografía especializada ha destacado los aspectos institucionales, en general los ha puesto en un segundo plano a la hora de formular recomendaciones sobre cuáles son las evaluaciones que América Latina necesita. Sobre todo es un resultado fundamental para el Uruguay, país que a pesar de haber creado por Ley un Instituto autónomo, aún no ha sido

instrumento y persisten dudas sobre su contribución al mejoramiento de la calidad de la educación.

REFERENCIAS BIBLIOGRÁFICAS

- Benavidez, V. (2010). Las evaluaciones de logros educativos y su relación con la calidad de la educación. *Revista Iberoamericana de Educación*, 53, pp. 83-96.
- Benveniste, L. (2000). Student Assessment as a Political Construction: The Case of Uruguay. *Education Policy Analysis Archives*, 8. Consultado el 7 de agosto de 2010 en <http://epaa.asu.edu/ojs/article/viewFile/423/546>.
- Benveniste, L. (2002). The Political Structuration of Assessment: Negotiating State Power and Legitimacy. *Comparative Education Review*, 46(1), pp. 89-118.
- Bertolini, M. et al. (2001). *Educación Secundaria: la reforma impuesta. Días visiones críticas*. Montevideo: FENAPES / NORDAN.
- Chubb, J.E. and Moe, T. (1990). *Politics, Markets and America's schools*. Washington: The Brookings Institution.
- DIEE (2002 a). *Los niveles de desempeño al inicio de la educación primaria. Estudio de las competencias lingüísticas y matemáticas. Evaluación nacional en el primer nivel de la escolaridad: inicial 5, 1er. y 2do año. Primer informe. Setiembre 2002*. Montevideo: General de Planemiento y Gestión Educativo, Gerencia de Investigación y Evaluación / ANEP.
- DIEE (2002 b). *Evaluación nacional de aprendizajes en lenguaje y matemática. 6to año de Enseñanza Primaria. - 2002. Primer Informe de devolución de resultados de la muestra nacional*. Noviembre 2002. Montevideo: General de Planemiento y Gestión Educativo, Gerencia de Investigación y Evaluación / ANEP.
- DIEE (2004). *Informe Nacional PISA 2003*. Montevideo: ANEP. Elaborado por Ravela, P.
- DIEE (2007). Uruguay en PISA 2006. Primeros resultados en ciencias, matemáticas y lectura. Montevideo: ANEP.
- Di Gropello, E. (1999). Los modelos de descentralización educativa en América Latina. *Revista de la CEPAL*, 68, pp.153-170.
- Di Gropello, E. (2004). La descentralización de la educación y las relaciones de rendición de cuentas en los países latinoamericanos. *Documento de trabajo*, 30. PREAL: Santiago de Chile.
- Eisenhardt, K. (1989). Agency Theory: An Assessment and a Review. *Academy of Management Review*, 14 (1), pp. 57-74.
- Fernández, T. (2004). De las escuelas eficaces a las reformas de segunda generación. *Estudios Sociológicos*, 64, pp. 377-408.
- Fernández, T. (2005) Persistent Inequalities in Uruguayan Primary Education 1996 - 2002. In Teese, Richard; Lamb, Stephen and Duru-Bellat, Marie *International Studies in Educational Inequality, Theory and Policy*, 2, pp. 177-206. Dordrecht: Springer.
- Fernández, T. (2008). *El Proyecto Instituto de Evaluación Educativa del Uruguay: ¿un paso hacia la rendición de cuentas?*. Segundo Congreso Uruguayo de Ciencia Política. Montevideo.

- Fernández, T. (2009). La instalación del nuevo Instituto de Evaluación Educativa de Uruguay: desafíos organizacionales e institucionales. En Caetano, G. *Encrucijada 2010: la política uruguaya a prueba. Observatorio Político. Informe de coyuntura n°9*. Montevideo: Editorial Fin de Siglo / Instituto de Ciencia Política/CLACSO.
- Fernández, T. and Midaglia, C. (2005). *Quiénes y cómo usan los informes generados por los sistemas de evaluación de aprendizajes en la educación primaria. Los casos de México y Uruguay*. Informe Final de Investigación. México: El Colegio de México.
- Fernández, T. and Midaglia, C. (2005). El uso de los informes generados por los sistemas de evaluación de aprendizaje en la educación primaria. Los casos de México y de Uruguay. En Cueto, S. (ed) *Uso e impacto de la información educativa en América Latina*. Santiago de Chile: PREAL.
- Ferrer, G. (2006). *Sistemas de evaluación de aprendizajes en América Latina. Balance y desafíos*. Santiago de Chile: PREAL.
- Gauri, V. and Vawda, A. (2003). Vouchers for Basic Education in Developing Countries. A Principal-Agent Perspective. Washington, *World Bank Policy Research Working Paper 3005*.
- Hevia Rivas, R. (1992). *Política de descentralización en la educación básica y media en América Latina*. Santiago de Chile: UNESCO/REDUC.
- Hirshman, A. (1970). *Exit, voice and loyalty*. MA: University of Cambridge Press.
- Knight, Jack (1997). *Institutions and social conflict*. MA: Cambridge University Press.
- Martínez Rizo, F. (2010). Políticas de evaluación educativa y organismos a cargo de ellas. Experiencias y tendencias recientes en América Latina. Ponencia a las *Jornadas de Educación: una mirada a 10 años*. Montevideo: Universidad Católica del Uruguay.
- Murillo, J. and Román, M. (2010). Retos de la evaluación de la calidad de la educación en América Latina. *Revista Iberoamericana de Educación, 53*, pp. 83-96.
- Narodowski, M., Nores, M. and Andrada, M. (2002). *La evaluación educativa en la Argentina. Desde los operativos nacionales hasta los boletines escolares*. Buenos Aires: Prometeo 3010.
- North, D. (1995). *Instituciones, cambio institucional y desempeño económico*. México: Fondo de Cultura Económica.
- Postlethwaite, N. and Kellaghan, T. (2009). National Assessment of educational achievement. *Education Policy Series n°9*. Ginebra: UNESCO.
- Ravela, P. (2001). *¿Cómo presentan sus resultados los sistemas de evaluación de aprendizaje de América Latina?*. Documento de trabajo-Grupo de Estándares y Evaluación de Aprendizajes. Santiago de Chile: Grade/PREAL.
- Ravela, P., Picaroni, B., Fernández, T., Gonet, D., Haretche, C., Loureiro, G. and Luaces, O. (2000). *Evaluaciones nacionales de aprendizajes en Educación Primaria en el Uruguay (1995-1999)*. Montevideo: Ed. UMRE-MECAEP-ANEP.
- Ravela, P., Arregui, P., Valverde, G., Wolfe, R., Ferrer, G., Martínez, F., Aylwin, M. and Wolf, L. (2008). Las evaluaciones de aprendizaje que América Latina necesita. *Documento de Trabajo n°40*. Santiago de Chile: PREAL. En: http://www.preal.org/Biblioteca.asp?Id_Carpeta=64&Camino=63|Preal%20Publicaciones/64|PREAL%20Documentos.

- Ruiz, G. (2003). Proyectos de evaluación de la educación básica en la década 1990-2000. En Zorrilla, Margarita (ed.) *La evaluación de la educación básica en México 1990-2000. Una mirada a contraluz*. Aguascalientes: Universidad Autónoma de Aguascalientes.
- UMRE (1996a). *Evaluación nacional de aprendizajes. Fundamentos*. Montevideo: ANEP.
- UMRE (1996b). *Primer informe público de resultados. Evaluación Nacional de Aprendizajes en Lengua Materna y Matemática. 6to año de Enseñanza Primaria 1996*. Montevideo: ANEP.
- UMRE (1997a). *Segundo Informe público de resultados. Evaluación Nacional de Aprendizajes en Lengua Materna y Matemática. 6to año de Enseñanza Primaria 1996*. Montevideo: ANEP.
- UMRE (1997b). *Tercer Informe Público. Evaluación Nacional de Aprendizajes en Lengua Materna y Matemática. 6tos año de Enseñanza Primaria 1996*. Montevideo: ANEP.
- UMRE (1999). *Evaluación Nacional de Aprendizajes en sextos años de Educación Primaria. Primer Informe*. Montevideo: ANEP.
- Winkler, D. and Gershberg, A. (2000). Los efectos de la descentralización del sistema educacional sobre calidad de la educación en América Latina. *Documento de Trabajo n° 17*. Santiago de Chile: PREAL.
- Winters, M. (2006). La elección de escuelas en Estados Unidos. En Corvalán, J. and McMeekin, R.W. (ed.) *Accountability educational: posibilidades y desafíos para América Latina a partir de la experiencia internacional*. Santiago: CIDE/PREAL.
- Wolff, L. (2004). Educational Assessment in Latin America: the State of the Art. *Applied Psychology*, 53(2), pp.192-214.
- Zorrilla, M. (ed.) (2003). *La evaluación de la educación básica en México 1990-2000. Una mirada a contraluz*. Universidad Autónoma de Aguascalientes: Aguascalientes.

ANEXO 1

Participación en evaluaciones internacionales

TIMSS 1995	TIMSS 1999- R	LLECE 1997	CIVIC EDUC. (1994-2002)	PISA 2000 PISA PLUS	PIRLS 2001	TIMSS 2003	PISA 2003	LLECE 2006	PISA 2006	TIMSS 2007	PISA 2009	Iccs 2009	TOTAL POR PAIS
Argentina		Argentina		Argentina	Argentina	Argentina		Argentina	Argentina		Argentina		8
		Bolivia											1
		Brasil		Brasil			Brasil	Brasil	Brasil		Brasil		6
	Chile	Chile	Chile	Chile		Chile		Chile	Chile		Chile	Chile	10
Colombia		Colombia	Colombia		Colombia			Colombia	Colombia	Colombia	Colombia	Colombia	9
		Costa Rica						Costa Rica					2
		Cuba						Cuba					2
								Ecuador					1
								El Salvador		El Salvador			2
								Guatemala				Guatemala	2
		Honduras											1
México		México		México			México	México	México		México	México	8
								Nicaragua					1
		Paraguay						Paraguay					2
								Panamá			Panamá		2
		Perú		Perú				Perú			Perú		4
		Rep. Dom.						Rep. Dom.			Rep. Dom.	Rep. Dom.	3
							Uruguay	Uruguay	Uruguay		Uruguay		4
		Venezuela											1
3	1	13	2	5	2	2	3	16	6	2	9	5	69

Fuente: elaboración propia con base en el relevamiento comparado. En anexos se detallan las fuentes específicas para este punto.