



Construyendo juntos la Nueva Escuela Pública Paraguaya



Ficha técnica

Presidente de la República

Fernando Armindo Lugo Méndez

Ministerio de Educación y Cultura

Autoridades

Ministro de Educación y Cultura

Luis Alberto Riart Montaner

Viceministra de Educación para la Gestión Educativa

Diana Carolina Serafini Fernández

Viceministro de Educación para el Desarrollo Educativo

Héctor Salvador Valdez Alé

Viceministra de la Juventud

Diana Beatriz García Galeano

Viceministro de Culto

Hugo Antonio Britez Ibarra

Directora General de Planificación Educativa

Mirna Vera Notario

Directora de Evaluación de la Calidad Educativa

Graciela Sosa Salgueiro

MINISTERIO DE EDUCACIÓN Y CULTURA

Sede Central "Ramón Indalecio Cardozo"

15 de agosto 629 entre Gral. Díaz y Eduardo V. Haedo

Asunción, Paraguay

Telefax: (59521) 450-014/015

www.mec.gov.py

ISBN: 978-99953-99-27-6

Publicación © Setiembre 2011



Evaluación educativa: un tema central en la agenda de la Nueva Escuela Pública Paraguaya

Coordinador:

Luis Alberto Riart M.

Equipo técnico:

Mirna Vera Notario
Graciela Sosa Salgueiro
Carlos A. Pardo Adames

Colaboración especial de:

Juan Carlos Tedesco

Editora:

Adriana Pessoa Nardi





Evaluación educativa: un tema central
en la agenda de la Nueva Escuela
Pública Paraguaya

4





Índice

Introducción

Dr. phil. Luis Alberto Riart Montaner, Ministro de Educación y Cultura de Paraguay 7

Capítulo I: Evaluación educativa: un tema central para el mejoramiento de la calidad 11

- 1. La historia de la evaluación 12
- 2. Los últimos años de la evaluación 19
- 3. Evaluaciones internacionales. 21
 - 3.1. Association for the Evaluation for Educational Achievement – IEA 21
 - 3.2. Educational Testing Service (ETS) – IAEP 22
 - 3.3. Programa para la Evaluación Internacional de Alumnos – PISA 23
 - 3.4. Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación – LLECE 24
- 4. Pertinencia de participar en evaluaciones internacionales . . . 25
- 5. Una breve historia de los desarrollos técnicos. 26

Capítulo II: Reflexiones sobre la evaluación. 29

- 2.1 Reflexiones sobre evaluación de los aprendizajes. 29
- 2.2 Reflexiones sobre la evaluación educativa en Paraguay. . . . 32

Capítulo III: Las evaluaciones educativas en Paraguay 35

- 3.1. Sistema Nacional de Evaluación del Proceso Educativo – SNEPE 35
- 3.2. Análisis, difusión y uso de los resultados 40
- 3.3. Aplicación 2010: los primeros resultados 41

Capítulo IV: Innovación en la evaluación educativa: evaluaciones mediante el uso de celulares 45

- 4.1 ¿Cómo opera el sistema? 47
- 4.2 Primera aplicación: principales resultados. 48
- 4.3 Conclusiones sobre la primera experiencia 50



Capítulo V: Evaluaciones internacionales de las que participa Paraguay	51
5.1. Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE – UNESCO).	51
5.2. Estudio Internacional de Educación Cívica y Ciudadana (International Civic and Citizenship Education Study - ICCS)	52
5.3. Programa de Evaluación y Monitoreo de la Alfabetización.	57
Capítulo VI: Medición de impacto del preescolar en el primer ciclo de la Educación Escolar Básica	63
Introducción	63
6.1. Marco del estudio	64
6.2. Objetivo del estudio.	65
6.3. Hipótesis del estudio.	65
6.4. La muestra	66
6.5. Metodología del estudio	68
6.6 Fases de estudio	70
6.7. Análisis de resultados	73
6.8. Comentarios finales	84
Bibliografía	87



Introducción

La Nueva Escuela Pública Paraguaya asumió el desafío de la educación de calidad. Contrariamente a las tendencias de las reformas de los '90 en América Latina de priorizar el acceso a la educación por sobre la calidad, esta administración determinó que propiciar y garantizar la universalización de la educación no puede transitar un carril separado de la calidad de la misma.

Ella es un tema central en la agenda educativa, y desde el compromiso con esa centralidad se realizan intervenciones relacionadas al mejoramiento de la formación de los educadores y educadoras, así como de las condiciones de aprendizaje de las y los estudiantes a través de la dotación de textos, bibliotecas de aulas, incorporación de TIC, entre otros. En este proceso es indispensable contar con información sobre la calidad del aprendizaje del alumnado, y los factores asociados a dichos aprendizajes de tal forma a diseñar políticas públicas que contribuyan al mejoramiento de la calidad de la educación.

En este marco y dada la relevancia del tema, el Ministerio de Educación y Cultura ha puesto de nuevo en el debate público la importancia de la evaluación de la calidad de la educación, sin perder de vista el camino ya transitado. Se reconocen avances importantes en materia de evaluación en nuestro país, sin embargo, se requiere de un fortalecimiento y de una redefinición de los objetivos y uso de los resultados de la evaluación, atendiendo a que una de las principales debilidades observadas en todo este tiempo ha sido la disociación entre los resultados de las evaluaciones y el diseño de las políticas.

Entendemos que la evaluación en sí misma no mejora la calidad, pero nos sustentamos en que evaluamos para mejorar, asumiendo que para lograrlo es indispensable tener conciencia de la magnitud del problema y hacerse cargo de él. Los resultados deben ayudar a tomar decisiones, por ende el objetivo central de las evaluaciones es el de contribuir al mejoramiento de la calidad de los aprendizajes



y reducir la desigualdad. En este sentido, es imprescindible no sólo evaluar los aprendizajes, sino también evaluar el impacto de las políticas.

Para que la evaluación cumpla con su objetivo se debe lograr la participación de los actores pedagógicos en el proceso de evaluación, análisis y uso de los resultados. Por lo tanto, es fundamental que quienes toman las decisiones de la política pública, quienes investigan, quienes ejercen la docencia, las universidades, directores y directoras, y las familias se sientan partícipes y responsables de los resultados educativos. Los resultados de las evaluaciones deben estar asociados con propuestas curriculares acordes a las necesidades detectadas; así mismo, deben ser la base para la elaboración de materiales para las y los docentes de tal manera que puedan mejorar la práctica y analizar sus efectos.

Lo que emerge con absoluta claridad en torno al tema de la evaluación es que el sentido de la misma es aportar datos, estadísticas, información vital para la toma de decisiones en torno a las estrategias de la política educativa, a fin de que la misma responda con pertinencia y eficiencia a las necesidades, al diagnóstico y a las metas de mejoramiento de la calidad de los aprendizajes y de la gestión. Este es el sentido que la sociedad toda tiene el desafío de incorporar en relación al tema de la evaluación, lo que representa también para el MEC un enorme reto, más aún si tenemos en cuenta el carácter punitivo con el que se ha asumido el tema de la evaluación por los distintos actores del sistema educativo.

En torno a las reflexiones sobre la evaluación, surge también la pregunta ¿Qué institución es la que tiene que medir y evaluar? Nos planteamos entonces la necesidad de definir si las evaluaciones deben ser desarrolladas por el MEC o por un organismo externo. Ambas posibilidades representan ventajas y desventajas: si son realizadas por instancias del MEC, las mismas pudieran perder niveles de objetividad, pero ganar en las posibilidades reales de incorporación de estrategias al interior del sistema educativo, dado que están cerca de la acción. En contrapartida, si quienes evalúan están fuera del sistema educativo, su lejanía de la acción cotidiana



podría impedir o dificultar la incorporación de acciones al interior del sistema educativo, pero podría ganarse en objetividad.

Actualmente estamos evaluando la posibilidad de la creación de un Instituto de Evaluación de la Calidad Educativa, con autonomía pero vinculado a la acción, y hemos incorporado por primera vez, en el Anteproyecto de Presupuesto 2012, una línea presupuestaria que permita la sostenibilidad del Sistema de Evaluación.

Este libro busca aportar al análisis, al debate público y a las consideraciones sobre la evaluación de la calidad en el sistema educativo, además de dar cuenta de los trayectos recorridos, las metas alcanzadas y los desafíos pendientes en torno a la temática. En ese sentido, el documento reúne información acerca de la evaluación en nuestro país, y de las evaluaciones regionales y estudios internacionales en los que Paraguay participa. Se presentan los marcos conceptuales, metodológicos y los procesos técnicos desarrollados, así como reflexiones en torno a la evaluación educativa y la evolución histórica de la misma desde sus inicios hasta la actualidad. En la línea de la transparencia de la gestión y de la información, se comparten los primeros resultados de aprendizaje de los estudiantes del 1º, 2º y 3º ciclo de la Educación Escolar Básica en la aplicación llevada a cabo en noviembre del año 2010, resultados que muestran las deudas históricas que nos quedan por saldar y los retos que nos mueven a ajustar las estrategias pedagógicas y a innovar permanentemente en la búsqueda de la calidad de los aprendizajes.

Fortaleciendo la transparencia en la gestión y promoviendo la construcción participativa basada en el conocimiento, la información y el intercambio de opinión en una retroalimentación continua entre los agentes educativos y la sociedad civil, se hace entrega de este material sobre las experiencias educativas en materia de evaluación emprendidas por el MEC, iniciando así un proceso de socialización de la información, proceso que debe llevarnos además a realizar alianzas con comunidades académicas, universidades, poniendo a disposición la base de datos e informaciones requeridas para la investigación y/o proyectos educativos, entre otros temas.



En la construcción de la Nueva Escuela Pública Paraguaya, el MEC asume su responsabilidad e incorpora estrategias concretas en la búsqueda del mejoramiento de la calidad de la educación y al mismo tiempo entiende que las acciones educativas no deben ser aisladas, por lo que indefectiblemente deberán articularse los esfuerzos educativos con otras políticas públicas y con todos los actores sociales, basados en el principio de la participación plural para construir las respuestas educativas, pero al mismo tiempo comprometidos con la obligación del Estado de ser garante del inalienable derecho de cada ser humano a una educación de calidad a lo largo de toda la vida.

Dr. phil. Luis Alberto Riart Montaner
Ministro de Educación y Cultura
República del Paraguay



Capítulo I

Evaluación educativa: un tema central para el mejoramiento de la calidad

Mirna Vera¹

Carlos A. Pardo Adames²

En la década de los años 90 el auge de las reformas educativas en América Latina ha tenido como principal foco la expansión de los sistemas educativos, pero esta expansión se ha dado sin la calidad deseada. Es así que los resultados de las evaluaciones realizadas han generado la creciente preocupación de los países por mejorar la calidad de la educación e instalado la evaluación educativa en el centro del debate.

Este debate cobra aún mayor importancia en el contexto educativo nacional donde los resultados de las evaluaciones aplicadas a estudiantes de los diferentes niveles educativos develan debilidades importantes en cuanto al desempeño de los estudiantes. Estos resultados lejos de descalificar a estudiantes y docentes, constituyen un aporte importante al proveer información sobre el estado y situación del sistema educativo.

En este marco, la evaluación educativa aporta información objetiva sobre la situación educativa vinculada a las condiciones y variables que inciden en los aprendizajes y al mismo tiempo, apoya la toma de decisiones. La evaluación, por tanto, contribuye a la retroalimentación, la reorientación y planificación de las propuestas y prácticas educativas.

1 Directora General de Planificación Educativa del MEC.

2 Especialista en Evaluación. Consultor Internacional.



La evaluación ha experimentado transformaciones importantes empujadas por una la realidad educativa y social cada vez más compleja, convirtiéndose actualmente en una disciplina científica y una práctica profesional.

1. La historia de la evaluación

Se puede decir que la educación existe desde el propio inicio de la humanidad y que la evaluación aparece desde el inicio de la educación. Es por esto, que no es raro encontrar casos de evaluación educativa en toda la historia de la humanidad y en todas las diversas prácticas educativas que se han realizado. Igualmente, la evaluación ha adquirido sus propias características que la hacen diferente en cada nivel y contexto educativo; es así que se puede hablar de la evaluación en el aula de clase como diferente a la realizada a nivel institucional.

Es en China, en la antigüedad, donde podríamos encontrar la mayor cantidad de ejemplos de evaluación educativa antes de la época moderna. La historia al respecto se remonta a 2200 años a. C.³ cuando el emperador chino hacía que sus funcionarios presentaran pruebas para determinar si eran aptos o no para desempeñarse en el servicio civil. Esos exámenes se refinaron hasta que se introdujeron las pruebas escritas en la Dinastía Han (202 a. C. hasta EL 200 d. C.), fecha en la cual se empezó a evaluar cinco tópicos:

- Leyes civiles.
- Cuestiones militares.
- Agricultura.
- Impuestos.
- Geografía.

La evaluación en China tomó su forma más compleja a partir de 1374 hasta la Dinastía Ching (1644 - 1911)⁴, en donde, además de

3 DuBois, P. H. (1970). *History of Psychological Testing*. Boston. Allyn & Bacon.

4 Greaney, V. Y Kellaghan T. (1995) *Equity Issues in Public Examinations in Developing Countries*. Washington. World Bank.



los temas anteriores, se enfatizaba el conocimiento que se tenía sobre las ideas de Confucio. Los candidatos a ocupar cargos públicos pasaban por tres etapas de evaluación⁵:

- En la primera, el evaluado debía escribir ensayos sobre temas predeterminados y un poema. Pasaban entre el 1% y el 7% de quienes se presentaban.
- Las evaluaciones del nivel de exámenes de distrito tenían las mismas características de las anteriores pero eran más rigurosas y extenuantes. Pasaban entre el 1% y el 10% de quienes alcanzaban este nivel.
- Los que pasaban las etapas anteriores se presentaban en Pekín a la evaluación final. Sólo pasaba el 3% de quienes se presentaban.

El sistema chino de evaluación influyó notablemente en la aparición de pruebas escritas en las escuelas europeas del siglo XVI y, un par de siglos más tarde, se instituyeron en Europa sistemas de selección a la universidad y al servicio civil, que recordaban las prácticas desarrolladas en China. Estos exámenes públicos han sido una característica esencial de los sistemas educativos europeos desde entonces, idea que posteriormente se fue expandiendo en sus colonias: África, Asia y el Caribe. Hoy en día son una práctica muy frecuente en la mayoría de los países y una fuente rica de información para la transformación de los sistemas educativos.

Estas ideas, en conjunción con la aparición de la psicología experimental, son las que dieron forma a la actual evaluación educativa. En la segunda mitad del siglo XIX, Wilhelm Wundt funda el primer laboratorio de psicología en Leipzig (1879) y experimenta, con su **metro de pensamiento**, un aparato que pretende establecer la diferencia entre lo percibido por un sujeto y lo que acontece en la realidad. Sus resultados aportan esencialmente al concepto de medición de las diferencias individuales, lo que contribuyó posteriormente al desarrollo del uso de pruebas en psicología para evaluar las diferencias individuales.

⁵ DuBois P. H. Op. Cit.



Casi de inmediato se desarrolla la medición psicológica conocida como la Época de Cobre, debido al material del cual estaban hechos la mayoría de los instrumentos de medición, cuyos principales exponentes son Sir Francis Galton en Gran Bretaña y James McKeen Cattell.

Sir Francis Galton establece, en 1884, el primer laboratorio de psicometría. Allí se encuentran una serie de instrumentos que miden dos aspectos del ser humano, considerados como esenciales por la ciencia de entonces: aspectos físicos y conductuales.

Luego, en 1890, James M. Cattell, alumno de Galton y dedicado a la medición de diversos aspectos del ser humano, utiliza, por primera vez, las palabras “Test Mentales” relacionadas con la utilización de 10 pruebas que propone aplicar al público en general. Algunas de las pruebas de Cattell se mantienen en la tradición de medición de Galton, como por ejemplo, la fuerza requerida para exprimir, el tiempo de reacción a un sonido, el tiempo utilizado para nombrar colores, el grado de presión necesaria para ocasionar dolor⁶, sensibilidad a la diferencia de pesos, entre otras.

A partir de estos trabajos, se presenta un especial desarrollo en la psicometría que realiza estudios de evaluación del ser humano que se originan especialmente en el sistema educativo, y que dan inicio a una nueva época en evaluación conocida como de pruebas (testing). La primera evaluación de este tipo se realiza en Francia en 1905⁷. La Sociedad Libre para el Estudio Psicológico del Niño se encontraba interesada en mejorar la efectividad y eficiencia de las escuelas identificando las causas del fracaso escolar. Desde su perspectiva consideraban que habían dos clases de niños que fracasaban: aquellos que podían aprender pero no lo hacían y aquellos que no podían aprender. Alfred Binet, quien se encontraba vinculado con la sociedad mencionada en la época (1904), se dio a la tarea de determinar si el bajo nivel académico de un niño se debía a retardo mental o a alguna otra causa.

6 Galton había mostrado que personas retardadas mentales eran relativamente insensibles al dolor, lo que llevó a Cattell a asumir que la sensibilidad al dolor era buen predictor de la inteligencia.

7 Thorndike, R. (1997). The Early History of Intelligence Testing. En Contemporary Intellectual Assessment. New York. The Guilford Press.



En 1905, un comité formado por el ministerio público francés, e integrado por Binet y otros miembros de la sociedad mencionada, presentaron la escala Binet-Simon, de 30 tareas cortas, ordenadas por dificultad. El nivel intelectual del niño se definía por la tarea más difícil que pudiera ejecutar correctamente. El cociente intelectual que introdujo más tarde, que relaciona este nivel de habilidad con la edad.

El siguiente paso importante en el desarrollo de la evaluación fue la posibilidad de aplicación de pruebas a grupos grandes de personas. Hasta el momento las pruebas o evaluaciones importadas de Francia eran esencialmente de aplicación individual o, en algunos casos en que se podían aplicar a grupos, la calificación dependía del juicio del examinador. Con el ingreso de Estados Unidos a la I Guerra Mundial⁸ se desarrolló a pasos agigantados la posibilidad de aplicación y calificación masiva de pruebas.

En esa época, el gobierno de los Estados Unidos encomendó a Robert Yerkes la evaluación de inteligencia de 1.750.000 nuevos reclutas del ejército. Yerkes formó un comité que desarrolló los conocidos Army Test (Alpha y Beta). El test Alpha se basó en trabajos no publicados de Otis y consiste en ocho sub-pruebas:

- Seguimiento de instrucciones orales.
- Razonamiento aritmético.
- Razonamiento práctico.
- Sinónimos – antónimos.
- Oraciones en desorden.
- Series de números.
- Analogías.
- Información.

La noción de pruebas de lápiz y papel -esquema de evaluación desarrollado para los Army test- se aplicó, inmediatamente después de la I Guerra Mundial, en la industria y en la educación⁹. Se desarrollaron pruebas que se aplicaron a millones de personas durante la década del 20, como el National Intelligence Test.

⁸ DuBois Op. Cit.

⁹ Aiken, Lewis. (1996). Test Psicológicos y Evaluación. México D. F. Prentice Hall.



El inicio de la evaluación en educación tal y como se la conoce hoy en día se da cuando el College Entrance Examination Board, diseña, en 1925, el Scholastic Aptitude Test (SAT) compuesto de preguntas de selección múltiple relacionadas con las oraciones incompletas, analogías y secuencia de números. Esta prueba se administraba a los estudiantes que quisieran ingresar a las universidades en Estados Unidos. La función del Board mencionado fue asumida posteriormente por el Educational Testing Service, entidad que ha diseñado, estandarizado y validado otras pruebas igualmente conocidas¹⁰.

En 1925 se elaboran pruebas en forma de test con preguntas de selección múltiple, lo que contribuye con el desarrollo de cuestionarios de fácil calificación ocasionando el desarrollo de tecnología que permita la calificación de cuestionarios por medios mecánicos como las máquinas de lectura óptica inventadas en la década del 30. Estos hechos han marcado especialmente el camino que ha tomado la evaluación educativa en el resto de siglo.

En 1957 sucedió un hecho que marcó definitivamente el desarrollo de la educación y de la evaluación educativa, principalmente en Estados Unidos¹¹ y como consecuencia, en el resto del mundo. La Unión de Repúblicas Socialistas Soviéticas (URSS) lanzó al espacio el primer satélite artificial en la historia de la humanidad: el Sputnik I; algunos meses más tarde lanzó al primer ser viviente: la perra Laika y un poco más tarde, al primer ser humano: el soviético Yuri Gagarin; en realidad no demoró mucho el viaje de la primera mujer: Valentina Teleskova.

Esto llevó a que, de inmediato, se iniciara un cuestionamiento del sector educativo en Estados Unidos, lo que tuvo como consecuencia inmediata la evaluación de sus diferentes componentes: los estudiantes, los currículums, los docentes, los programas, la parte administrativa, etc. Se destinaron enormes cantidades de dinero a la evaluación y a la educación y se promulgaron leyes como la Ley de Educación Elemental y Secundaria de 1965, que favorecía el

10 DuBois Op. Cit.

11 Wolf, Richard. (1987). Educational Evaluation. En International Journal of Educational Research. Vol. 11. Londres. Pergamon Press.



desarrollo de propuestas educativas en los distritos escolares pero sólo si se evaluaban los desarrollos educativos previos. Como consecuencia, los administradores educativos buscaron afanosamente el tipo de evaluaciones educativas que se realizaba en la época para implementarlas lo más pronto posible: las pruebas objetivas.

Toda esta fortaleza que se le dio a la evaluación en educación ocasionó que se convirtiera en una especialidad de la educación desde principios de la década del 60 y que perdiera el sentido que tenía hasta el momento en el contexto educativo: que la evaluación es parte integral del proceso educativo, sentido que está recuperando en los últimos años.

En la década del 60 se retoman tres perspectivas que unidas se constituyen en el paradigma de la época¹². El libro de Ralph Tyler (1950) “Principios Básicos del Currículo y la Instrucción”, en el que se propone que la evaluación debe ser parte integral del proceso educativo y debe hacerse para determinar el grado en el cual un programa evaluado promueve el logro de objetivos educacionales.

Estos planteamientos se relacionan con los de Robert Mager (1962) en su libro “Preparando Objetivos para la Instrucción Programada”, donde se plantean los criterios para la elaboración y formulación de objetivos educacionales. Fue tal el impacto de estos dos libros que muchos evaluadores educativos de principios de los 60, asumieron que la única forma de evaluar un programa educativo era constatar si se habían logrado los objetivos educacionales planteados en forma medible.

El tercer elemento surge de los planteamientos de B. F. Skinner (1958) en su libro “Máquinas de Enseñanza”, donde se plantean principios de aprendizaje establecidos en el laboratorio, para la enseñanza de niños en las escuelas. Esencialmente se plantea la necesidad de dar material instruccional de manera secuencial, con refuerzos positivos para los aprendices.

12 Popham, James. (1987). Two-Plus Decades of Educational Objectives. En: International Journal of Educational Research. Vol. 11. Londres. Pergamon Press.



Adicionalmente, se integran, desde el punto de vista de la evaluación, los postulados de Benjamín Bloom (1956) que aparecen en su libro “Taxonomía de los Objetivos Educativos: Volumen I, el Dominio Cognitivo”, en donde se presentan las categorías cognitivas relacionadas con el aprendizaje educativo: información, comprensión, aplicación, análisis, síntesis y evaluación.

Estas categorías, en conjunto con la propuesta de Tyler¹³ de evaluar los objetivos educativos, forman parte de la cultura evaluativa occidental en el aula de clase y se han constituido en el eje principal de la evaluación escolar hasta nuestros días.

Dentro de este contexto, y gracias a la inyección económica para el desarrollo de la evaluación educativa, aparecen otros planteamientos que contribuyen a grandes avances en relación con el concepto mismo de evaluación y con los procedimientos para realizarlos. Una de estas miradas es la que plantea Robert Glaser (1963) en su documento “Tecnología Educativa y la Medición del Aprendizaje”. Plantea que la evaluación implica una comparación y que esta comparación puede darse de dos maneras diferentes:

- **Interindividual:** Los resultados de una persona (o una institución, o programa) se comparan con los resultados de las demás personas que presentan la evaluación con ella. Las inferencias se hacen a partir del porcentaje de personas que se supera.
- **Intraindividual:** Los resultados de una persona se comparan con lo evaluado. Las inferencias se hacen a partir de lo que la persona puede o no hacer.

El primer tipo de comparación es conocido como Evaluación con Referencia a la Norma y el segundo, como Evaluación con Referencia a Criterio. Desde ese momento hay una verdadera explosión de pruebas que evalúan con referencia a criterio, puesto que se supone, dan mejor información desde el punto de vista educativo. Sólo había un problema, los modelos matemáticos utilizados para el análisis de datos no permitían un manejo matemático efectivo.

13 Tyler es el primero en utilizar los términos “Evaluación Educativa”, en contraposición con los de Medición Educativa.



Durante muchos años se desarrollaron índices, con problemas ya sean conceptuales o matemáticos, para abordar la problemática planteada por este tipo de evaluación. Sólo a finales de la década del 70 se hace pública la Ítem Response Theory (Teoría Respuesta Pregunta) cuyos modelos matemáticos logit han dado buena cuenta de la Evaluación con Referencia a Criterio.

A finales de la década del 60, Michael Scriven (1967), en su artículo “La Metodología de la Evaluación”, establece la diferencia entre lo que es Evaluación Sumativa y Evaluación Formativa. Ante todo plantea que la meta del evaluador es emitir juicios bien informados y que lo esencial del proceso evaluativo es un juicio de valor. Para él, las dos funciones mencionadas se realizan para calcular el valor del objeto luego de estar completamente desarrollado (evaluación sumativa) y para ayudar a desarrollar los objetos (evaluación formativa).

Los anteriores son sólo dos casos que reflejan el impacto de las concepciones de los 60 en la evaluación educativa y en la educación misma, y que han guiado el desarrollo de la evaluación luego de su aparición, tanto que aún hoy se reflexionan sobre esos tópicos.

Esta discusión se cristaliza en diferentes sistemas de evaluación en diferentes niveles, desde el de aula de clase que adopta la evaluación a través de pruebas objetivas, pero sin la parafernalia técnica en la mayoría de los casos, hasta los sistemas de evaluación nacionales e internacionales.

2. Los últimos años de la evaluación

En 1989 Robert Glaser¹⁴ propone como conclusión de diversas tendencias de investigación en educación, la evaluación de competen-

14 Glaser, R. (1989). New Concepts of Achievement and Reasoning. En International Journal of Educational Research. Londres. Pergamon Pres.



cias, entendida ésta como la puesta en práctica en situaciones de la vida cotidiana, de lo aprendido en la escuela.

Esta concepción se deriva directamente de las tendencias de la psicología en relación con la cognición humana, especialmente a partir de los trabajos de Howard Gardner relacionados con la inteligencia del ser humano y supera las concepciones de la época (inicios de la década del 80) que relaciona la competencia con la eficiencia y efectividad en relación con los estándares de calidad establecidos¹⁵.

A partir de este momento y durante toda esta década, se profundiza en el concepto de evaluación educativa y su integración e interacción con todo el contexto educativo, y se involucran conceptos como los de evaluación de competencias, evaluación alternativa¹⁶, evaluación con portafolios, evaluación auténtica, evaluación de ejecuciones, todos relacionados con aspectos cualitativos de la evaluación.

El desarrollo de estas perspectivas implica la vinculación de los elementos cognitivos en la evaluación, tales como la resolución de problemas, la creatividad, el pensamiento crítico, entre otros, historia que aún se encuentra en pleno desarrollo.

A partir de ese momento se inicia un gran esfuerzo en torno a la evaluación educativa en el mundo ya que ésta se constituye en uno de los pilares que pueden transformar el proceso educativo al proporcionar información pertinente y válida de lo que ocurre en un contexto particular.

Es así como aparecen, tímidamente en un comienzo, los SISTEMAS INTERNACIONALES DE EVALUACIÓN y, también, los SISTEMAS NACIONALES DE EVALUACIÓN.

15 Measuring Competence: defining Performance and Mastery.

16 Worthen, B. (1993). Critical Issues that will Determine the Future of Alternative Assessment. En: Phi Delta Kappan.



3. Evaluaciones internacionales

3.1. Association for the Evaluation of Educational Achievement – IEA

En cuanto a las primeras evaluaciones internacionales se puede decir que ya en 1970, la International Association for the Evaluation of Educational Achievement (IEA), realiza la primera evaluación internacional de ciencias naturales (conocida como FISS). Esta evaluación aportó información importantísima en relación con el cumplimiento de las metas educativas en los países participantes, de tal manera que al retroalimentar al sector, se evidenciaron cambios que redundaron en el desarrollo de políticas particulares.

Esta misma organización realizó la segunda evaluación internacional en ciencias entre 1979 y 1988 (cuando presentó los resultados finales), conocida como SISS. Esta segunda evaluación se realizó sobre 57 tópicos de contenidos, para cada uno de los cuales se evaluaron 3 tipos de habilidades:

- Conocimiento.
- Comprensión de un principio.
- Aplicación de la información y de los principios a la resolución de problemas.

Las pruebas se aplicaron a 3 poblaciones diferentes: 10 años y 4º grado, 14 años y 9º grado y a los estudiantes de último año de la educación media. En esta evaluación, y por primera vez en el contexto de evaluaciones masivas, se utilizan cuestionarios que recogen datos de variables potencialmente asociadas al logro de los estudiantes. Estos cuestionarios tienen el propósito de obtener información del contexto en donde ocurre el proceso educativo para entender y determinar la influencia de estos factores en relación con la calidad, con la intención de generar políticas que mejoren la educación.

Entre 1993 y 1995, la IEA realizó la tercera evaluación internacional (más conocida como TIMSS) de ciencias naturales incluyendo el



área de matemáticas en esta ocasión. Por primera vez participa un país latinoamericano (Colombia) y se introduce el tema de la necesidad de incluir la evaluación como elemento fundamental en los países de la región como generador de políticas educativas.

Este estudio es de investigación colaborativa entre diferentes instituciones y tiene su centro en The Center for the Study of Testing, Evaluation and Educational Policy del Boston College. Dada la importancia de los resultados obtenidos y del reconocimiento de la evaluación como parte fundamental de la educación a partir del TIMSS, la IEA decide institucionalizar el nombre y la evaluación, y estandarizar algunos de sus procesos y la realización de esta evaluación. Con el tiempo hay una mayor cantidad de países que se involucran y participan en el TIMSS.

Desde 1990 la Asamblea General de la IEA decidió realizar evaluaciones cada 4 años. Se han realizado la de 1995, 1999 (TIMSS repetido), 2003, 2007 y se está terminando la de 2011.

3.2. Educational Testing Service (ETS) - IAEP

En 1988, el Educational Testing Service (ETS) inicia la primera Evaluación Internacional del Progreso Educativo en matemáticas y ciencias (IAEP I)^{17 18 19}. El enfoque de estas pruebas es esencialmente con referencia a criterio, identificando niveles de habilidad en la población evaluada, por ejemplo, en matemáticas y para la población de 13 años, estos niveles son:

- Sumas y restas simples.
- Uso de operaciones básicas para resolver problemas.
- Uso de habilidades matemáticas para resolver problemas de dos pasos.
- Comprensión de los conceptos de medición y geometría, y resolver problemas más complejos.

17 Es casi una ampliación National Assessment of Educational Progress que realiza el ETS en Estados Unidos. Inclusive se utilizaron preguntas del Banco de Ítems en la Evaluación Internacional.

18 Lapointe, A., Mead, N. y Phillips G. (1988) A World of Differences. ETS.

19 Bertrand, R. Dupuis, F. (1989). A World of Differences. Technical Report. Université Laval.



- Comprensión y aplicación de conceptos matemáticos más avanzados.

En esta primera evaluación participaron 12 poblaciones de 9 países y se utilizaron cuestionarios para recolectar información de factores que potencialmente se relacionan con el logro educativo. En 1994 se realiza la IAEP II con la participación de 24 poblaciones.

Es de resaltar que, por primera vez, se incluyen resultados por niveles de desempeño o logros en una evaluación de esta naturaleza. Hoy en día este tipo de resultados es práctica común ya que proporciona información muy útil para entender los procesos de aprendizaje.

Esta evaluación no se realizó de nuevo. El ETS ha apoyado las evaluaciones de la IEA y de la OECD.

3.3. Programa para la Evaluación Internacional de Alumnos – PISA

La organización para la Cooperación y el Desarrollo Económico (OECD) creó en 2000 el Programa para la Evaluación Internacional de Alumnos – PISA (por sus siglas en inglés), que tiene por objeto evaluar si los alumnos que se encuentran cerca de la finalización de la educación básica o media tienen los aprendizajes (conocimientos) y habilidades que les permitan la participación efectiva en la sociedad actual o del saber.

Las pruebas de PISA son aplicadas cada tres años. Examinan el rendimiento de alumnos de 15 años en áreas temáticas clave y estudian otro tipo de resultados educativos como la actitud y motivación por aprender de parte del estudiante, su visión de sí mismos y las estrategias de aprendizaje que utilizan.

La evaluación se realiza cada 3 años, en cada una de ellas se concentra en algún aspecto en particular:

- 2000 lectura.
- 2003 matemáticas.



- 2006 ciencias.
- 2009 lectura.
- 2012 matemáticas.
- 2015 ciencias.

3.4. Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación – LLECE

En la década de los 90, en Latinoamérica, se vio el surgimiento de sistemas nacionales de evaluación de la calidad de la educación, especialmente a raíz de la reunión sobre este tema realizada en Santiago de Chile en 1992 y patrocinada por el Banco Mundial. Muy pocos países tenían para la época alguna forma de evaluación de la calidad de la educación, los más avanzados eran Colombia (tiene el sistema más antiguo en la región, desde 1968), Chile, México y Brasil. Esta reunión permitió conocer diversas experiencias, sus logros y, particularmente, las dificultades que enfrentaron los diferentes sistemas.

A partir de entonces diferentes países inician la conformación de sus propios sistemas nacionales de evaluación de la calidad de la educación sin que exista una coordinación entre ellos que facilite el conocimiento de la calidad educativa en el contexto Latinoamericano. Por esto la UNESCO, gracias a la importancia que le confiere al desarrollo de sistemas de evaluación, propicia la creación del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), hecho que se concreta en 1994 en Ciudad de México. La coordinación la asume la Oficina Regional de Educación de dicha organización con base en Santiago de Chile. El Laboratorio propicia, entonces, la realización del I Estudio Comparativo de la Situación Educativa en Latinoamérica, cuyos primeros resultados se divulgaron en diciembre de 1998.

Trece países participan en este primer estudio: Argentina, Bolivia, Brasil, Colombia, Costa Rica, Cuba, Chile, Honduras, México, Paraguay, Perú, República Dominicana y Venezuela²⁰.

20 En el informe final no aparecen resultados ni de Costa Rica ni de Perú por diversos motivos.



A partir del año 2000, el LLECE trabaja en la idea de realizar un segundo estudio que denomina SERCE (Segundo Estudio Regional Comparativo y Explicativo), cuyos resultados se han venido divulgando desde 2009 y han tenido un gran impacto en los países participantes.

En la actualidad, se está desarrollando el TERCE o Tercer Estudio Regional Comparativo y Explicativo.

4. Pertinencia de participar en evaluaciones internacionales

Como se mencionó anteriormente, las evaluaciones internacionales como las de la IEA (TIMSS, ICCS, PIRLS), la OECD (PISA) y LLECE, se han dedicado a evaluar diversos aspectos de la educación y la cultura para determinar el grado en el cual los estudiantes se encuentran dentro de los parámetros establecidos e indagar sobre las posibles variables que intervienen en los resultados.

Cada una de ellas se realiza bajo la perspectiva de un marco teórico que explica y especifica los propósitos y objetivos de la evaluación. Algunos se concentran en la evaluación de disciplinas importantes para el contexto social y cultural del momento y otros en la evaluación de aspectos importantes para la contribución efectiva del estudiante a la sociedad. En algunos casos, se introducen elementos novedosos y nuevos a la evaluación como el uso de nuevas tecnologías o la introducción de nuevos formatos para la práctica evaluativa.

También se presentan diferencias en cuanto al grupo específico de estudiantes que se pretende evaluar ya que en algunos casos se concentran en estudiantes de 4º grado o en los de final de la educación media, pues se pretende explorar el desarrollo de habilidades importantes para momentos posteriores de esos estudiantes, como en su vida adulta.



Todos han introducido el uso de cuestionarios que buscan información del contexto del estudiante, de la familia y los procesos adelantados por sus maestros.

Es decir, la diversidad es mucha. No obstante, con un pequeño estudio es posible indagar sobre el marco conceptual que sustenta cada evaluación y, por lo tanto, es posible determinar la conveniencia o no de participar en uno de estos estudios.

Por supuesto que siempre habrá ventajas en la participación, no tanto por las posibilidades de entender los procesos educativos de un país en contextos más universales, sino que existe una verdadera posibilidad de aprender para desarrollar la técnica empleada en el contexto de la evaluación en el sistema nacional.

5. Una breve historia de los desarrollos técnicos

También existe una historia de los desarrollos técnicos y teóricos relacionados con la medición de atributos del ser humano. En términos generales, la psicometría ha abordado el problema desde dos perspectivas teóricas: la Teoría clásica de los test (mucha de su historia la acabamos de ver) y la Teoría Respuesta al Ítem (IRT). Esta última surge como consecuencia de los problemas que aún no puede resolver la Teoría Clásica de los Test (TCT) como son: el diseño de test, la identificación de preguntas sesgadas o con Funcionamiento Diferencial de Preguntas y el equating (comparabilidad) entre los puntajes de pruebas.

Como dicen Hmableton y Swaminathan (1985) la TCT no provee información acerca del lugar, en la escala de puntajes, de máxima discriminación de los ítems, tampoco ha tenido éxito en detectar el sesgo de las preguntas. Debido a ésta y otras razones, en los últimos tiempos se han desarrollado aproximaciones teóricas más apropiadas para solucionar los problemas en la medición de atributos del ser humano.



El propósito de cualquier teoría de la medición es “describir la forma en la cual pueden hacerse inferencias, a partir de las respuestas de una persona a unas preguntas, de características no observables de los examinados, o rasgos medidos por un test”²¹. Las propuestas desarrolladas en el presente se agrupan alrededor de la Teoría del Rasgo Latente o IRT, como se conoce hoy en día.

La historia de esta teoría y sus planteamientos no es tan reciente. Ya para 1936 se planteaban algunos conceptos como el de parámetro y se establecían relaciones con la TCT. Un empujón bastante fuerte a estos modelos lo dio Frederick Lord en 1952 en su tesis de doctorado donde describe el modelo de ojiva normal de dos parámetros, seguido por la propuesta de Birbaum en el sentido de sustituir los modelos logísticos por los de la ojiva normal con su correspondiente tratamiento estadístico.

En 1960 Georg Rasch, un matemático Danés, da otro giro a la problemática general de la medición educativa al plantear su modelo logístico de un parámetro o modelo simple, que aparece en el libro *Probabilistic models for some intelligence and Attainment test*. Su trabajo influye en diversas personas relacionadas con la psicometría quienes desarrollan y popularizan el modelo, entre quienes encontramos a Wright en Estados Unidos, Andrich en Australia, Andersen en Europa, Choppin en Inglaterra.

Durante la década de los 70, se lleva a cabo la gran divulgación de estos modelos. Su dificultad principal radica en la complejidad de los modelos matemáticos empleados y en la imposibilidad de desarrollarlos manualmente, a diferencia de los de la TCT. Es por esto, quizás, que deba esperarse a la aparición de los computadores personales para que en realidad se difundieran a muchas personas. Instituciones como el ETS (Educational Testing Service) contribuyen a la reflexión sobre el uso de esta teoría de forma amplia.

Es indudable la contribución que los manejos conceptuales y matemáticos de la IRT han tenido en la medición educativa, especialmente en la medición psicológica en general. La literatura disponible en la actualidad es bastante amplia y el software especializado

21 Hambleton y Swaminathan (1985). *Item Response Theory*. Kluwer. Bosaton.





para los procesos estadísticos es bastante preciso, de muy buena calidad y funcional. En los diversos estudios internacionales como el TIMMS o el IAEP, el Estudio de Cívica y Democracia de la IEA y en evaluaciones realizadas por diferentes países como el NAEP en Estados Unidos, las pruebas de logro cognitivo de Colombia, el SI-MECAL en Bolivia, el sistema de evaluación de Paraguay, las evaluaciones nacionales en Australia, las evaluaciones de Holanda, Dinamarca y muchas otras, se utiliza decididamente algún modelo de la IRT para el análisis de ítem, el análisis de prueba, la generación de escalas de calificación y los sistemas de resultados específicos, lo que ha permitido abordar situaciones de evaluación mucho más complejas que las que puede resolver la TCT, como la integración de procesos cualitativos a la medición.

Aún hay mucho que decir sobre las posibilidades de uso de la medición y evaluación educativas y sobre las perspectivas que se presentan hacia el futuro, pero esa historia aún se está escribiendo.





Capítulo II

Reflexiones sobre la evaluación

2.1 Reflexiones sobre evaluación de los aprendizajes

Juan Carlos Tedesco ²²

El ciclo de procesos de transformación educativa que comenzó en casi todos los países durante la década de los años 90 estuvo basado en la idea según la cual era prioritario modificar el diseño organizativo e institucional de los sistemas educativos. Uno de los argumentos sobre el cual se sostenía esta prioridad era el bajo nivel de responsabilidad por los resultados con los cuales operaban las administraciones educativas tradicionales. Según el enfoque dominante en esos años, para aumentar la responsabilidad por los resultados era preciso otorgar mayor autoridad a los actores del proceso pedagógico (y en ese sentido se ubicaban las estrategias de descentralización y de autonomía a los establecimientos escolares) y establecer mecanismos que permitieran medir sistemáticamente los logros de aprendizaje y difundir masivamente la información para provocar mayor y mejor demanda por parte de las familias, la opinión pública y el propio gobierno.

En el caso de América Latina, prácticamente todos los países crearon sistemas nacionales de evaluación de la calidad; se instalaron mecanismos regionales de medición de resultados y varios países se incorporaron a las mediciones internacionales. De esta forma, los sistemas educativos comenzaron a disponer de instrumentos técnicamente confiables para conocer sus resultados, especialmente aquellos referidos a las asignaturas básicas: lengua, mate-

²² Especialista en Educación y Ex Ministro de Educación de la Nación Argentina. Conversatorio sobre evaluación de aprendizajes realizado en Paraguay.



máticas, ciencias sociales y naturales. Una discusión importante acerca de los sistemas de evaluación se refiere al tipo de institución responsable de la medición. En un extremo, están los sistemas que dependen de los Ministerios de Educación. En el otro, están las agencias privadas independientes. Entre un extremo y otro se ubican diferentes fórmulas mixtas de organismos públicos con participación de sectores no gubernamentales (agencias de evaluación, institutos, etc.).

Después de más de una década de la instalación de estos instrumentos, parece llegado el momento de evaluar a los evaluadores. Las razones que justifican este ejercicio son muy diferentes. Por un lado, las mediciones indican que la calidad de la educación no está mejorando. En general, todas las evaluaciones indican que no hay cambios significativos en los resultados de aprendizaje de los alumnos y que esos resultados están directamente asociados al origen social y a las condiciones de vida de las familias. Estos resultados podrían ser vistos como un dato positivo si los consideramos en relación al crecimiento de la matrícula. Desde este punto de vista, podríamos afirmar que los sistemas educativos de la región mantuvieron sus resultados, aún a pesar de incorporar alumnos de sectores con mayores dificultades socioculturales. Sin embargo, tampoco aparecen cambios sensibles en aquellas zonas en las que el aumento de matrícula no fue significativo. En síntesis, lo cierto es que, a pesar de los enormes esfuerzos realizados por los procesos de reforma educativa y, en algunos casos, por la significativa inversión financiera realizada, los resultados de los procesos de evaluación no parecen registrar cambios en los logros de aprendizaje de los alumnos.

Este dato abre varios interrogantes: ¿Las reformas educativas no han modificado lo que sucede en las escuelas o los instrumentos no están midiendo correctamente los resultados? La instalación de estos mecanismos ¿ha provocado mayor responsabilidad por los resultados o ha servido, al contrario, para segmentar aun más la demanda educativa?

Un hecho es innegable: los sistemas de medición han convertido los resultados educativos en un motivo de debate público. El impac-



to que tiene la difusión masiva de los resultados de las pruebas nacionales o internacionales a través de los medios de comunicación de masas es muy significativo, especialmente si se lo analiza desde el punto de vista político. En este sentido, los instrumentos técnicos de medición de resultados han adquirido un significativo papel político y la comunicación se ha convertido en una de las principales dimensiones que debe manejar toda gestión de política pública en educación. En síntesis, ya no es posible pensar la problemática de la evaluación exclusivamente desde el punto de vista técnico-pedagógico, sino que se requiere otro tipo de mirada, mucho más amplia y compleja.

En esta ampliación de la mirada sobre los instrumentos de evaluación, debemos considerar aspectos internos y aspectos externos a las acciones educativas. Desde el nivel interno, es necesario incluir la relación de las políticas de evaluación con los actores del sistema. Más allá de las intenciones, parece evidente que los docentes percibieron estos dispositivos más como una presión o una amenaza que como un insumo para el mejoramiento de las estrategias de enseñanza y aprendizaje. En ese contexto, la información de los resultados no se transformó en un insumo para el mejoramiento de las políticas, los datos no se elaboraron en las escuelas y el impacto sobre las estrategias de enseñanza fue muy débil.

En relación a los actores externos al sistema, es preciso considerar todo lo relativo a la comunicación de los resultados. ¿A quién, cómo y cuándo se deben comunicar los resultados? ¿Qué puede y debe hacer cada actor con esta información? Estas preguntas aparecen en el momento en que transformamos a las evaluaciones de la calidad y sus resultados en un producto de la comunicación masiva, en el momento en que interactúan con el conjunto de la información transmitida por los medios masivos de comunicación. El modo de presentar la información, la propia información presentada, los interlocutores a los que se la presentamos, el hecho de que sea el propio gobierno y su ministerio de educación quienes la legitimen, obliga a otro nivel de análisis de la problemática. La comunicación masiva de esos resultados es uno más de los elementos que han transformado a las políticas de evaluación en un hecho político, además de una construcción técnico-pedagógica.



En este contexto, la participación en evaluaciones internacionales agrega algunas cuestiones específicas. Por un lado, la ventaja de permitir comparar la situación nacional con la de otros países, lo cual enriquece el diagnóstico. Pero, al mismo tiempo, reduce el grado de autonomía para manejar tanto los criterios de la evaluación como las formas de difusión de las informaciones.

Por último, también es preciso proponer una mirada que vincule la evaluación con los objetivos de las políticas sociales. ¿Evaluamos para mejorar la eficiencia de la inversión educativa?, ¿Para orientar la demanda? ¿Para compensar las diferencias sociales? ¿Para comprobar los niveles de segmentación social? Los resultados de las evaluaciones dependen de las decisiones que tomemos en relación a “para qué” evaluamos, y esas decisiones suponen definir el escenario social para el cual estamos trabajando. Detrás de esta mirada, ha crecido un debate ideológico y político. El “para qué” se ha vuelto una pregunta central en el análisis, porque define una idea sobre la sociedad que queremos construir. Modelos que ponen la idea de equidad en el centro del debate, otros que ponderan el concepto de competitividad, aquellos que se centran en la idea de eficiencia, por ejemplo, tendrán distintas respuestas a la hora de pensar las políticas de evaluación.

2.2 Reflexiones sobre la evaluación educativa en Paraguay

Carlos A. Pardo

La evaluación educativa cobra especial interés en el contexto paraguayo debido a la multiplicidad de procesos evaluativos realizados en poco tiempo, al interés por parte del Ministerio de Educación y Cultura, pero especialmente, porque los resultados de la evaluación nacional realizada por el SNEPE se encuentran por darse a conocer. Es decir, el contexto es propicio para aprovechar la información en beneficio de la calidad educativa, tema central en la actualidad.



En los últimos días se realizaron, con el auspicio del Ministro de Educación, algunos eventos que permiten reflexionar en torno a la educación, la calidad, la evaluación.

Pienso que algo importante que es necesario resaltar es la intención (y ojalá verdadero uso en el futuro) de utilizar la evaluación como fuente primordial, no sólo para la toma de decisiones sino para entender lo que ocurre en el proceso educativo. Este hecho implica que las decisiones que se van a tomar tendrán una base firme y fuerte ya que se preguntará a los resultados sobre lo que ellos quieren decir. No es como ocurría en otros contextos, es decir, que las decisiones ya estaban tomadas y, en ocasiones, no estaban en sintonía con la realidad. Ahora se tendrán bases fuertes y muy completas de lo que ocurre en educación y sobré qué aspectos se debe actuar para mejorar la calidad. Los resultados de SNEPE tienen todo el potencial para ser la base de estas decisiones y diseño de política educativa.

Por otro lado, y con igual importancia, se encuentra la alta calidad de los procesos de evaluación alcanzados por el Ministerio de Educación y Cultura a través del SNEPE. En todos los sectores y el ámbito social actual es vital la calidad de la información para el diseño de políticas y esto es igual en el sector educativo. Por lo general la información no ha sido buena, es decir, se basa en pruebas creadas rápido y sin cumplir los requerimientos mínimos para que las interpretaciones que derivan de sus resultados sean válidas. Pero ese no es el caso del SNEPE que ha incluido en la obtención de información procedimientos modernos y muy robustos, no sólo para cualificar la información sino para evitar falsas expectativas en relación con lo que ocurre en la educación del país. Adicionalmente, se han incluido los cuestionarios de factores asociados ya que no sólo es importante saber si se está bien o no, sino a qué se debe y qué factores influyen en la calidad de la educación.

Por último, pero no menos importante, es la participación de personas no sólo muy conocedoras del sector educativo del país sino muy interesadas en encontrar la clave que permita mejorar la calidad de la educación, ya que, parece unánime, todos pensamos que la educación es el gran motor que hace avanzar el desarrollo de un país en todos los aspectos, desde el social hasta el económico.



En relación con los resultados y las discusiones realizadas en los eventos mencionados queda la certeza que se mencionaron los aspectos importantes y pertinentes sobre los cuales trabajar para mejorar la calidad. Las preguntas se concentran en dos: qué es lo que los estudiantes paraguayos no aprenden y por qué no lo aprenden.

Entendiendo que el contexto social y educativo del Paraguay es único e incomparable, es claro que los estudiantes no aprenden los elementos generales de las disciplinas que les permitirían alcanzar niveles de conocimiento avanzado. Esto no quiere decir que no aprendan nada o que todos los estudiantes aprendan poco. Por el contrario, hay evidencia de altos logros en el desempeño de algunos estudiantes. La pregunta que surge es: qué se puede hacer para que este desempeño sea mucho más generalizado y lo alcancen muchos más estudiantes, la gran mayoría.

De ahí la importancia de la evaluación que permitirá conocer quiénes tienen estos logros y por qué los tienen, cuáles son las condiciones que los favorecen y qué elementos contextuales existen, de tal manera que se pueda generalizar.

Aunque algunas respuestas aparentemente ya se conocen, es necesario confirmar las hipótesis con los resultados de la evaluación. Debemos preparar las preguntas adecuadas y tener en mente las mejores estrategias para encontrar las soluciones.

Ahora no sólo debe existir esperanza de cambiar el futuro, se tienen las herramientas necesarias para hacerlo de verdad.



Capítulo III

Las evaluaciones educativas en Paraguay²³

En este apartado se presenta una síntesis de las principales evaluaciones realizadas a nivel nacional, así como las evaluaciones internacionales de las cuales Paraguay participa.

3.1 Sistema Nacional de Evaluación del Proceso Educativo - SNEPE

La evaluación de la calidad en educación se incorpora con fuerza en América Latina en la década de los 90 coincidiendo con las Reformas Educativas de los diferentes países y mediante la creación de los sistemas nacionales de evaluación de la calidad.

En Paraguay, al igual que otros países de Latinoamérica, el énfasis puesto en la necesidad de contar con sistemas nacionales de evaluación, como base para mejorar la calidad de la educación en condiciones de equidad, cobró fuerza en muy poco tiempo y aglutinó los consensos sociales y políticos indispensables para el desarrollo de los sistemas educativos en los países latinoamericanos. Es así, que en el año 1996 se crea el Sistema Nacional de Evaluación del Proceso Educativo – SNEPE.

Los antecedentes de su creación mencionan, entre otros temas, que los planes de mejoramiento de la calidad y la gestión educativa requieren de la generación de sistemas de información y evaluación como bases necesarias para la definición de políticas, la pla-

²³ MEC, Dirección General de Planificación Educativa.



nificación y la gestión. De ahí la necesidad de contar con un Sistema Nacional de Evaluación de la Calidad que produzca y difunda información relevante sobre el desempeño del sistema educativo, especialmente cuánto aprenden y qué niveles de aprendizajes logran los estudiantes durante su permanencia en el sistema educativo formal, y acerca de cuáles son los factores asociados a dichos aprendizajes.

Desde sus inicios el SNEPE ha evaluado a los estudiantes de diferentes niveles educativos con la aplicación de pruebas a una muestra representativa a nivel nacional, en los grados o cursos del final de cada ciclo educativo; 3º, 6º y 9º grados de la Educación Escolar Básica (EEB) y 3º curso de la Educación Media, si bien en el 2006 se aplicó a estudiantes del 2do. curso; también se han realizado mediciones a estudiantes del 3º año de la Formación Inicial de Docentes y en el preescolar de la Educación Inicial.

El SNEPE, desde su instalación, ha tenido limitaciones para institucionalizarse y siempre ha operado con recursos provenientes de fuentes de financiamiento externo, generalmente préstamos, lo cual ha imposibilitado garantizar la periodicidad de las evaluaciones. Otras de las limitaciones ha sido la falta de la comparabilidad de los resultados entre las diferentes aplicaciones.

En los 15 años de vigencia de este Sistema se han registrado transformaciones importantes en el SNEPE, la primera refiere a su objetivo, pasando de sólo ofrecer información con fines de diagnóstico a medir, actualmente, los niveles de desempeño de los estudiantes así como los factores asociados al mismo y proveer información permanente, confiable y oportuna que permita incorporar acciones de política a fin de mejorar la calidad de la educación paraguaya. De igual forma ha incorporado nuevos modelos para el diseño, la construcción y evaluación de pruebas, tal como la Teoría de Respuesta al Ítems (TRI), cuya aplicación complementa a la Teoría Clásica del Test (TCT), utilizada por el sistema desde sus inicios.



3.1.1. Etapas para la aplicación

El proceso de aplicación reúne toda la rigurosidad técnica, científica y metodológica de las evaluaciones internacionales utilizadas para la medición de la calidad de los aprendizajes; las pruebas comprenden varias etapas y en promedio, todo el operativo hasta contar con los resultados de la evaluación, tiene una duración aproximada de tres años. A fin de ofrecer una mejor ilustración se describen a continuación las etapas desarrolladas para la aplicación.

- **Construcción y características de las pruebas:** los instrumentos tienen como objetivo básico dimensionar las competencias adquiridas por los estudiantes en el espacio escolar para el grado/curso conforme a lo establecido en el currículo en cada una de las áreas del conocimiento a evaluar. Así mismo, pretende valorar la competencia de los estudiantes en problemáticas que exigen el diálogo de algunos saberes propios de las disciplinas.

Desde la perspectiva de la medición, el SNEPE considera, la competencia como “**Un saber hacer en contexto**, es decir, el conjunto de acciones que un estudiante realiza en un contexto particular y que cumple con las exigencias específicas del mismo”. En la prueba, las competencias se circunscriben a las acciones de los procesos cognitivos que demandan el nivel de desempeño del estudiante, en cuanto a “conocimiento, comprensión y aplicación”.

En la dimensión de los procesos cognitivos se han tomado, tradicionalmente, tres de los seis planteamientos taxonómicos de Benjamín Bloom; se construyen ítems a efecto de medir el desempeño que hace referencia al dominio cognitivo. El conocimiento se define como la recordación de lo aprendido previamente y representa el nivel más bajo de los desempeños del nivel cognitivo. En este estadio, el estudiante recuerda y reconoce información e ideas, además de principios en la misma forma en que los aprendió.



La comprensión se concibe como la habilidad de aprehender mediante el esclarecimiento e interpretación de la información en base al conocimiento previo. La aplicación hace referencia a la capacidad de utilizar el material aprendido en situaciones concretas y nuevas. Esto requiere un nivel de entendimiento mayor que el de la comprensión, para ello, el estudiante selecciona, transfiere y utiliza datos y principios para realizar una tarea o solucionar un problema.

Las pruebas de rendimiento académico son elaboradas sobre la base de las capacidades básicas establecidas en el Diseño Curricular vigente para el nivel y las áreas de estudio. A partir del análisis de este documento se elabora una tabla de especificaciones para cada una de las áreas a evaluar. La tabla de especificaciones constituye la estructura de la prueba que representa la integración de las distintas dimensiones que se desean medir.

• **Estructura de las pruebas:** Las pruebas se estructuran por bloques balanceados, presentes dos veces en diferentes cuadernillos de manera a ampliar los temas sin extender mucho la prueba, evaluando así con mayor detalle los ejes temáticos, las competencias y los procesos cognitivos. Al presentar en cada bloque un número reducido de ítems, disminuye la cantidad de ítems que cada estudiante debe responder, lo que permite aumentar el número de pruebas (cuadernillos construidos con dos bloques de un número determinado de ítems en cada uno) a responder en el grupo curso. De esta forma se hace una evaluación más eficiente.

Los instrumentos aplicados son validados a través de juicio de expertos y de una aplicación experimental. Estos resultados son analizados para determinar la validez, la confiabilidad de las pruebas y la calidad técnica de los ítems.

La evaluación es aplicada a una muestra representativa a nivel nacional y para estratos previamente definidos. La muestra es seleccionado en forma aleatoria, el marco muestral



utilizado es la base de instituciones educativas del Sistema de Información de Estadística Continua (SIEC).

• **Instrumentos Aplicados:** Los instrumentos son aplicados en las áreas de Matemática y Comunicación/Lengua y Literatura Castellana, además se aplican cuestionarios de factores asociales al aprendizaje. Estos cuestionarios están dirigidos a estudiantes, docentes, directores y padres de las instituciones participantes de la evaluación.

El objetivo de aplicar los cuestionarios es identificar los factores que mejor explican las varianzas entre los logros alcanzados en las pruebas de rendimiento académico.

Las preguntas que conforman los cuestionarios se refieren a factores que reúnen un conjunto de variables que podrían incidir directa o indirectamente en los procesos de aprendizaje. Estos factores hacen referencia a condiciones internas y externas de la escuela que actúan sobre los actores del proceso educativo y, en consecuencia, podrían influir en el desempeño de los estudiantes.

• **Procesos de validación de los instrumentos**²⁴ : La validez es un concepto técnico que se refiere al grado en el cual el proceso de evaluación permite una buena interpretación de los resultados, es decir, informa sobre la calidad de todo el proceso de evaluación. Así, el proceso de evaluación involucra en general tres aspectos:

- a. Diseño y elaboración de los instrumentos.
- b. Aplicación de los instrumentos.
- c. Procesamiento de datos y obtención de resultados.

En el diseño y elaboración de los instrumentos, se controla la validez a través del juicio de expertos quienes revisan en detalle el planteamiento de los ítems y verifican que las claves sean las correctas.

24 Pardo, Carlos. Especialista en evaluación educativa.



La aplicación piloto de los instrumentos permite recoger información para construir instrumentos libres de errores de medida para la aplicación definitiva con propósitos de difusión de resultados, para ello, se hacen controles a partir de las respuestas de los estudiantes a cada pregunta planteada.

Durante el procesamiento de datos se realiza un análisis de ítem tanto en el pilotaje como en la aplicación definitiva utilizando modelos de análisis estadísticos rigurosos que permitan la detección de falencias en los ítems. En el presente caso se ha utilizado el Modelo de Rasch de la Teoría Psicométrica de Respuesta al Ítem (TRI) que produce indicadores de alta calidad para detectar los ítems que no cumplen con criterios preestablecidos.

La producción de resultados, igualmente se realiza con modelos estadísticos tanto de la TRI (el modelo de Rasch) como de la Teoría Clásica de las Pruebas (los puntajes en escala).

Adicionalmente, cada procesamiento de datos tanto para el análisis de ítem como para la producción de resultados se verifica con un segundo procesamiento de datos realizado de manera independiente.

Para todos los aspectos de la validez mencionados se cumplen y se tienen en cuenta estándares internacionales de calidad de evaluaciones educativas.

3.2 Análisis, difusión y uso de los resultados

Si bien se han dado avances importantes en cuanto al diseño, construcción y validación de los instrumentos, en materia de análisis, difusión y uso de información se registran aún desafíos importantes.



Históricamente todos los informes de resultados se han limitado a dar cuenta sobre el porcentaje del logro de aprendizajes sin indagar sobre las probables causas de los bajos niveles de aprendizaje de los estudiantes, tampoco se ha logrado establecer la vinculación de los resultados de la evaluación con investigaciones educativas.

Otra de las debilidades detectadas en este proceso es que los resultados no han llegado a nivel de las instituciones educativas y de los docentes, sumado a ello, tampoco han sido utilizados para generar políticas públicas que contribuyan a mejorar la calidad de los aprendizajes.

Ante esta situación, el MEC ha iniciado un debate sobre la importancia de la evaluación educativa como herramienta que aporte insumos para mejorar los aprendizajes de los estudiantes. En este marco se ha convocado a especialistas en el área de educación e investigadores a quienes se les han presentado los resultados y se ha puesto a disposición la base de datos.

Así mismo, con la aplicación realizada en el 2010 por primera vez el MEC en una evaluación nacional en la Educación Escolar Básica, realiza el análisis de los resultados por niveles de desempeño y se enfatiza en la necesidad de profundizar el análisis de los factores asociables al aprendizaje, desde distintas perspectivas. Además se buscará establecer un perfil de los estudiantes para cada nivel de desempeño de tal forma que los resultados sirvan de base para trabajar con los docentes en base a los requerimientos de los estudiantes.

3.3 Aplicación 2010: los primeros resultados

En el año 2010 se han aplicado pruebas y cuestionarios a 59.556 estudiantes del 3º, 6º y 9º grados de la EEB, correspondientes a 900 instituciones educativas. Así mismo fueron aplicados cuestionarios a 51.263 padres/madres, 3.457 docentes y directores de las



900 instituciones participantes de la evaluación. Algunos de los resultados se presentan a continuación.

Los niveles de desempeño son una forma de resultados con referencia al criterio, es decir, informan lo que un estudiante es capaz o no de hacer en una disciplina o área académica particular, de acuerdo con su nivel. Debido a lo anterior, son muy informativos y de fácil uso por parte de los docentes para detectar los procesos cognitivos y el área disciplinar en que los estudiantes presentan falencias de tal manera, que puedan concentrarse todos los esfuerzos de enseñanza en estudiantes específicamente identificados.

Los niveles de desempeño tienen la característica de ser crecientes e inclusivos, es decir, a mayor nivel es mayor la capacidad de un estudiante en un área disciplinar particular y, si un estudiante se encuentra en un nivel alto, quiere decir que superó con éxito los niveles anteriores. No obstante, su desventaja es que no informa sobre los temas o tópicos en los que una persona tiene falencias y para ello, es necesario obtener resultados complementarios. Cabe señalar que los mismos están siendo procesados actualmente y serán complementados con un análisis de factores asociados al aprendizaje.

En general, los primeros resultados dan cuenta que tanto en matemáticas como en comunicación no se han dado variaciones importantes en la calidad del aprendizaje de los estudiantes que, en mayor medida, se ubican en los niveles I y II lo que revela bajos niveles de aprendizajes.

Así mismo, es importante destacar que de los tres grados evaluados, los estudiantes del 3º son los que lograron ubicarse, en mayor porcentaje, en el nivel IV; más del 8% en matemáticas y 7.3% en comunicación, atendiendo que este es el nivel de mayor dificultad, es decir, se ubican aquí aquellos estudiantes con mayor habilidad. En contrapartida en este mismo grado, un 17% se ubica por debajo del nivel I, que son estudiantes que no logran siquiera responder correctamente los ítems del nivel de menor dificultad. Los resultados en matemáticas evidencian además que a medida que aumenta el grado disminuye el porcentaje de estudiantes que se ubican



por debajo del nivel I, es así que más del 17% de los estudiantes del tercer grado se encuentran por debajo del nivel I, reduciéndose prácticamente a la mitad en el noveno grado.

Un aspecto importante a destacar es que las pruebas de comunicación en el sexto grado fueron aplicadas en guaraní y en castellano, es decir, los alumnos completaron la misma cantidad de ítems en ambos idiomas y los resultados a escala nacional evidencian tendencias similares; a priori, pareciera que no se aprecian diferencias significativas en el nivel de desempeño de los estudiantes asociadas al idioma, sin embargo, se deberá profundizar en el análisis de estos resultados, incorporando otras dimensiones relacionados al contexto como ser el área de residencia, el idioma hablado en el hogar, el idioma de enseñanza, entre otras.

Cuadro 1: Matemática resultados por grado, según niveles de desempeño de los estudiantes (en porcentaje)

NIVELES	3° grado	6° grado	9° grado
Por debajo del I	17.2	10.2	7.9
I	35.4	30.4	28.2
II	26.2	45.0	47.3
III	12.8	11.5	14.5
IV	8.5	3.0	2.0

Fuente: MEC, DGPE, SNEPE 2010.

Cuadro 2: Comunicación resultados por grado, según niveles de desempeño de los estudiantes (en porcentaje)

NIVELES	3° grado	6° grado		9° grado
		Castellano	Guaraní	
Por debajo del I	17.0	11.8	9.1	10.5
I	23.4	21.9	25.5	27.0
II	32.0	44.2	44.7	42.6
III	20.3	18.7	17.6	17.4
IV	7.3	3.4	3.2	2.5

Fuente: MEC, DGPE, SNEPE 2010.



Estos resultados interpelan y plantean una serie de interrogantes:

- ¿Por qué los bajos niveles de desempeño de los estudiantes?
- ¿Cuáles son los factores que explican estos resultados?
- ¿Qué estamos evaluando y cómo estamos evaluando?
- ¿Qué estamos midiendo y qué estamos enseñando?
- ¿Qué estamos haciendo en la formación docente?
- ¿Qué hacer para cambiar estos resultados?

Estas y otras preguntas deberán responderse para lograr el mejoramiento de la calidad de los aprendizajes de los estudiantes.



Capítulo IV

Innovación en la Evaluación Educativa: evaluaciones mediante el uso de celulares ²⁵

El MEC, en el marco de la implementación del Plan Nacional de Educación 2024 que tiene como uno de sus ejes estratégicos el mejoramiento de la calidad de la educación, propone el uso de los celulares como una metodología innovadora para recabar información pertinente y oportuna, sin descuidar los procesos técnico-científicos y las actividades propias de la evaluación. Esta metodología no pretende reemplazar a las evaluaciones que se vienen desarrollando sino más bien, la información recogida a través de este mecanismo es utilizada para complementar las generadas por el SNEPE.

La evaluación por celular no es un sistema de calificación de alumnos, docentes o instituciones educativas. Los directores de las instituciones educativas y los estudiantes participan en forma voluntaria; los estudiantes que deciden participar de la misma no requieren contar con un celular propio, por otro lado, es importante señalar que esta evaluación no implica costos para los directores ni para los estudiantes.

El objetivo de esta aplicación es contar con información pertinente y oportuna sobre el aprendizaje de los estudiantes en temas específicos en las áreas de Lengua y Literatura Castellana y Matemática, de manera a realizar acciones que contribuyan al mejoramiento de los aprendizajes.

25 MEC, Dirección General de Planificación Educativa



Esta implementación piloto, en una primera etapa aborda temas básicos del currículo en el área de Lengua y Literatura Castellana y Matemática, suministrando información sobre el logro de algunas capacidades y/o competencias adquiridas. Gradualmente, se irá ampliando el alcance de las evaluaciones.

Además se evalúa, entre otros aspectos, la receptividad por parte de los estudiantes y docentes de este nuevo mecanismo de evaluación, la tecnología utilizada como plataforma para los reportes de los datos, así como la validez de los datos recabados.

La experiencia es pilotada en una muestra intencionada de instituciones educativas de gestión oficial que ofrecen Educación Media. Cabe señalar que la misma no constituye un sistema de calificación de alumnos, docentes o institución educativa.

Del piloto participan los estudiantes del primero, segundo y tercer año de la Educación Media que asisten a instituciones educativas de gestión oficial y que forman parte de la muestra. Los directores de las instituciones educativas orientan a sus estudiantes en la metodología de uso de los celulares para este fin.

Para la aplicación fueron capacitados los directores de las instituciones educativas en el uso de la tecnología, de manera que puedan guiar a los estudiantes para que los mismos puedan responder los ítems.

Las pruebas son diseñadas por especialistas del área correspondiente cuidando el rigor técnico establecido por el SNEPE, las mismas incluyen ítems de selección múltiple sobre una muestra de contenidos específicos y básicos establecidos en el currículo nacional vigente.

A fin de disminuir los sesgos que pudieran producirse al no tener aplicadores externos se ha calculado el tiempo promedio de duración de cada ítem, con esto se limita la posibilidad del estudiante de consultar materiales, docente y/o compañeros.

Las pruebas se aplican en tres etapas:



- **1ra. Etapa:** al inicio del año lectivo, con el propósito de identificar las condiciones iniciales en que se encuentran los estudiantes para fortalecer el aprendizaje inicial (etapa finalizada).
- **2da. Etapa:** a mediados del año, con el propósito de monitorear el avance del aprendizaje con relación a los temas específicos del currículo para la retroalimentación oportuna (en proceso).
- **3ra. Etapa:** al final del año lectivo, con el propósito de verificar el aprendizaje de los estudiantes en temas específicos del currículo correspondiente al curso.

La primera etapa fue aplicada en abril y la segunda en setiembre, la aplicación de la tercera etapa está prevista para noviembre del 2011.

Cabe señalar que esta actividad no ha originado ningún tipo de gastos al MEC atendiendo que la misma fue desarrollada en el marco de la responsabilidad social de las empresas, a través de un convenio con la telefonía móvil Tigo.

4.1. ¿Cómo opera el sistema?

Los directores de las instituciones educativas que optaron participar de la evaluación fueron capacitados por técnicos de la Dirección de Evaluación de la Calidad Educativa (DECE) y por los responsables del desarrollo del sistema utilizado para la evaluación.

El director de la institución educativa es el responsable de habilitar las pruebas a los estudiantes a través de un código aleatorio facilitado por el sistema informático utilizado para la evaluación, este código es único y está vinculado a la institución educativa, el curso y el área que será evaluada. Cabe señalar que a fin de garantizar la validez de los resultados obtenidos el código tiene una vigencia



determinada, en promedio se estimó un tiempo de exposición de tres minutos por cada ítem de la prueba.

Una vez que el sistema provee el código al director, este facilita dicho código a los estudiantes quienes llaman al menú estudiante, el sistema solicita se ingrese el código en ese momento y los estudiantes incorporan el código proveído por el director, así empiezan a recibir las pruebas en forma de mensaje y disponen de un tiempo determinado para responder cada ítem. El sistema no permite realizar dobles marcaciones en un mismo ítem, una vez que el estudiante marca una opción, automáticamente pasa al siguiente ítem hasta culminar la prueba. Es importante señalar que una vez terminada la prueba el alumno ya no puede acceder a la misma.

Los resultados se van teniendo en línea, entre otros temas, sobre la cantidad de instituciones que habilitan las pruebas, sobre cuántos alumnos están participando de la evaluación y el porcentaje de respuestas correctas por cada curso.

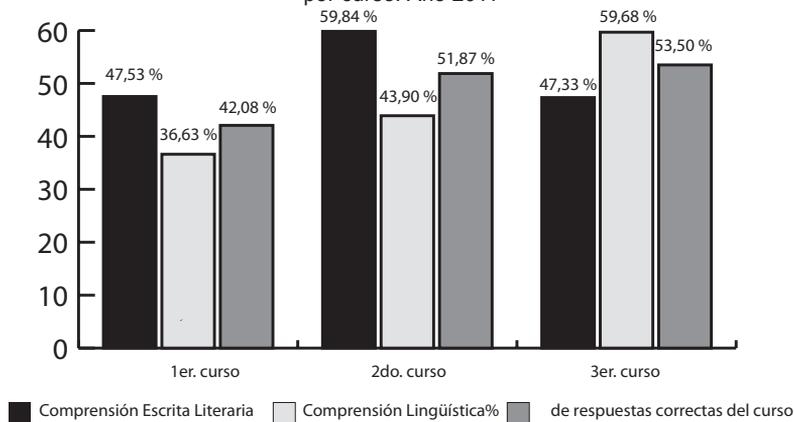
La base de datos resultante es compatible con cualquier sistema de procesamiento de datos lo cual permite establecer diferentes cruces de variables y análisis de los resultados.

4.2. Primera aplicación: principales resultados

Si bien este primer año de aplicación fue con la intención de pilotar la metodología, la misma aporta informaciones importantes para las áreas evaluadas. En Lengua y Literatura Castellana se evaluaron comprensión escrita y literaria y la comprensión lingüística, la comparación entre cursos permite apreciar, por un lado, que los del 1° curso son los que registraron el menor porcentaje de respuestas correctas, y por otro lado, que tanto en el 1° curso (36,63%) como en el 2do. (43,90%) la comprensión lingüística requiere ser fortalecida, en contrapartida, en el 3° curso el área más deficitaria es la comprensión escrita y literaria.



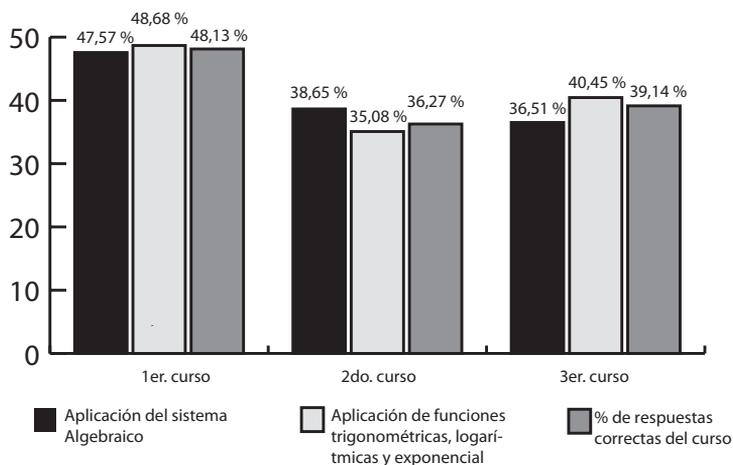
Gráfico 1: Lengua y Literatura Castellana. Porcentaje de respuestas correctas por curso. Año 2011



Fuente: MEC, DGPE, Evaluación por celular.

En matemática, a diferencia de los resultados de las pruebas de comunicación, es en el 1º curso donde se registran los niveles más altos de porcentajes de respuestas correctas, sin embargo, el mismo es inferior a 50%. Resultan preocupantes los resultados obtenidos. Por otra parte en los cursos 2do. y 3ro. el porcentaje es inferior a 40%, esto lleva a concluir en la necesidad de reforzar la aplicación de sistemas algebraicos, así como la aplicación de funciones trigonométricas, logarítmicas y exponenciales.

Gráfico 2: Matemática. Porcentaje de respuestas correctas por curso. Año 2011



Fuente: MEC, DGPE, Evaluación por celular.



Los resultados fueron entregados a las instituciones educativas participantes de la prueba, en porcentaje promedio de respuestas correctas por curso y materia, según las respuestas dadas por los estudiantes en forma general y por competencias según los ejes temáticos correspondientes a cada curso y materia evaluada.

4.3. Conclusiones sobre la primera experiencia

Si bien la muestra estuvo compuesta por un total de 300 instituciones educativas, participaron 185 instituciones del operativo experimental, el resto de las instituciones no pudo participar por problemas de señal de la telefonía que apoyó este emprendimiento.

En total 17.000 estudiantes participaron de la prueba, de los cuales 8.416 estudiantes respondieron la totalidad de las preguntas de Lengua Castellana y Literatura y 13.746 estudiantes respondieron la totalidad de las preguntas de Matemática.

La primera aplicación permite comprobar: i) que el uso de los celulares es una metodología válida para recabar información pertinente y oportuna sobre el aprendizaje de los estudiantes, ii) alta receptividad por parte de los estudiantes y directores de este nuevo mecanismo de evaluación, y iii) la validez de la tecnología utilizada como plataforma para los reportes de los datos.

Por último, el uso de los aparatos celulares aporta ventajas adicionales en cuanto a la utilización de los recursos materiales, humanos y de logística, teniendo en cuenta la inversión que insumen los sistemas nacionales de evaluación. Además, por su bajo costo y rapidez en dar la información, ofrece la posibilidad de realizar varias aplicaciones en el año y retroalimentar los aprendizajes.



Capítulo V

Evaluaciones internacionales de las que participa Paraguay²⁶

5.1. Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE – UNESCO)

El primer estudio se realizó en 1997 (PERCE), el segundo se concreta en octubre del 2006 en escuelas de la Capital y de los 17 departamentos del país (SERCE). El tercer estudio (TERCE), está en etapa de preparación y los ítems serán pilotados en el año 2012. La aplicación principal está prevista para el 2013.

Sus objetivos son generar conocimiento sobre los resultados de aprendizaje de los estudiantes de educación primaria en las áreas de lenguaje, matemática y ciencias, así como de los factores asociados que explican las diferencias de rendimiento entre los estudiantes.

Busca mejorar la calidad y la equidad de la educación ofreciendo información relevante para las políticas educativas y las prácticas en las escuelas y las aulas. Es una prueba estandarizada para América Latina y el Caribe.

El SERCE involucró a un total de 10.283 estudiantes de los grados 3° y 6° de la EEB, correspondiente a 204 instituciones educativas, docentes, directores y padres/madres de los grados evaluados.

²⁶ MEC, Dirección General de Planificación Educativa.



Los principales resultados del SERCE ubican a nuestros estudiantes en los niveles más bajos de desempeño tanto en matemáticas como en comunicación.

5.2. Estudio Internacional de Educación Cívica y Ciudadana (International Civic and Citizenship Education Study – ICCS)

Esta evaluación es llevada adelante por la Asociación Internacional para la Evaluación de Logros Educativos – IEA (Alemania, Hamburgo) y a nivel regional estuvo coordinada por el Centro Regional para el Fomento del Libro en América Latina y el Caribe (CERLALC) y el Sistema Regional de Evaluación y Desarrollo de Competencias Ciudadanas (SREDECC).

Por primera vez en el año 2009 Paraguay participó en un estudio internacional sobre ciudadanía, conjuntamente con países europeos, asiáticos y latinoamericanos.

El objetivo principal del estudio es investigar cómo los jóvenes están preparados para asumir su rol de ciudadanos, en materia de conocimientos, competencias y actitudes.

Paraguay hace parte del conjunto de los treinta y ocho países que participaron del Estudio Internacional sobre Educación Cívica y Ciudadana, grupo que cuenta con seis países latinoamericanos: Chile, Colombia, Guatemala, México, Paraguay y República Dominicana.

El estudio involucró a un total de 140.000 estudiantes, 62.000 docentes y 5.300 instituciones educativas. En Paraguay participaron alumnas y alumnos del noveno grado de la Educación Escolar Básica, cuyas edades oscilan entre 13,5 y 17 años, docentes del octavo grado que enseñan en la institución educativa desde el inicio del año escolar y los directores de las instituciones educativas incluidas en el estudio. La muestra fue seleccionada por el Centro de



Procesamientos de Datos de la IEA, DPC y estuvo compuesta por 3.399 estudiantes del noveno grado de la Educación Escolar Básica (Tercer Ciclo), 1.176 profesores del octavo grado y directores de 149 instituciones educativas de gestión oficial, privada y privada subvencionada.

La prueba del ICCS abarcó cuatro ejes temáticos: sociedad y sistemas cívicos, principios cívicos, participación cívica e identidades cívicas. Recolectó y analizó como variables adicionales a los resultados, información sobre las actividades, disposiciones y actitudes de los estudiantes relacionados con la Educación Cívica y Ciudadana.

5.2.1. Principales resultados

Los resultados obtenidos por los estudiantes de nuestro país en el Estudio Internacional sobre Educación Cívica y Ciudadana generan desafíos importantes para la educación paraguaya.

El puntaje promedio de conocimiento en Cívica y Ciudadanía, obtenido por los estudiantes de nuestro país –424–, es inferior al promedio internacional ICCS –500–, y nos sitúa en el penúltimo lugar entre los treinta y seis países cuyas muestras fueron representativas. Por su parte, el puntaje promedio de las mujeres –438– es significativamente superior al de los varones –408–, con una diferencia de 29 puntos a favor de las mujeres.

En cuanto al tipo de gestión de la institución educativa, el puntaje promedio de los estudiantes que asisten a instituciones educativas de gestión privada supera a las de gestión oficial; este mismo patrón de comportamiento se observa a escala regional e internacional. Es importante considerar en este punto la necesidad de profundizar en el análisis de los resultados de los factores asociados atendiendo que el tipo de gestión de las instituciones puede estar solapando otros factores que explican en mayor medida estos resultados.

El análisis de los resultados por niveles de desempeño ubica a la mayor parte de los estudiantes de nuestro país en los niveles más bajos de desempeño. A escala nacional, el 38% de los estudiantes



se ubican por debajo del nivel 1, esto indica que carecen de las competencias mínimas en Cívica y Ciudadanía; sólo el 7% de los estudiantes alcanzaron el nivel 3, es decir el nivel de desempeño más alto. El desempeño del 46% de los estudiantes que asisten a instituciones educativas de gestión oficial es inferior al nivel 1. Este comportamiento se registra tanto en las pruebas administradas a nivel nacional a través del SNEPE, así como en las administradas por el LLECE.

Si bien los factores asociables al desempeño de los estudiantes fueron considerados en forma separada, existe una asociación positiva entre el nivel educativo de los padres, la categoría ocupacional de los mismos y la presencia de libros en el hogar y los resultados obtenidos por los estudiantes en Educación Cívica y Ciudadana.

Las opiniones y actitudes de los estudiantes sobre la sociedad, sistemas y principios cívicos indican que la confianza hacia las instituciones públicas, por parte de los estudiantes de nuestro país, es inferior al promedio regional. Las instituciones que les generan mayor grado de confianza son las escuelas y los medios de comunicación; en cambio, los tribunales de justicia, la policía y los partidos políticos generan menos confianza a nuestros estudiantes.

En lo que respecta a la igualdad de derechos, los estudiantes de nuestro país demostraron una actitud positiva hacia la igualdad de género, los inmigrantes y grupos étnicos. El puntaje promedio de los estudiantes de nuestro país, con respecto a igualdad de género y valores democráticos es superior al promedio regional.

La participación cívica de los estudiantes de nuestro país supera el promedio regional en las actividades desarrolladas dentro y fuera de la escuela, así como en futuras protestas ilegales y discusión sobre temas políticos y sociales, caso contrario se da con la participación en futuras protestas legales y expectativa de participación política informal.

La percepción de los estudiantes de nuestro país sobre la importancia de la ciudadanía como movimiento social, ejercida por medio de la participación en protestas pacíficas, actividades que puedan



beneficiar a la gente o que promuevan los derechos humanos y protejan el medioambiente, se posiciona en la media regional.

Los estudiantes opinan que los docentes generan un clima escolar favorable para el aprendizaje, la mayoría de ellos expresaron que son tratados en forma justa, escuchados y ayudados por sus docentes, y además perciben el interés de los docentes en el bienestar de los estudiantes. También manifiestan que rara vez los docentes incentivan a los estudiantes a expresar su opinión, a discutir temas o presentar diferentes puntos de vista, durante las clases.

Según la opinión de los docentes, la violencia escolar, el acoso sexual, el vandalismo y la drogadicción afectan a un grupo minoritario de los estudiantes de nuestro país, el ausentismo representa el principal problema. En general, los docentes coinciden en que los estudiantes demuestran una actitud positiva hacia la escuela y que se sienten parte de la comunidad escolar.

Los docentes de nuestro país opinan que los objetivos más importantes de la Educación Cívica y Ciudadana son la promoción del conocimiento de los derechos y responsabilidades de los ciudadanos, el respeto y protección al medio ambiente y la promoción del pensamiento crítico e independiente de los estudiantes. Asimismo, consideran que los principales responsables de la Educación Cívica y Ciudadana deberían ser los directores de la institución educativa, la escuela en su totalidad y los docentes de esta asignatura específica.

La escala de confianza que los docentes de nuestro país tienen en sí mismos para enseñar Educación Cívica corresponde a la media regional. La economía y negocios e instituciones jurídicas y tribunales son los temas en los que se sienten con menor confianza.

Sobre los métodos de enseñanza, enfoques pedagógicos y resultados de la evaluación para fortalecer el aprendizaje de los estudiantes, el puntaje promedio de los docentes de nuestro país se posiciona por debajo de la media regional.



Los docentes manifiestan que, para mejorar la Educación Cívica y Ciudadana, es fundamental disponer de libros de texto y capacitar a los docentes en métodos pedagógicos y en conocimiento de la materia.

Con base a la opinión y percepción de los directores, la autonomía de la escuela se ubica en la media regional, mientras que la participación de los docentes y de los padres en la gestión escolar está por debajo de la media regional. Con relación a la disponibilidad de recursos en la comunidad local, el promedio de nuestro país está por debajo del promedio regional e internacional.

El Módulo Latinoamericano muestra que un alto porcentaje de estudiantes de nuestro país opina que la concentración de poder en una persona garantiza el orden, y que las dictaduras se justifican cuando ofrecen orden, seguridad y beneficios económicos; esta tendencia también se presenta a escala regional. Sin embargo, la mayoría de los estudiantes de nuestro país no está de acuerdo con que un gobierno cierre los medios de comunicación que lo critiquen o que disuelva el Congreso Nacional.

En relación al ejercicio del servicio público y el Gobierno, más del 70% de los estudiantes expresaron su desacuerdo con que un funcionario público acepte sobornos y utilice los recursos de la institución para su beneficio personal, y más del 80% está totalmente de acuerdo con que el Gobierno debe rendir cuentas a la sociedad y manejar con transparencia los recursos; sin embargo, resulta preocupante que entre el 30% y 70% de los estudiantes de nuestro país opina que se puede desobedecer la ley bajo ciertas circunstancias.

Los temas que se abordan con mayor frecuencia en las escuelas son las consecuencias del consumo de drogas ilegales, la discriminación hacia las personas con orientación sexual distinta y las ventajas o desventajas del funcionamiento de organizaciones no gubernamentales en un país democrático.

Finalmente, a pesar de los resultados, no se debe perder de vista que la escuela es una de las instituciones en la que más confían los estudiantes, esto nos brinda la oportunidad de cambiar la historia,



pero este cambio no podrá lograrse sólo a través de la escuela, sino en articulación con otros sectores y otros organismos, pues la formación cívica y ciudadana nos corresponde a todos.

5.3. Programa de Evaluación y Monitoreo de la Alfabetización

El Instituto de Estadística de la UNESCO (UIS) inició el Programa de Evaluación y Monitoreo de la Alfabetización (LAMP) en 2003 mediante un trabajo conjunto con un número de países y organizaciones a efectos de desarrollar un enfoque metodológico para la medición de las habilidades de lectura y uso de números entre los jóvenes y adultos, de modo a que se mejore el cuerpo de evidencia estadística existente.

Su diseño se inspiró en encuestas conducidas en países de la Organización para la Cooperación y el Desarrollo Económico (OCDE), la Encuesta Internacional de Alfabetización de Adultos (IALS por sus siglas en inglés) y la Encuesta sobre Alfabetización de Adultos y Habilidades para Vida (ALL por sus siglas en inglés). Estas representaron los esfuerzos más importantes en la medición directa de competencias de alfabetismo y uso de números a nivel internacional. Aprender de sus fortalezas y debilidades contribuyó a definir LAMP en sus primeros años.

5.3.1. Marco conceptual y metodológico del LAMP²⁷

Las sociedades contemporáneas tienen la visión de la educación como un derecho humano fundamental, desde que se proclamara formalmente la Declaración Universal de los Derechos Humanos (1948) tras la devastadora experiencia de la Segunda Guerra Mundial. A partir de entonces, el alcance y las especificaciones de este

²⁷ Extraído del Documento técnico n° 1 del Instituto de Estadística de la UNESCO, "La Nueva Generación de Estadísticas sobre Competencias en Alfabetismo: Implementando el Programa de Evaluación y Monitoreo de la Alfabetización (LAMP)", UNESCO-UIS, 2009, Montreal, Canadá.



derecho no cesaron de crecer, basados en sucesivas redefiniciones de la problemática educativa. En este sentido, el concepto de educación propugnado por la UNESCO se basa en los cuatro pilares de la educación presentados en 1996 por el Grupo de Trabajo sobre Educación para el Siglo XXI: (UNESCO, 1996) aprender a ser, aprender a vivir juntos, aprender a conocer y aprender a hacer.

Más tarde cobró fuerza la idea de una educación a lo largo de toda la vida, “de la cuna hasta la tumba”, como un “marco filosófico y conceptual y un principio organizativo de todas las formas de educación”²⁸, donde la educación de adultos se reconoce como un componente esencial de este derecho. Esto ante la constatación de que cada vez más personas se encuentran limitadas en sus oportunidades de participación en la perspectiva de una sociedad basada en la información y el conocimiento, por su limitado acceso a la educación formal o no formal, especialmente en los países en desarrollo.

Por consiguiente, la educación se entiende como intrínsecamente importante para el desarrollo humano, ya que se encuentra inserta en el proceso de mejoramiento de las oportunidades y libertades de cada individuo, que le permitirán llevar adelante la clase de vida que valora, respetando los derechos de los demás.

Por su parte, la alfabetización es un componente central de la educación en general, particularmente con relación a la operación de los sistemas educativos nacionales que fueron originalmente concebidos como una forma de garantizar que cada ciudadano esté dotado de las competencias básicas en alfabetismo. Asegurar estas competencias para todas las personas es un objetivo central de todo sistema educativo. Sin embargo, es importante tener en cuenta que las visiones de la alfabetización han evolucionado con el pasar del tiempo. En este sentido, la UNESCO, luego de sucesivas redefiniciones ha declarado lo siguiente:

- El alfabetismo es la habilidad de identificar, comprender, interpretar, crear, comunicar y calcular, utilizando materia-

28 Marco de Acción de Belém. Aprovechar el poder y el potencial del aprendizaje y la educación de adultos para un futuro viable. CONFINTEA VI, Belem, 2009.



les impresos y escritos asociados con diversos contextos. El alfabetismo involucra un continuo aprendizaje que habilita a las personas a alcanzar sus objetivos, desarrollar sus conocimientos y potenciales y participar plenamente en la comunidad y en la sociedad ampliada. (Reunión de expertos de 2003 y publicada en UNESCO, 2005); y

- Adicionalmente, la noción de “pluralidad de la alfabetización” (2004) enfatiza su dimensión social tanto con relación a su adquisición como aplicación. En consecuencia, la alfabetización se concibe como un conjunto que abarca prácticas diversas insertas en contextos socioeconómicos, políticos, culturales y lingüísticos, y adquiridas dentro y fuera de la escuela. También involucra el contexto de la familia y comunidad, los medios de comunicación a través de distintas tecnologías, las competencias para seguir aprendiendo, el mundo laboral y la vida en general. Por ende, este concepto de alfabetización enfatiza que los desafíos implican alfabetizar no solamente a los individuos sino también a las sociedades (UNESCO, 2004).

La evolución de estos conceptos tiende a enfatizar ciertos elementos esenciales que deben ser tomados en cuenta en cualquier discusión sobre la alfabetización y su medición: i) la importancia de utilizar textos; ii) la necesidad de hacerlo con comprensión, es decir, en forma competente; iii) no restringirse a los textos, sino incluir la utilización de números; iv) la necesidad de examinar estos temas en función de experiencias de la vida diaria donde los múltiples y diversos aspectos de la vida social se concretan; y v) comprender que, en cualquier entorno, la vida cotidiana nunca será una realidad aislada que sólo afecta a las personas a nivel individual.

Esta definición hace hincapié en dos elementos: las competencias (lectura, escritura y utilización de números) y los distintos escenarios de la vida social donde las competencias de alfabetismo se manifiestan.

LAMP representa un intento por proporcionar un enfoque viable que permita generar información sobre el fenómeno de la alfabetización y ha sido diseñado tomando en cuenta la noción de “pluralidad de



la alfabetización” propuesta por la UNESCO. Esta idea enfatiza el entronque del alfabetismo con los aspectos sociales, económicos y culturales. De hecho, la intención del LAMP es generar información sobre lo que las personas saben y son capaces de hacer con relación a un conjunto definido de aspectos críticos asociados con el manejo de textos en situaciones de la vida cotidiana. De esta manera, LAMP se enfoca en los elementos relativos a la lectura y no aborda la escritura.

LAMP intenta entregar una imagen sólida de los elementos clave que deben tenerse en cuenta en el diseño, implementación y evaluación de las políticas públicas. Entrega una visión detallada de la distribución de competencias en alfabetismo (textos continuos, textos esquemáticos y utilización de números) y examina los factores que podrían estar asociados con los bajos desempeños (componentes de lectura). Esta información es interpretada en función del contexto socioeconómico.

LAMP tiene como propósito brindar a los tomadores de decisión información robusta y evidencia crítica sobre el perfil de la población adulta con relación a habilidades de lectura y de uso de números a los efectos de alimentar los debates públicos y el diseño de programas de alfabetización y educación de adultos en el marco de las políticas educativas y los programas sociales. LAMP explora la distribución de diferentes habilidades entre la población y la necesidad de usar esas habilidades para operar en situaciones de la vida diaria. Por lo tanto, LAMP puede informar de modo significativo acerca de elementos centrales relativos a los derechos de las personas a la educación.

LAMP es un esfuerzo para dotar a los países con las herramientas metodológicas que requieren para conducir evaluaciones directas de las habilidades de alfabetismo y uso de números y por fortalecer las capacidades nacionales. Al combinar estos dos elementos, LAMP busca asegurar que los países que lo implementen lleguen a producir información sobre el alfabetismo de modo robusto y sostenido. Está particularmente interesado en asegurar la apropiación nacional. Así, éste no es un estudio que se conduce una vez según su propio interés y agenda, sino que ha sido diseñado para contribuir al desarrollo de capacidades nacionales, reposando en la ex-



periciencia ya existente y promoviéndola como el mejor mecanismo para respaldar iniciativas nacionales.

LAMP examina el alfabetismo en tres dominios diferentes: la lectura de textos continuos (prosa); la lectura de textos discontinuos o esquemáticos (documentos); y el uso de números. Las habilidades vinculadas a la prosa son definidas como la capacidad para tratar textos organizados en párrafos, con sangrías y encabezados, que revelan su estructura y propósito. Las habilidades vinculadas a los documentos son definidas como la capacidad para usar textos no continuos en diferentes formatos (tablas, horarios, gráficos, mapas, etc.) que permite al lector usar diferentes estrategias para ingresar y extraer información de éstos. Finalmente, las habilidades vinculadas al uso de números incluyen completar tareas breves con contenido matemático. Éstas pueden demandar el uso de ciertas funciones como calcular, estimar y entender nociones de forma, longitud, volumen, unidades monetarias y otras funciones de medida. Se busca que las tareas usadas para la evaluación sean significativas para los entrevistados en términos de su relación con situaciones diarias como: el hogar y la familia, salud y seguridad, comunidad y ciudadanía, consumo, trabajo, ocio y recreación.

Adicionalmente, LAMP explora factores vinculados con el bajo desempeño lector, brindando información clave para el diseño de intervenciones vinculadas a la alfabetización. De hecho, se ha realizado un esfuerzo adicional a los efectos de establecer los marcos para la evaluación de los componentes de lectura. Asumiendo que leer implica tanto el reconocimiento de las palabras como su comprensión, este módulo evalúa cómo las personas aplican estas habilidades a efectos de procesar textos escritos. La capacidad para nombrar letras y números, para procesar frases y no sólo decodificar palabras, y la lectura fluida de párrafos, son indicadores que ayudan a describir a la población de acuerdo a su capacidad lectora.

Finalmente, LAMP genera información sobre las características socioeconómicas de los entrevistados y su uso de materiales escritos en diferentes ámbitos de la vida diaria, de modo que se cuente con elementos analíticos clave para caracterizar a la población.



La información reportada también puede ser utilizada para comparaciones internacionales dada la necesidad de monitorear el progreso tanto a nivel nacional como internacional en esta temática. Por esta razón, niveles comparables de habilidades son usados para el reporte de información.

5.3.2. Instrumentos utilizados por LAMP

Un cuestionario de antecedentes que genera información sobre las características socioeconómicas de los entrevistados incluyendo su uso de materiales escritos. Esto es clave para el análisis de la información generada por las herramientas cognitivas.

Una prueba filtro, orientada a brindar una primera aproximación al desempeño esperado de cada respondiente de modo que se pueda decidir qué conjunto de instrumentos deben ser usados para lograr una imagen más detallada de sus habilidades.

Un módulo para aquéllos con menor desempeño esperado, que establecerá con mayor precisión la ubicación del respondiente y explorará los elementos que pueden estar impidiendo que tenga un mejor desempeño (componentes de lectura).

Un módulo para aquéllos con mayor desempeño esperado, que establecerá, nuevamente y con mayor precisión la ubicación del respondiente. Estos instrumentos están acompañados por una documentación detallada que brinda suficiente guía acerca de las diferentes fases de implementación.



Capítulo VI

Medición de impacto del preescolar en el primer ciclo de la Educación Escolar Básica²⁹

Introducción

La evaluación de la calidad del preescolar en Paraguay forma parte de las acciones del Programa de Mejoramiento de la Educación Inicial y Preescolar (2003-2008), impulsado por el Ministerio de Educación y Cultura en el marco del Plan Nacional de Educación Inicial (2003-2012). Esta evaluación ha sido coordinada por la Dirección General de Planificación Educativa (DGPE) a través de la Dirección de Evaluación de la Calidad Educativa (DECE).

A fin de llevar adelante las acciones propuestas y medir la efectividad del preescolar, se seleccionó una muestra de 100 instituciones educativas y un total de 600 niños y niñas; de los cuales 400 estaban matriculados en el preescolar (grupo experimental o grupo de tratamiento) y 200 que no asistían al preescolar (grupo control). Se ha realizado el seguimiento de ambos grupos (experimental y de control) desde el preescolar hasta el 3° grado de la EEB. La primera prueba de medición de efectividad del preescolar se aplicó en el 2005, cuando el grupo estaba matriculado en el preescolar, y la última en el 2008.

29 MEC-BID Programa de Mejoramiento de Educación Inicial y Preescolar – Informe Final.



En el año 2005 el estudio de efectividad se realizó, en tres etapas: el primero en los meses de marzo-abril, el segundo en agosto y el tercero en octubre; las pruebas aplicadas en marzo del año 2007 midió las competencias del 1° grado, si bien fueron aplicadas a alumnos y alumnas del 2° grado. Asimismo, la aplicación al 3° grado del 2008 evaluó las competencias del 2° grado, atendiendo que la medición se llevó a cabo en el mes de marzo.

Cabe mencionar que de la muestra original, conformada por 600 niños, 437 fueron inicialmente evaluados, pero las series que fueron completadas y procesadas correspondieron a 425 niños/as. Esto debido a diferentes factores como: la migración de familias, niños controles que posterior a la primera visita (marzo-abril) ingresaron al preescolar, error en la toma de datos inicial; es decir, se seleccionaron niños que estaban fuera del rango de edad establecido y con necesidades educativas especiales. Considerando estos aspectos, finalmente el análisis se realizó sobre la base de 425 casos (niños/as y sus respectivas familias).

6.1. Marco del estudio

En la historia de la Educación Inicial en el Paraguay se identifican los procesos de reforma educativa como los momentos que marcan las etapas de su evolución. Tuvo su comienzo en 1922, con la llamada reforma educativa de Ramón Indalecio Cardozo, a través de la incorporación del kindergarten en algunas escuelas normales, posteriormente, con la reforma educativa de 1957, se organizó en dos etapas, el jardín y el pre-primario, luego con las innovaciones educativas 1973, cambia la denominación de las etapas y las edades reglamentarias. Con la reforma educativa 1992 se asumen los principios generales de la educación básica y el Plan de Acción de la Educación para Todos en el 2000, la organización del nivel inicial comprende al jardín maternal para los niños de 0 a 2 años; jardín infantil, para los de 3 a 4 años y preescolar para los de 5 años, el cual articula a la educación inicial con la escolar básica.



En el 2004 comienza la ejecución del Programa de Universalización y el Programa de Mejoramiento de la Educación Inicial y Preescolar, los cuales forman parte de las estrategias de acción del Plan Nacional de Educación Inicial (2003-2012), en ese contexto se elabora el marco curricular de la Educación Inicial y el Programa de Estudios del Jardín de Infantes y Preescolar.

En el marco curricular de la Educación Inicial se sustenta la importancia de la Educación Inicial como una concepción amplia e innovadora que apunta al desarrollo integral de los niños y niñas menores de 6 años, atendiendo sus necesidades bio-psicomotoras, cognitivas, sociales, emocionales y afectivas, así como la prevención y detección de dificultades e interferencias en su desarrollo y crecimiento. Así también, como una oportunidad de las madres y los padres de familia para mejorar y enriquecer sus prácticas de crianzas y a su vez un compromiso del personal docente y de apoyo para cumplir con los propósitos planteados.

La implementación del nuevo Programa de Estudios-Jardín de Infantes y Preescolar en el aula se inicia en el 2005, año en que arranca el Estudio de Efectividad del Preescolar y la Medición del Impacto del Preescolar en el rendimiento en el primer ciclo de la Educación Escolar Básica.

6.2. Objetivo del estudio

Medir el impacto del Preescolar en el rendimiento del primer ciclo de la Educación Escolar Básica.

6.3. Hipótesis del estudio

La hipótesis del estudio es que existe una relación positiva entre la educación Preescolar y el rendimiento en el primer ciclo de la Educación Escolar Básica. Esto es, si todos los niños y niñas cursaran el Preescolar, habría un incremento en el rendimiento de la Educa-



ción Escolar Básica (EEB), que no existiría si los niños y niñas no cursaran el Preescolar.

6.4. La muestra

6.4.1. Población

La población estuvo constituida por niños y niñas de 5 años de edad matriculados en el preescolar de modalidad formal³⁰.

Cabe destacar que para la selección de las instituciones educativas no fueron considerados los siguientes casos:

- Instituciones educativas con preescolar funcionando en una sección plurigrado³¹.
- Instituciones educativas indígenas³².
- Instituciones educativas que ofrecen el preescolar con matrícula menor a 10 alumnos.
- Instituciones educativas de los departamentos de Boquerón y Alto Paraguay³³ (Chaco) y Misiones.

Las instituciones que no formaron parte del marco muestral representan el 6,9% del total de la matrícula del preescolar.

30 La educación inicial y preescolar formal es la atención sistemática, graduada y normada brindada a los niños menores de 6 años en las instituciones educativas, sobre la base de orientaciones específicas, un programa de estudios, y un sistema de evaluación, explicitados y aprobados por el MEC. MEC, 2004. Op. Cit, pág. 21.

31 Comprenden dos o más combinaciones de etapa/grados consecutivos o no, y funcionan simultáneamente en un aula, a cargo de un maestro, en un turno escolar y desarrollan programas educativos diferentes. MEC, 2008. Estadísticas Educativas: Educación Inicial. p. 10.

32 Se tiene previsto realizar un estudio particular con los preescolares indígenas.

33 Debido a que dichos departamentos presentan ciertas características como ser: población dispersa, difícil acceso y bajo nivel poblacional. DGGEC, 2003. Encuesta Permanente de Hogares.



6.4.2. Selección de instituciones educativas

Una vez definido el marco muestral, se realizó un muestreo aleatorio proporcional de acuerdo a la distribución de la matrícula por zona (rural-urbana) y sector (oficial-privado/privado subvencionado).

Tabla 1: Distribución de la muestra por zona, según sector.

SECTOR	ZONA		Total
	Urbana	Rural	
Oficial	38	41	79
Privado *	19	2	21
Total	57	43	100

Fuente: MEC- DGIMPCE, SIEC 2003. *Incluye el sector privado subvencionado.

La distribución de la muestra por departamento estuvo conformada de la siguiente manera: Asunción 5, Concepción 3, San Pedro 16, Cordillera 2, Guairá 4, Caaguazú 5, Caazapá 2, Itapúa 11, Paraguari 8, Alto Paraná 17, Central 23, Ñeembucú 2, Amambay 1 y Canindeyú 1.

6.4.3. Selección de secciones del preescolar

Se realizó un sorteo para elegir una sección del preescolar por cada institución educativa, en el caso de contar con más de una sección.

6.4.4. Selección de niños del grupo experimental por sección

Las personas encargadas de aplicar los instrumentos solicitaron al docente del preescolar la lista de alumnos presentes, para realizar la selección al azar (por sorteo) de 2 niños y 2 niñas. Cabe señalar que del proceso de selección no participaron niños con las siguientes características:

- Con alguna discapacidad evidente para sus docentes.
- Menores de 5 años.
- Mayores de 5 años 5 meses.



El grupo experimental estuvo compuesto por 352 niños/niñas.

6.4.5. Selección de niños del grupo control

Para el grupo control se procedió a identificar a niños/niñas de la comunidad que no estuviesen asistiendo al preescolar (para este procedimiento se solicitó la ayuda de los docentes y directivos de las instituciones educativas) para lo cual las personas encargadas de aplicar los instrumentos tuvieron en cuenta las mismas características consideradas para la selección del grupo experimental. El grupo control estuvo compuesto por 85 niños/niñas.

6.4.6. Pérdida de casos

La muestra original, si bien estuvo conformada por 600 niños y niñas, el estudio se realizaba sobre un universo de 437 niños y niñas y sus familias; esta diferencia puede ser atribuida entre otros motivos a: i) la migración de familias, ii) niños del grupo control que posterior a la primera visita (marzo-abril) se matricularon en el preescolar, iii) inconvenientes en la toma de datos inicial que fueron detectados en el proceso; tales como, niños que estaban fuera del rango de edad establecido y con necesidades educativas especiales.

6.5. Metodología del estudio

Para la medición del impacto del preescolar en el rendimiento en el primer ciclo de la Educación Escolar Básica se formuló el problema en términos de lo que en la literatura de análisis causal se denomina Efecto Promedio del Tratamiento (EPT) y Efecto Promedio del Tratamiento en los Tratados (EPTT); en inglés Average Treatment Effect (ATE) y Average Treatment Effect on the Treated (ATT), respectivamente.

Más específicamente, el EPT se formuló del siguiente modo: si en la cohorte incluida en el estudio, todos los niños y niñas hubiesen



cursado el preescolar ¿Habría sido significativamente mayor el rendimiento promedio de la población de alumnos?

El EPTT se formuló del siguiente modo: Si los niños y niñas que cursaron el preescolar no lo hubiesen hecho ¿Habría sido significativamente menor el rendimiento promedio de estos niños?

La palabra “significativamente”, debe entenderse en el sentido de significación estadística.

6.5.1 Los instrumentos

Las pruebas utilizadas para el estudio son:

- Madurez del Preescolar.
- Disponibilidad léxica.
- Vocabulario.
- Cálculo.
- Problemas.

La Prueba de Madurez del Preescolar (PMPE) mide las funciones cognitivas de los preescolares directamente a través de pruebas. Se divide en 5 sub-escalas que están dirigidas a capturar las capacidades cognitivas de los niños en las áreas de reconocimiento, comprensión, establecimiento de relaciones, establecimiento de operaciones de clasificación y de ordenamiento temporal.

La Escala de Disponibilidad Léxica mide el grado de disponibilidad de un término en la mente de un hablante, esto es, la facilidad o dificultad con que un vocablo aparece en la conciencia del niño en un momento dado. Este índice procesa las frecuencias de aparición de vocablos y permite un examen del dato léxico en áreas o centros de interés.

Los instrumentos vocabulario sobre dibujos, cálculo y problemas aplicados forman parte de un conjunto de pruebas para medir el desarrollo cognitivo “Batería Woodcock-Muñoz” (Batería-R). En la prueba de Vocabulario sobre dibujos, el/la niño/a debe identificar imágenes de acuerdo a sus nombres, por páginas completas, hasta



que el niño haya respondido incorrectamente a 6 ítems consecutivos. El instrumento consta de 58 ítems.

En la prueba de Cálculo, el/la niño/a debe ir efectuando los ejercicios presentados en el instrumento hasta que haya respondido incorrectamente a 6 ítems consecutivos, el mismo consta de 58 ítems. Atendiendo al objetivo del estudio han sido utilizados 18 ítems. Similarmente, en la prueba de Problemas Aplicados, el/la niño/a debe ir respondiendo a los planteamientos presentados por el aplicador hasta que haya respondido incorrectamente a 6 ítems consecutivos. El instrumento consta de 59 ítems, pero los/as niños/as llegaron a responder hasta el ítems 26.

Los resultados de las pruebas citadas, aplicados al grupo de tratamiento y de control, junto a covariables tales como entorno familiar y formación de los maestros, que han sido consignados para cada niño en el proceso de recolección de datos, constituyen el insumo de información disponible para efectuar el análisis de impacto del preescolar en el primer ciclo de la Educación Escolar Básica.

Los resultados de las pruebas han sido expresados en una escala compuesta de cinco niveles: Excelente, Muy Bueno, Bueno, Aceptable e Insuficiente. Para construir la escala, se procedió a identificar el máximo puntaje de cada prueba y luego se dividió el rango de cero al máximo en cinco segmentos iguales, a los que se aplicó las denominaciones mencionadas, según su orden.

6.6 Fases del estudio

El estudio comprendió dos fases: Fase Descriptiva y Fase Causal.

Fase descriptiva: El objetivo de esta fase es proveer información sobre las relaciones que existen entre las variables que conforman la base de datos. Si bien la información obtenida en esta fase no posee carácter causal, pues “relación no implica causalidad”. Sin embargo, las relaciones obtenidas mediante estadísticas descripti-



vas, pueden ayudar a guiar el análisis causal así como la interpretación de los resultados.

Para el análisis descriptivo de la información, se utilizó la técnica de redes bayesianas y la técnica de regresión lineal. La primera, debido a que provee instrumentos de análisis visual de la información, lo cual permite apreciar de una manera simple las relaciones más importantes; y la segunda, debido a que la regresión lineal es una de las herramientas más comunes para el análisis de datos.

La herramienta de red bayesiana utilizada en este estudio es GENIE³⁴. La herramienta de análisis estadístico es R³⁵. La elección de las citadas herramientas se realizó debido a que cuentan con las funcionalidades requeridas para el trabajo establecido y continuamente incorporan otras nuevas, lo cual las ubica en la primera línea de las herramienta de su tipo; además, se hallan disponibles en la modalidad de software libre, por lo que no hay restricción de carácter presupuestario que impida el uso de estas herramientas en el MEC.

Fase causal: Está orientada a establecer si la relación observada, en la fase descriptiva, posee o no carácter causal, así como el grado de intensidad de la misma.

Para el estudio, se utilizó la base de datos construida por el MEC en base a un diseño cuasi-experimental, cuya estructura básica está descrita en la introducción de este documento; más detalles sobre la fuente de datos y su recolección se encuentra en el documento "Resultados de la Evaluación de Efectividad del Preescolar", MEC, Agosto 2009. El Apéndice de este documento provee información sobre el proceso seguido para la puesta a punto de la base de datos, a partir de los archivos originales.

Dado que la fase descriptiva mostró la existencia de una relación positiva entre la realización del preescolar y el rendimiento en los grados siguientes, la fase explicativa se orientó a determinar si la

34 Decision Systems Laboratory, University of Pittsburgh, Version 2.0.3568.0.

35 The R Foundation for Statistical Computing, Version 2.4.1.



relación observada posea o no carácter causal. Al efecto, se procedió en dos etapas.

Primero, se realizó un pre-procesamiento de los datos, empleando la técnica de matching con el objetivo de que la información a ser utilizada para el cálculo de impacto permita que el grupo de tratamiento y el de control sean lo más homogéneos posible, previo al tratamiento; atendiendo que en la base de datos sin procesar dichos grupos poseen características diferentes, tales como el entorno familiar, condición socioeconómica. Con esto se buscó reducir el sesgo en la estimación causal que produce el empleo de datos que provienen de observaciones o cuasi-experimentos, como es el caso de este estudio.

Luego, se procedió a la estimación del EPT y el EPTT, con el empleo de un modelo lineal. El matching se realizó con el paquete *Machtl*³⁶ y la estimación del EPT y el EPTT con el paquete *Zelig*³⁷. Ambos paquetes son adiciones al R estándar.

Debido a que facilita la obtención de una buena solución de matching, el método empleado es la búsqueda por algoritmo genético³⁸, implementada en el paquete *Machtl*.

Sobre los datos procesados mediante el procedimiento de matching, fue aplicado el modelo lineal, en el que el tratamiento es identificado mediante una variable binaria, que indica si el niño o la niña asistieron o no al preescolar; y el efecto está dado por el resultado de la prueba correspondiente, en el 2008.

36 Desarrollado por Daniel E. Ho, Kosuke Imai, Gary King y Elizabeth A. Stuart.

37 Desarrollado por Kosuke Imai, Gary King y Olivia L.

38 Diamond, A. and Sekhon, J. (2005), "Genetic Matching for Estimating Causal. Effects: A New Method of Achieving Balance in Observational Studies". <http://jsekhon.fas.harvard.edu/>



6.7 Análisis de resultados

Como se ha mencionado anteriormente el análisis se realizó en dos fases (descriptiva y causal), los resultados se presentan en la secuencia en que fue realizado dicho análisis.

Primeramente, se presenta un conjunto de estadísticas descriptivas, que busca proveer indicios que permitan identificar las relaciones entre las variables del estudio, de modo a proveer una guía para el posterior análisis causal.

El objetivo principal de la fase descriptiva es determinar si existe una relación positiva entre el rendimiento y la asistencia al preescolar; luego, es la identificación de posibles factores que puedan explicar la relación observada.

Posteriormente, se presenta el análisis efectuado para determinar si la relación observada entre el rendimiento y la asistencia al preescolar es tal que:

- i) si en la cohorte incluida en el estudio todos los niños y niñas hubiesen cursado el preescolar el rendimiento promedio de la población de estudiantes hubiese sido significativamente superior;
- ii) si los niños y niñas que cursaron el preescolar no lo hubiesen hecho hubiese sido menor el rendimiento promedio de estos niños y niñas.

En el gráfico 1, puede verse una descripción de los datos agregados que incluye al grupo de tratamiento y de control, de la muestra; en la que puede leerse la distribución de los resultados así como el promedio de cada prueba efectuada, en la secuencia de su realización.

Este gráfico permite apreciar el rendimiento en las diferentes pruebas aplicadas, de todos los niños y niñas que formaron parte de la muestra, sin discriminar si asistió o no al preescolar.



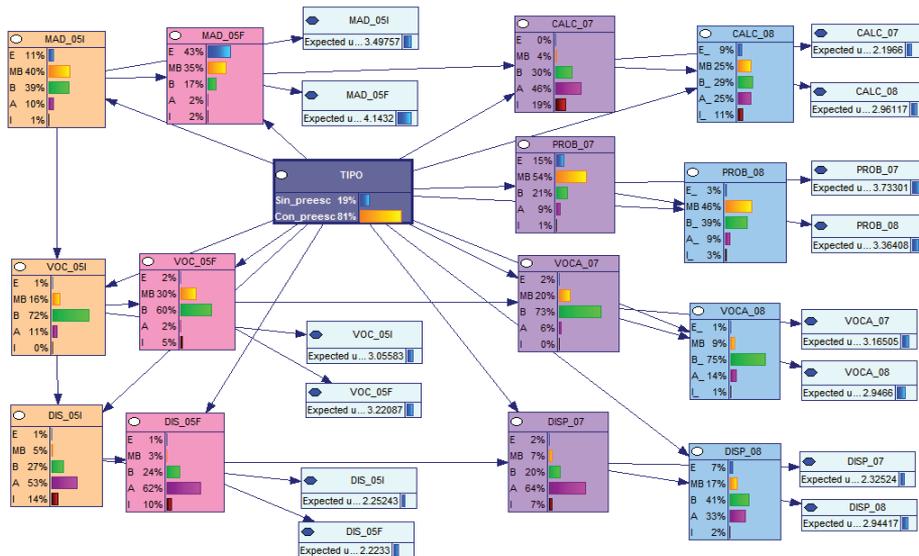


Gráfico 1: Resultado de las pruebas aplicadas a los niños y niñas de la muestra.

Para la lectura de los resultados, las siguientes equivalencias son importantes:

MAD_05I y MAD_05F	Madurez al inicio y al final del 2005, respectivamente.
VOC_05I y VOC_05F	Vocabulario al inicio del 2005 y al final del 2005.
DIS_05I y DIS_05F	Disponibilidad léxica al inicio del 2005 y al final del 2005.
CALC_07 y CALC_08	Rendim. en la realización de cálculos, en el 2007 y 2008.
PROB_07 y PROB_08	Rendim. en la solución de problemas, en el 2007 y 2008.
VOCA_07 y VOCA_08	Vocabulario en el 2007 y el 2008.
DISP_07 y DISP_08	Disponibilidad léxica en el 2007 y 2008.

En los gráficos de barras, la **E**, **MB**, **B**, **A** e **I**, representan Excelente, Muy Bueno, Bueno, Aprobado e Insuficiente, respectivamente; que son las categorías en las que se clasifican el rendimiento de los niños y las niñas. Se recuerda que para construir las categorías, se procedió a identificar el máximo puntaje de cada prueba y luego se dividió el rango de cero al máximo en cinco segmentos iguales, a los que se aplicó las denominaciones mencionadas.



Las cajas con barras muestran la distribución del rendimiento. Las cajas sin barras muestran el rendimiento promedio; para cuyo cálculo se asigna el valor 5 a la categoría E, 4 a MB y así sucesivamente hasta 1. De esta manera puede verse, por ejemplo, que en la primera prueba de madurez, MAD_05I, realizado al inicio del año escolar 2005, el 11% de los niños y niñas de la muestra, sin distinguir entre grupo de tratamiento y control, obtuvo Excelente, 40% obtuvo Muy Bueno, 39% Bueno, 10% consiguió un rendimiento aceptable y 1% se encuentra en la categoría Insuficiente; el rendimiento promedio de esta prueba fue 3,49.

Los resultados muestran la evolución del rendimiento de los niños y niñas de la muestra en las diferentes pruebas aplicadas.

Las conexiones señalan una relación, que podría además tener carácter causal. Así, puede verse que VOC_05I antecede a VOC_05F, y que ésta a antecede a VOCA_07, la que a su vez antecede a VOCA_08.

En la secuencia de las pruebas de vocabulario, señaladas en el párrafo anterior, además de existir un orden temporal, es razonable suponer que exista un patrón en los resultados de las pruebas sucesivas.

Resulta razonable suponer que el niño que obtenga cierto rendimiento en una prueba, tienda a obtener un rendimiento similar en pruebas futuras.

Por ejemplo, si un niño obtiene un rendimiento excelente al inicio, es menos probable que este niño ingrese a la categoría insuficiente, comparado a un niño que en la misma prueba inicial haya obtenido insuficiente. Similar patrón cabe esperar en las demás pruebas, lo cual es expresado por las fechas que se observan en el gráfico 1.

En el centro del gráfico se encuentra la caja (variable) con la etiqueta TIPO, la cual cuenta con dos categorías; Sin_preesc y Con_preesc, que indican sin un determinado niño cursó (Con_preesc) o no cursó (Sin_preesc) el preescolar. Aquí puede apreciarse que



el 19% de los niños que componen la muestra no han cursado en preescolar y que el 81% si lo ha hecho.

Debe notarse que de la variable TIPO parten flechas que apuntan a todas las demás variables que componen el gráfico, esto es debido a que es precisamente la incidencia del preescolar – cuya realización o no es señalada por las categorías, Con_preesc y Sin_preesc, que conforman la variable TIPO – el objetivo del análisis. El sentido de la flecha indica la orientación de la relación bajo estudio.

Los resultados que se observan en el gráfico 1 corresponden a toda la muestra. Los gráficos siguientes, gráfico 2 y gráfico 3, muestran los resultados obtenidos por los niños en el grupo de tratamiento (Con_preesc) y en el de control (Sin_preesc).

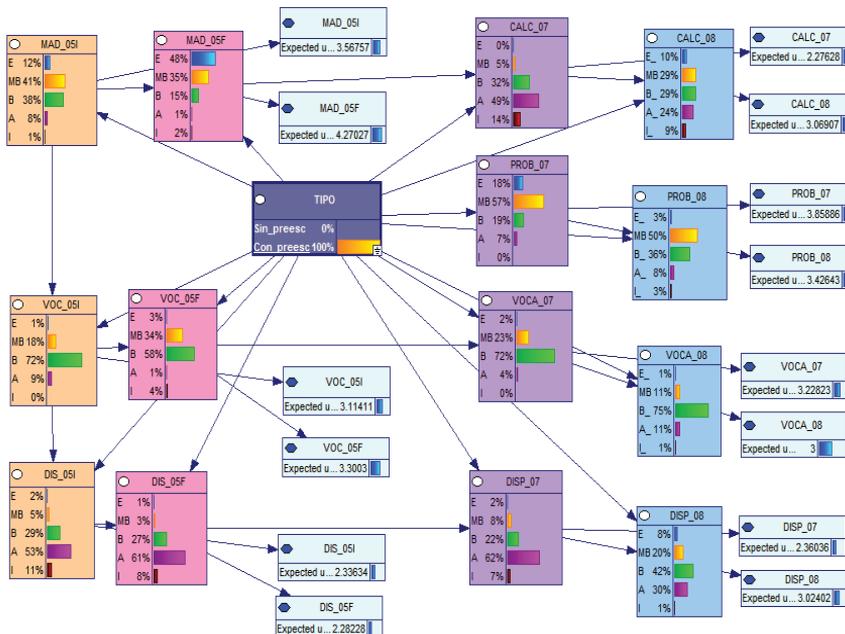


Gráfico 2: Resultados de las pruebas aplicadas a los niños y niñas que asisten al preescolar.

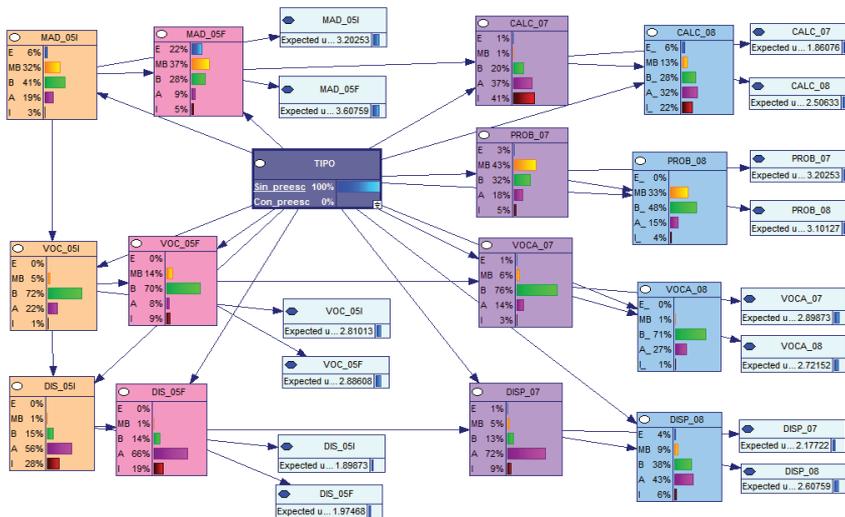


Gráfico 3: Resultados de las pruebas aplicadas a los niños y niñas sin preescolar.

Al comparar los resultados de los dos últimos gráficos, puede apreciarse que existe diferencia en el rendimiento de los dos grupos a favor de los niños y niñas que cursaron el preescolar; en este punto no se puede afirmar que dicha diferencia esta explicada por el tratamiento, es decir, que exista una relación causal.

Tabla 2: Resultados 2008.

TIPO	Cálculo	Problemas	Disponibilidad Léxica	Vocabulario
Con_preesc	3,069	3,43	3,024	3,00
Sin_preesc	2,51	3,10	2,61	2,72

Tabla 3: Resultado 2007.

TIPO	Cálculo	Problemas	Disponibilidad Léxica	Vocabulario
Con_preesc	2,27	3,85	2,36	3,22
Sin_preesc	1,86	3,20	2,17	2,89

Al final del 2005 se obtuvieron los siguientes resultados:

Tabla 4: Resultado final 2005.

TIPO	Madurez	Disponibilidad Léxica	Vocabulario
Con_preesc	4,27	2,28	3,30
Sin_preesc	3,60	1,97	2,88

Así pues, en la prueba del 2008, del 2007 y de fin del 2005, se ve que los niños y niñas que han asistido al preescolar tienen consistentemente un rendimiento superior al de los niños y niñas que no lo hicieron. La pregunta que surge en este punto es si la diferencia se debe al hecho de haber cursado el preescolar o a alguna otra razón.

Para orientarse hacia una respuesta a la pregunta anterior, es importante tener en cuenta el resultado de la primera prueba, a inicios del 2005, antes de que los niños y las niñas del grupo de tratamiento hayan cursado el preescolar, obtuvieron mejores resultados en los tres aspectos evaluados.

Tabla 5: Resultado inicio 2005.

TIPO	Madurez	Disponibilidad Léxica	Vocabulario
Con_preesc	3,56	2,33	3,11
Sin_preesc	3,20	1,89	2,81

Los niños y niñas del grupo de tratamiento (Tabla 5) ya al inicio se encuentran en mejores condiciones que los del grupo de control. Lo que queda por averiguar es si la diferencia de rendimiento del 2008 puede ser atribuido, al menos en parte, al hecho de que los niños y niñas hayan asistido al preescolar (tratamiento) o podría ser consecuencia de otros factores, tal como una ventaja inicial, que luego se mantiene, explicada por el entorno familiar u otros factores.

La fase causal busca dar una respuesta a la pregunta anterior; pero antes; abajo se presentan los resultados de la estimación de modelos de regresión lineales, aplicados a los datos disponibles.



Para cada uno de los aspectos evaluados se utilizó el modelo de regresión lineal aplicados a los datos disponibles con el propósito de ilustrar las relaciones exploradas previamente con las técnicas de las redes, con una técnica alternativa habitualmente utilizada en este tipo de estudios. Sin embargo, debe remarcarse que los resultados del análisis de regresión tienen exclusivamente un carácter descriptivo y no deben interpretarse en un sentido causal.

```

CALC_08
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.601508    1.066687   0.564  0.5731
CANT_HNO     -0.042643    0.024603  -1.733  0.0839
CANT_PER     -0.006113    0.011677  -0.524  0.6009
COMP_HAB_N   0.011044    0.293042   0.044  0.9632
COMP_HAB_O   -0.091083    0.192171  -0.474  0.6358
COMP_HAB_P   0.060361    0.201459  -0.300  0.7646
CONT_MED     -0.146612    0.137065  -1.070  0.2854
DIS_OSI      0.120786    0.077750   1.554  0.1211
ESP_JUG      0.247247    0.800852   0.309  0.7577
ESTU_MA_PrI  -0.053729    0.253981  -0.212  0.8326
ESTU_MA_Sec  -0.029239    0.276972  -0.106  0.9160
ESTU_MA_Ter  -0.131489    0.331343  -0.397  0.6917
ESTU_PA_PrI  0.460447    0.259388   1.775  0.0767
ESTU_PA_Sec  0.416294    0.270351   1.540  0.1244
ESTU_PA_Ter  0.583027    0.333737   1.747  0.0814
MAD_OSI      0.091008    0.074156   1.227  0.2205
PARTI_ZO     0.065937    0.118973   0.554  0.5798
SECTOR       0.263016    0.150891   1.743  0.0821
TIENE_MA     0.635763    0.618700   1.028  0.3048
TIENE_PA     -0.456509    0.429627  -1.063  0.2886
TIPO         0.305835    0.153016   1.999  0.0463 *
TIPO_VIV     -0.155211    0.142767  -1.087  0.2776
TRAB_MA      0.030004    0.122811   0.244  0.8071
TRAB_PA      0.190977    0.339968   0.562  0.5746
VIVE_CON_M   -0.028437    0.222910  -0.128  0.8986
VIVE_CON_MYP 0.070353    0.125409   0.561  0.5751
VIVE_CON_P   0.390373    0.748207   0.522  0.6021
VOC_OSI      0.214041    0.117797   1.817  0.0700
ZONA         0.347913    0.134210   2.592  0.0099 **

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.066 on 383 degrees of freedom
Multiple R-Squared:  0.2026, Adjusted R-squared:  0.1443
F-statistic: 3.476 on 28 and 383 DF,  p-value: 2.298e-08

```

Tabla 6. Prueba de Cálculo. Aplicación 2008.

```

PROB_08
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.5032305    0.7828027   1.920  0.0556
CANT_HNO     0.0009349    0.0180555   0.052  0.9587
CANT_PER     0.0113235    0.0085680   1.322  0.1871
COMP_HAB_N   0.0542133    0.1856982   0.292  0.7703
COMP_HAB_O   -0.0608060    0.1410272  -0.431  0.6666
COMP_HAB_P   -0.0322940    0.1478436  -0.218  0.8272
CONT_MED     0.0154964    0.1005874   0.154  0.8776
DIS_OSI      0.0930700    0.0570580   1.631  0.1037
ESP_JUG      0.1455731    0.5877164   0.248  0.8045
ESTU_MA_PrI  0.0494501    0.1863875   0.265  0.7909
ESTU_MA_Sec  0.0779650    0.2032598   0.384  0.7015
ESTU_MA_Ter  0.0853134    0.2431610   0.351  0.7259
ESTU_PA_PrI  0.2395585    0.1903558   1.258  0.2090
ESTU_PA_Sec  0.1712965    0.1984007   0.863  0.3885
ESTU_PA_Ter  0.1003398    0.2449177   0.410  0.6823
MAD_OSI      0.0859776    0.0544205   1.580  0.1150
PARTI_ZO     -0.0610260    0.0873098  -0.699  0.4850
SECTOR       0.0919983    0.1107334   0.831  0.4066
TIENE_MA     0.6139492    0.4540413   1.352  0.1771
TIENE_PA     0.5785194    0.3152878   1.835  0.0673
TIPO         0.2338026    0.1122929   2.082  0.0380 *
TIPO_VIV     0.1040044    0.1047718   0.993  0.3215
TRAB_MA      0.0434426    0.0901266   0.482  0.6301
TRAB_PA      0.1558399    0.2494904   0.625  0.5326
VIVE_CON_M   -0.1520048    0.1635857  -0.929  0.3534
VIVE_CON_MYP -0.0012504    0.0920322  -0.014  0.9892
VIVE_CON_P   -0.2293394    0.5490818  -0.418  0.6764
VOC_OSI      0.1114913    0.0864470   1.290  0.1979
ZONA         0.2051990    0.0984922   2.083  0.0379 *

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.782 on 383 degrees of freedom
Multiple R-Squared:  0.1241, Adjusted R-squared:  0.06007
F-statistic: 1.938 on 28 and 383 DF,  p-value: 0.003408

```

Tabla 7. Prueba de Problemas. Aplicación 2008.

```
DISP_08
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.950904  0.879907  3.354 0.000877 ***
CANT_HNO     -0.016129  0.020295  -0.795 0.427277
CANT_PER     -0.005403  0.009631  -0.561 0.575112
COMP_HAB_N   -0.186481  0.208733  -0.893 0.372207
COMP_HAB_O   -0.037645  0.158521  -0.237 0.812413
COMP_HAB_P   -0.058875  0.166183  -0.354 0.723327
CONT_MED     -0.047296  0.113065  -0.418 0.675958
DIS_05I      0.186874  0.064136  2.914 0.003781 **
ESP_JUG      0.388996  0.660621  0.589 0.556320
ESTU_MA_Pri  -0.237831  0.209508  -1.135 0.257007
ESTU_MA_Sec  -0.160973  0.228474  -0.705 0.481513
ESTU_MA_Ter  -0.374049  0.273324  -1.369 0.171952
ESTU_PA_Pri  0.102848  0.213969  0.481 0.631026
ESTU_PA_Sec  0.060122  0.223012  0.270 0.787619
ESTU_PA_Ter  -0.119851  0.275299  -0.435 0.663556
MAD_05I     -0.041186  0.061171  -0.673 0.501171
PARIE_ZO    -0.044957  0.098140  -0.458 0.647150
SECTOR      0.018266  0.124470  0.147 0.883409
TIENE_MA    -0.892403  0.510364  -1.749 0.081168 .
TIENE_PA    -0.025125  0.354398  -0.071 0.943520
TIPO        0.403203  0.126222  3.194 0.001517 **
TIPO_VIV   -0.113543  0.117768  -0.964 0.335594
TRAB_MA     0.201481  0.101306  1.989 0.047431 *
TRAB_PA     0.039456  0.280439  0.141 0.888187
VIVE_CON_M  0.050234  0.183878  0.273 0.784768
VIVE_CON_MYP 0.044737  0.103499  0.432 0.665655
VIVE_CON_P  0.518221  0.617194  0.840 0.401634
VOC_05I     0.004370  0.097170  0.045 0.964155
ZONA        0.340778  0.110710  3.078 0.002233 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.879 on 383 degrees of freedom
Multiple R-Squared: 0.1538, Adjusted R-squared: 0.09179
F-statistic: 2.487 on 28 and 383 DF, p-value: 6.383e-05
```

Tabla 8. Prueba de Disponibilidad Léxica. Aplicación 2008.

```
VOCA-08
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.773142  0.491029  5.648 3.18e-08 ***
CANT_HNO     0.005291  0.011326  0.467 0.640646
CANT_PER     -0.009630  0.005374  -1.792 0.073961 *
COMP_HAB_N   -0.230181  0.116483  -1.976 0.048862 *
COMP_HAB_O   -0.081318  0.088462  -0.919 0.358547
COMP_HAB_P   -0.220693  0.092738  -2.380 0.017813 *
CONT_MED     0.013129  0.063096  0.208 0.835279
DIS_05I      0.040898  0.035791  1.143 0.253879
ESP_JUG     -0.700337  0.368657  -1.900 0.058224 .
ESTU_MA_Pri  0.025674  0.116915  0.220 0.826303
ESTU_MA_Sec  0.143888  0.127499  1.129 0.259797
ESTU_MA_Ter  0.111381  0.152528  0.730 0.465692
ESTU_PA_Pri  0.044140  0.119405  0.370 0.711834
ESTU_PA_Sec  0.091232  0.124451  0.733 0.463960
ESTU_PA_Ter  -0.120440  0.153630  -0.784 0.433548
MAD_05I     0.076073  0.034136  2.228 0.026428 *
PARIE_ZO    0.017776  0.054767  0.325 0.745674
SECTOR      0.092410  0.069460  1.330 0.184177
TIENE_MA    -0.095682  0.284807  -0.336 0.737090
TIENE_PA    0.378517  0.197771  1.914 0.056376 .
TIPO        0.204028  0.070438  2.897 0.003989 **
TIPO_VIV    0.038175  0.065720  0.581 0.561671
TRAB_MA     -0.094888  0.056534  -1.678 0.094078 .
TRAB_PA     -0.471220  0.156498  -3.011 0.002776 **
VIVE_CON_M  0.181337  0.102612  1.767 0.077990 .
VIVE_CON_MYP 0.047355  0.057729  0.820 0.412559
VIVE_CON_P  -0.179608  0.344423  -0.521 0.602337
VOC_05I     0.122237  0.054226  2.254 0.024747 *
ZONA        0.230031  0.061781  3.723 0.000226 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4905 on 383 degrees of freedom
Multiple R-Squared: 0.2618, Adjusted R-squared: 0.2078
F-statistic: 4.85 on 28 and 383 DF, p-value: 2.284e-13
```

Tabla 9. Prueba de Vocabulario. Aplicación 2008.



En todos los casos se puede apreciar que existe una relación significativa entre el TIPO, que es la variable binaria que indica si el alumno asistió o no al preescolar, y la variable dependiente en consideración, que en este caso puede ser CALC_08, PROB_08, DISP_08 o VOCA_08. Pero, debido a que los datos no provienen de un experimento aleatorio, sino de un cuasi-experimento, el resultado de la regresión no debe interpretarse directamente en un sentido causal. Para ello se requiere un procedimiento, tal como el matching, que reduce el sesgo en la estimación de parámetros causales a partir de datos obtenidos vía observación o cuasi-experimentos.

Para el matching fueron utilizados los siguientes factores, que han sido registrados como parte del estudio.

ZONA	Zona.
SECTOR	Sector.
CDIS_05I	Disponibilidad Léxica 2005 (aplicación inicial).
CVOC_05I	Vocabulario 2005 (aplicación inicial).
CMAD_05I	Madurez 2005 (aplicación inicial).
TIENE_MA	Tiene mamá.
TIENE_PA	Tiene papá.
ESTU_MA	Estudios de la madre.
ESTU_PA	Estudios del padre.
TRAB_MA	Trabaja la madre.
TRAB_PA	Trabaja el padre.
VIVE_CON	Con quien vive el niño.
CANT_HNO	Cantidad de hermanos.
COMP_HAB	con quien comparte la habitación.
CANT_PER	Cantidad de personas que viven en la casa.
ESP_JUG	Tiene espacio para jugar.
CONT_MED	Tiene acceso a un control médico.
TIPO_VIV	Tipo de vivienda.
PARIE_ZO	Tiene parientes en la zona.

Es importante notar que las variables arriba listadas podrían influir la decisión del tratamiento; esto es, la decisión de que un niño asista o no al preescolar, pero estas variables no pueden ser afectadas por esta decisión. Se aplica este procedimiento, pues el propósito



del matching es que el grupo de tratamiento y el de control sean lo más homogéneos posible, previo al tratamiento.

Abajo se muestran los resultados obtenidos en la estimación de los valores causales, EPT y EPTT.

Tabla 10: Efecto del Preescolar en Cálculo.

CALC_08	Promedio	Límite Inferior (2,5%)	Límite Superior (97,5%)
EPTT	0,0002236	-0,1120046	0,1121530
EPT	0,131	-0,09394	0,39934

Aunque en la prueba de cálculo efectuado en el 2008 se observa un efecto positivo del tratamiento en el valor de los parámetros muestrales, tanto para el conjunto de los niños incluidos en el estudio, EPT = 0,131; como para los niños que asistieron al preescolar, EPTT= 0,0002236; no se puede afirmar que estos valores muestrales observados provean evidencia suficiente como para decir que el correspondiente parámetro poblacional sea positivo; es decir no se puede afirmar con significación estadística que el preescolar incrementa el rendimiento escolar posterior.

Esto es debido a que, en ambos casos, el rango en el que pudiera encontrarse el parámetro poblacional incluye el 0. Así, se observa que el rango, con 95% de confianza, del EPTT es [-0,1120046; 0,1121530]. Similarmente, el rango del EPT es [-0,09394; 0,39934].

Los efectos estimados a partir de las pruebas realizadas en el 2008 para la resolución de problemas (PROB_08), disponibilidad léxica (DISP_08) y vocabulario (VOCA_08), se muestran en las tablas siguientes:

Tabla 11: Efecto del Preescolar en la Resolución de Problemas.

PROB_08	Promedio	Límite Inferior (2,5%)	Límite Superior (97,5%)
EPTT	-0,001433	-0,090446	0,077968
EPT	0,1195	-0,07415	0,35111

Tabla 12: Efecto del Preescolar en la Disponibilidad Léxica.

DISP_08	Promedio	Límite Inferior (2,5%)	Límite Superior (97,5%)
EPTT	0,001229	-0,087631	0,089621
EPT	0,2015	-0,07513	0,52187

Tabla 13: Efecto del Preescolar en el Vocabulario.

VOCA_08	Promedio	Límite Inferior (2,5%)	Límite Superior (97,5%)
EPTT	0, 0008508	-0, 0471625	0, 0482373
EPT	0, 08134	-0, 03842	0, 23072

En todos los casos puede notarse que tanto el EPT como el EPTT poseen valores que no pueden distinguirse significativamente de cero. Sin embargo, en las tablas 9 al 12 existen dos patrones que son importantes mencionar:

Primero, en todos los casos, el valor observado en la muestra del efecto promedio del tratamiento sobre toda la población - EPT - es positivo y además el intervalo de confianza está sesgado hacia los valores positivos. Esto último significa que aunque no se puede afirmar en las pruebas individuales que haya un efecto causal positivo hay sin embargo una probabilidad positiva de que el efecto positivo exista; más abajo se estima el valor de esta probabilidad.

Segundo, el efecto sobre toda la población tiende a ser mayor que el efecto sobre los niños que asistieron al preescolar, esto lleva a pensar que si los niños que no asistieron al preescolar lo hubiesen hecho, tendrían mejor rendimiento escolar.

Para evaluar la plausibilidad de los patrones mencionados, se adopta una posición conservadora³⁹ y se estima la probabilidad de que sean observados, si tanto un valor positivo como un valor negativo de los parámetros poblacionales EPT y EPTT fuesen igualmente posibles, 50% de probabilidad de valor menor o igual a cero, y si los resultados de las pruebas fuesen independientes.

39 Se dice que la posición es conservadora, pues el patrón apunta a que la probabilidad de que el efecto sea positivo es mayor.

Bajo la hipótesis mencionadas, se encuentra que la probabilidad⁴⁰ de observar el primer patrón; es decir, los cuatro valores muestrales, dadas por la cantidad de pruebas aplicadas en el 2008, EPT positivos, es 6,25%.

Si se adopta la hipótesis como un a priori, que luego se actualiza en función a la observación del primer patrón, empleando la fórmula de Bayes, se obtiene que la probabilidad de que el EPT sea positivo es del 83%.

Por tanto, sería improbable observar los cuatro valores EPT positivos si fuese cierto de que tanto un valor positivo como un valor negativo de los parámetros poblacionales EPT fuesen igualmente posibles; es decir, es más probable que el EPT tenga un valor positivo.

Bajo la misma hipótesis del párrafo anterior, se encuentra que la probabilidad de observar tres valores muestrales positivos y un negativo de EPTT es del 25%. Por tanto, el patrón observado tiene una probabilidad elevada de ocurrir, bajo la hipótesis de que el EPTT es nulo y los resultados independientes. Así pues, no hay evidencia que induzca a pensar que el EPTT tenga un valor positivo.

6.8 Comentarios finales

El análisis efectuado lleva a concluir que aunque las pruebas individuales no provean evidencia suficiente que permita afirmar que exista un efecto causal estadísticamente significativo del preescolar sobre el rendimiento en la Educación Escolar Básica, es razonable inferir que tal efecto pueda existir en forma de incremento en el rendimiento promedio del grupo de control. Sin embargo, es mucho menos probable que este efecto positivo se dé con el grupo de tratamiento.

⁴⁰ Utilizando la fórmula de probabilidad binomial.



La diferencia de efecto del preescolar entre el grupo de tratamiento y el grupo de control puede ser debido a que los niños que componen el grupo de tratamiento tienden a provenir de hogares con mejores condiciones de entorno que los del grupo de control. De ser así, se estaría comprobando que el preescolar tiende a mejorar el rendimiento de los niños que provienen de entornos menos favorecidos.

Debido a la importancia del tema, se recomienda en futuros estudios explorar los factores que podrían explicar la posible diferencia en el efecto entre los grupos de tratamiento y control.

Así mismo, se recomienda explorar las razones; por ejemplo, la falta de entrenamiento especializado de los maestros, la pertinencia de los materiales, la articulación entre el preescolar y la EEB que podrían explicar el bajo efecto del preescolar en el rendimiento.





Evaluación educativa: un tema central
en la agenda de la Nueva Escuela
Pública Paraguaya

86





Bibliografía

MINISTERIO DE EDUCACIÓN Y CULTURA: Plan Nacional de Educación Inicial. Convenio de cooperación Técnica. Programa de Fortalecimiento de la Educación Inicial y Preescolar MEC-BID. Asunción, Paraguay. 2002.

MINISTERIO DE EDUCACIÓN Y CULTURA: Marco Curricular de la Educación Inicial. Programa de Mejoramiento de la Educación Inicial y Preescolar MEC-BID. Asunción, Paraguay. 2005.

MINISTERIO DE EDUCACIÓN Y CULTURA: Programas de Estudios. Jardín de Infantes y Preescolar. Programa de Mejoramiento de la Educación Inicial y Preescolar MEC-BID. Asunción, Paraguay. 2006.

MINISTERIO DE EDUCACIÓN Y CULTURA: Reglamento de Educación Inicial y Preescolar. Modalidad formal y no formal. Programa Escuela Viva desde la Educación Inicial MEC-BID. Asunción, Paraguay. 2006.

MINISTERIO DE EDUCACIÓN Y CULTURA: Resultados de la Evaluación de la Efectividad del Preescolar. Programa de Mejoramiento de la Educación Inicial y Preescolar MEC-BID. Asunción, Paraguay. 2009.





Evaluación educativa: un tema central
en la agenda de la Nueva Escuela
Pública Paraguaya

88



