

EVALUACIÓN DEL IMPACTO DE LOS JUEGOS EDUCATIVOS EN LÍNEA DE INOMA

EVALUACIÓN DEL IMPACTO DE LOS JUEGOS EDUCATIVOS EN LÍNEA
DE INOMA

Preparada por Armando Chacón y Pablo Peña

Para Inoma, A.C.

Traducido por Inoma A.C.

Abril de 2013

La evaluación que se resume en este documento fue realizada con la información proporcionada a los autores por Inoma. Los autores son responsables exclusivamente de errores metodológicos. No fueron remunerados por Inoma ni por ninguna otra organización. Pueden solicitarse materiales suplementarios a los autores contactando a pablo@uchicago.edu.

I. Introducción

Inoma es una organización mexicana sin fines de lucro dedicada a mejorar los resultados educativos complementando la educación convencional con juegos educativos en línea. Este documento presenta los resultados de una evaluación del impacto de los juegos educativos en línea de Inoma en los alumnos de primaria del estado de Puebla expuestos a dichos juegos. En su primera etapa—objeto de esta evaluación—los juegos de Inoma se diseñaron con el propósito de desarrollar las habilidades matemáticas de los estudiantes medidas por el examen nacional Enlace¹.

Con el fin de evaluar el impacto de sus juegos en línea, Inoma realizó una Prueba de Control Aleatorizado (PCA) a los estudiantes de 3ero a 6to grado en un grupo de escuelas públicas en el Área Metropolitana de la Ciudad de Puebla (AMCP), en el estado de Puebla, en México. La PCA se llevó a cabo entre febrero y junio de 2012.

II. Muestra

Inoma contó con el apoyo de las autoridades locales del estado de Puebla para lograr que los estudiantes estuvieran expuestos de manera sistemática a sus juegos en línea. Para la PCA la Secretaría de Educación del estado permitió que los estudiantes de las escuelas de tratamiento jugaran los juegos de Inoma una hora a la semana durante el tiempo de clase. Las escuelas ya dedican una hora semanalmente a desarrollar las capacidades digitales de sus alumnos—aprender a usar una computadora y el Internet—en sus aulas de medios. En las escuelas de tratamiento, esa hora semanal podía dedicarse a jugar los juegos de Inoma, aunque no era obligatorio.

La Secretaría de Educación del estado consideró que había 184 escuelas primarias en el AMCP en condiciones para la PCA.² De acuerdo con los registros oficiales, todas ellas contaban con aulas de medios equipadas con computadoras conectadas a Internet. En una etapa inicial Inoma eligió aleatoriamente cuáles de las 184 escuelas de tratamiento formaría parte de la muestra de las 60 en las que se aplicaría la PCA. En una segunda etapa seleccionó aleatoriamente las 30 escuelas de tratamiento. Sin embargo, se consideró que nueve de las escuelas de tratamiento se dieron de baja de la prueba—sus representantes no asistieron a la primera sesión informativa—por lo que aleatoriamente se seleccionó el mismo número de escuelas adicionales para reemplazar a las que aparentemente se habrían dado de baja de la prueba. Como resultado, la muestra aumentó a 69 escuelas: 33 escuelas en el grupo de tratamiento y 36 en el grupo de control.

¹ El examen se aplica a los estudiantes de 3ero a 6to de primaria, de 1ero a 3ero de secundaria y en 3ero de bachillerato al final de cada año académico.

² Algunas escuelas tienen dos turnos, matutino y vespertino. Para la evaluación los turnos distintos en una misma escuela fueron considerados como escuelas distintas.

Posteriormente se hizo del conocimiento de Inoma que la mayoría de las escuelas no contaban con computadoras o con conectividad al Internet adecuadas. En consecuencia, se tuvieron que reparar las aulas de medios en las escuelas de tratamiento. Se eliminaron a nueve escuelas del grupo de tratamiento por falta de infraestructura. Así mismo, dos escuelas más del grupo en tratamiento se rehusaron a participar y se dieron de baja de la prueba. Con el propósito de mantener el equilibrio en el diseño de la PCA, algunas escuelas seleccionadas aleatoriamente fueron eliminadas del grupo de control. La muestra final fue de 44 escuelas, 22 de ellas en el grupo en tratamiento y 22 en el grupo de control.

El cuadro 1 presenta la muestra utilizada para la evaluación. La información ahí expuesta se basa en los resultados de Enlace. Solamente considera aquellas escuelas cuyos estudiantes presentaron Enlace 2012 en 4to, 5to o 6to grado, y Enlace 2011 en el grado anterior (3ero, 4to o 5to). De las 4,138 escuelas primarias en el estado de Puebla con alumnos a los que se aplicó Enlace, 184 están en el AMCP y se suponía que contaban con aulas de medios operativas: 36 se asignaron al grupo de control y 33 al grupo en tratamiento. Sin embargo, 14 escuelas del grupo de control y 9 del grupo en tratamiento fueron eliminadas de la prueba y 2 se rehusaron a participar.

Cuadro 1. Muestra de composición

	Escuelas	Estudiantes
Escuelas primarias en el estado de Puebla:	4,138	350,434
Fuera del AMCP o sin condiciones	3,954	309,978
En el AMCP y en condiciones:	184	40,456
Escuelas elegidas para experimento:	69	15,362
Control original:	36	7,167
Control final	22	4,396
Excluidas del control	14	2,771
Tratamiento originales:	33	8,195
Tratamiento finales	22	5,584
Tratamiento excluidas	9	2,274
Tratamiento dadas de baja	2	337

AMCP: área metropolitana de la ciudad de Puebla. Los turnos distintos en una misma escuela fueron considerados como escuelas distintas.

El cuadro 1 muestra el número de alumnos en 4to, 5to y 6to grados en 2012 que tomaron Enlace en 2012 y 2011—cuando estaban en 3ero, 4to y 5to grados. Más de 350,000 estudiantes tomaron Enlace, y más de 40,000 estuvieron en la AMCP. En el conjunto de escuelas consideradas en condiciones para la PCA había más de 15,000 estudiantes. En el grupo final de control hubo 4,396 estudiantes y en el grupo final de tratamiento hubo 5,584.

III. Tratamiento

Dado que las aulas de medios no eran operativas antes de la PCA, Inoma tuvo que reparar las pertenecientes a las escuelas de tratamiento. Es así que el tratamiento no se limitó a que los estudiantes estuvieran expuestos a los juegos educativos en línea; también implicó que se repararan las aulas de medios a fin de que tuvieran computadoras funcionando adecuadamente y con conectividad al Internet. En teoría, haber reparado las aulas de

medios podría haber tenido un impacto no atribuible a que los alumnos hayan jugado los juegos en línea de Inoma.

A fin de obtener resultados más rápidos, Inoma preparó versiones reducidas de las pruebas Enlace o “mini Enlaces”.³ Dichas pruebas constaban con entre 14 y 19 reactivos y se aplicaron en las escuelas tanto del grupo control como del grupo en tratamiento en febrero, abril y junio de 2012⁴. En principio, la aplicación de este tipo de pruebas basadas en exámenes Enlace anteriores por sí misma podría implicar un mejor desempeño en la prueba real—los estudiantes practican más y los maestros se hacen más conscientes del nivel de conocimiento de sus alumnos. En ambos grupos de escuelas, tanto del grupo control como del grupo de tratamiento, se aplicaron este tipo de exámenes. En consecuencia, al haber participado el grupo control en esta práctica adicional de la prueba Enlace puede considerarse que recibió de alguna manera también recibió un tratamiento.

Para los alumnos no era obligatorio jugar los juegos de Inoma. Debido a ello, únicamente se puede estimar el impacto del “intento de tratar” (*intent-to-treat*). Al grupo de tratamiento se le ofreció el tratamiento, pero no necesariamente todos los individuos lo tomaron. El impacto se promedió entre todos los alumnos a los cuales se les ofreció el tratamiento, incluidos aquellos quienes voluntariamente eligieron no tomarlo. La relevancia del intento de tratar estimado depende de la intervención real esperada en la práctica. Si la política actual no fuera obligatoria—o fuera obligatoria pero sin verificar su cumplimiento—entonces el intento de tratar estimado sería informativo respecto de la misma.

Inoma pudo monitorear qué estudiantes en las escuelas de tratamiento se registraban en su plataforma en línea. Al darse cuenta de que para mediados de abril sólo un modesto grupo de alumnos de las escuelas de tratamiento se había registrado, Inoma les dio incentivos para que aumentara su participación. Le ofreció computadoras *Notebook* al maestro y a la escuela que registrara la mayor participación. En un sentido estricto, este incentivo adicional para las escuelas menos responsivas también debe considerarse como parte del tratamiento.

Se desconoce el tipo de interacción que se dio entre maestros y estudiantes en las aulas de medios durante el tiempo que estos utilizaban los juegos en línea de Inoma. Los maestros pudieron haber dejado que los alumnos exploraran por su cuenta los juegos y decidieran el modo de jugarlos o los maestros pudieron haberlos asesorado activamente. Lo que los maestros hayan dicho o hecho durante las sesiones de juego pudo haber tenido un impacto en la actitud de los estudiantes y en su nivel de compromiso; por ejemplo, ¿las percibieron como una obligación o como una diversión? Este aspecto del tratamiento es totalmente desconocido —una caja negra.

³ Los resultados de Enlace están disponibles varios meses después de que las pruebas se aplican.

⁴ Las pruebas de Matemáticas de Enlace tuvieron entre 53 y 71 preguntas en 2012, dependiendo del grado.

IV. Métrica del impacto

La Secretaría de Educación del estado de Puebla compartió con Inoma las calificaciones de todos los alumnos de primaria correspondientes a las pruebas Enlace de 2011 y 2012. Dado que la intervención se dio entre febrero y junio de 2012, los cambios en las calificaciones de las pruebas entre 2011 y 2012 pueden utilizarse como métrica de impacto. No fue posible incluir a los alumnos de 3er grado en la evaluación porque Enlace se aplica a los estudiantes de 3 a 9 y de 12 grado. En 2011 los alumnos de 3er grado cursaban el 2do grado, por lo que no fueron evaluados.

Existen varias consideraciones respecto al uso de las calificaciones de las pruebas Enlace como métrica de impacto. Primero, el nivel de dificultad de Enlace varía año con año. En consecuencia, tanto la media como la varianza de las calificaciones de las pruebas podrían cambiar aunque las habilidades de los estudiantes se mantuvieran constantes. Con el fin de controlar las variaciones en el grado de dificultad, se definió la métrica de impacto como los cambios en las calificaciones estandarizadas de las pruebas. Es decir, a las calificaciones de las pruebas correspondientes a cada grado en cada año se les restó la media y después se dividieron por su desviación estándar. Entonces para cada alumno se calculó la diferencia entre su calificación estandarizada en 2011 y en 2012. La diferencia promedio es cero, y cualquier diferencia está definida en desviaciones estándar de Enlace.

Segundo, Enlace no es perfecta y proporciona únicamente una medida ruidosa de lo que trata de capturar. Con el propósito de que sea breve, se requiere seleccionar un pequeño conjunto de preguntas de un grupo mucho mayor de preguntas teóricamente equivalentes. Desde la perspectiva de los alumnos, existe un elemento de azar en relación a su desempeño en Enlace.

Algunos estudiantes resultan más afortunados que otros en un año particular porque estudiaron más algunos de los temas que aparecieron en la prueba. Es decir, hay ruido en Enlace que causa “regresión a la media” a lo largo de las pruebas: se espera que algunos de los estudiantes afortunados que se desempeñaron bien un año tengan un desempeño inferior el año siguiente. Lo mismo aplica al otro extremo de la distribución: se espera que algunos de los que tuvieron un mal desempeño tengan un mejor desempeño porque tuvieron mala suerte el año anterior. La estrategia empírica para evaluar el impacto de los juegos en línea de Inoma debe considerar esta regresión a la media.

Tercero, en algunos aspectos la unidad de análisis es el salón de clase—no el alumno. Los estudiantes del mismo salón comparten maestros y recursos. En consecuencia, los choques al desempeño en Enlace no son independientes entre los estudiantes en el mismo salón de clase. Con el fin de tomar en cuenta la correlación potencial de términos de error, los errores estándar en el análisis de regresión deben aglomerarse por salón de clase.

Cuarto, los contenidos de Enlace no son exactamente los mismos año con año. Es posible que la versión de un año esté más relacionada con los contenidos de los juegos de Inoma que la de otros años. Esa variación no puede ser tomada en cuenta con la evaluación de un año exclusivamente. Inoma pudo haber sido afortunado o desafortunado en 2012. Esa limitación debe tenerse en mente cuando se analicen los resultados.

Quinto, hay algunas habilidades medidas por Enlace que Inoma no pretende mejorar. Es decir, los juegos de Inoma se enfocan en un subconjunto de lo que mide Enlace. Sin embargo, utilizar la totalidad de la prueba no debe representar ningún problema. En todo caso, Enlace es una medida con ruido del conjunto de habilidades objetivo de Inoma. Dado que el ruido está en el lado izquierdo de la ecuación—la variable dependiente—no debería introducir un sesgo en las estimaciones del impacto⁵.

Las calificaciones de los mini Enlaces podrían haberse usado como la métrica de impacto de la evaluación. Sin embargo, se prefiere el uso de los resultados reales de la prueba Enlace por diversas razones. En primer lugar hubo una pérdida de casos no despreciable entre los tres mini Enlaces. Un número importante de los alumnos de una prueba a la siguiente no se identificó correctamente, lo que redujo la muestra que podría ser utilizada de forma no aleatoria.⁶ En el caso de Enlace los estudiantes están correctamente identificados por su Clave Única de Registro de Población (CURP). En segundo lugar, las calificaciones de los mini Enlaces carecen de validez externa. Inoma los creó, los aplicó y los calificó. Por el contrario, la Secretaría de Educación Federal crea y califica Enlace, y esta es aplicada por cada escuela del país. En tercer lugar, los mini Enlaces pueden no haber sido tomados muy en serio, especialmente en las escuelas del grupo de control. Dado que nada estaba en juego, no es claro que sea comparable con Enlace. Por el contrario, los resultados de Enlace sí tienen implicaciones para las escuelas—parte de la compensación de los maestros está determinada con base en los resultados de la prueba Enlace. Finalmente, los resultados utilizando los mini Enlaces no pueden interpretarse con facilidad. No se traducen directamente en unidades comúnmente empleadas. Utilizar Enlace significa que un impacto equivalente a x desviaciones estándar está bien definido. Al usar los mini Enlaces, un impacto de x desviaciones estándar no coincide con cambios en las desviaciones estándar de Enlace—lo cual es un punto de referencia para otras intervenciones.

V Estrategia empírica

Debido a los problemas que se suscitaron con el diseño de la PCA, que ocasionaron que algunas escuelas quedaran excluidas y que otras se rehusaran a participar, se decidió adoptar un enfoque cuasi experimental. En vez de comparar las medias de los grupos

⁵ Un error de medición produce un sesgo de atenuación cuando se presenta en variables del lado derecho.

⁶ Menos del 80% de los que tomaron el mini Enlace de febrero fueron identificados entre los que tomaron el mini Enlace de junio.

finales de tratamiento y de control, se optó por un análisis de regresión que incluyó a todas las escuelas en el estado de Puebla y se aplicó una técnica de diferencias en diferencias (DD). En resumen, se compara el cambio en el desempeño de los estudiantes del grupo de tratamiento con el cambio en el desempeño de los alumnos en el grupo de control, pero sin excluir a ninguna escuela.⁷

En el enfoque de DD, la primera diferencia se define entre años (el cambio en calificaciones de pruebas estandarizadas de 2011 y 2012) y la segunda se define entre grupos (de tratamiento versus de control). A fin de controlar por la regresión a la media en las calificaciones de cada estudiante, se utilizó en las regresiones un polinomio de tercer grado en la calificación estandarizada de la prueba de 2011, y se usaron todos los estudiantes del estado de Puebla. Los errores estándar se aglomeraron por salón—en algunas escuelas hay varios salones del mismo grado.

Se corrieron regresiones separadas para niños y niñas, y para cada grado disponible (4to, 5to y 6to). Como prueba de robustez, además de los resultados en Matemáticas, se exploran los resultados en Español.

Las regresiones consideran todas las escuelas del estado de Puebla, incluidas aquellas que habían sido excluidas de los grupos de tratamiento o de control, así como las dos que abandonaron el grupo de tratamiento. Las regresiones incluyen efectos fijos para cada una de esas categorías.

El enfoque de DD implica que no se necesitan controles para las características invariantes en el tiempo de los alumnos o de las escuelas. Algunos ejemplos de esas características de los estudiantes son: el nivel educativo de los padres, los recursos materiales del hogar, sus preferencias, su coeficiente intelectual. Por parte de la escuela, esas características podrían incluir: su ubicación, sus instalaciones, sus programas de estudio, la preparación de sus maestros, los incentivos a los docentes. Todos esos determinantes potenciales del desempeño desaparecen porque se restan al considerar cambios en Enlace de un año al siguiente para cada estudiante. La ecuación de la regresión es:

$$\Delta y_i = \beta_0 + \beta_1 y_i + \beta_2 y_i^2 + \beta_3 y_i^3 + \theta d_i^s + \phi d_i^{dc} + \gamma d_i^{dt} + \lambda d_i^t + \varepsilon_i \quad (1)$$

donde y es la calificación estandarizada de 2011 y Δy es el cambio en la calificación estandarizada entre 2011 y 2012.⁸ El subíndice indica al estudiante i . d^s es una *dummy* para estudiantes en escuelas en la muestra de 69 escuelas participantes en la PCA, d^{ec} es una *dummy* para los estudiantes en las escuelas que fueron excluidas del grupo de control, d^{et} es

⁷ Si la CPA fue implementada correctamente, entonces una comparación de medias y el enfoque DD identificarían el mismo parámetro. Ello no ocurriría si la CPA no fuera aplicada correctamente.

⁸ Las pruebas estandarizadas tienen una media de cero y una desviación estándar de 1. La estandarización se hace para cada prueba y para cada grado por separado.

una *dummy* para los estudiantes en las escuelas que fueron excluidas del grupo de tratamiento d^{dt} es una *dummy* para los estudiantes en las escuelas que abandonaron el grupo de tratamiento, y d^f es una *dummy* para los estudiantes en las escuelas incluidas en el grupo final de tratamiento.

El coeficiente λ es el parámetro de interés: el impacto del tratamiento. Se estima comparando los grupos finales de tratamiento y de control. Las estimaciones de φ y γ podrían interpretarse como “señuelos”. Si se implementaron correctamente la aleatorización y el tratamiento, las estimaciones de φ y γ no deberían ser estadísticamente diferentes de cero. De otro modo, podrían interpretarse como evidencia de selección en la muestra.

VI Resultados

El cuadro 2 muestra los resultados de la regresión para Matemáticas y el cuadro 3 muestra los de la regresión para Español. Ambas tablas presentan la información en forma similar. El panel superior muestra los resultados de los niños y el inferior de las niñas. Cada panel incluye tres columnas, una para cada grado analizado: 4to, 5to y 6to. Se incluyeron todos los estudiantes en el estado de Puebla que tuvieron calificaciones en 2011 y 2012 en las regresiones en los cuadros 2 y 3.⁹ La especificación de la regresión se describe en la ecuación (1). La significancia estadística de los coeficientes fue calculada utilizando errores estándar robustos aglomerados por salón de clases.¹⁰

Cuadro 2. Impacto en Matemáticas de Enlace como resultado de la exposición a los juegos de Inoma			
Variable explicativa	Variable dependiente: cambio en la calificación estandarizada del examen 2011-12		
	4th	5th	6th
Niños			
Tratamiento final	-0.004	0.180 **	0.114*
Excluidos del tratamiento	0.072	0.347 ***	0.152
Abandonaron tratamiento	-0.263**	0.089	-0.629 **
Excluidos del control	0.004	0.155 *	0.053
En la muestra PCA (control)	-0.024	-0.092 *	-0.078 *
R cuadrada	0.208	0.226	0.202
Observaciones	62,221	56,352	56,786
Conglomerados	5,683	5,385	5,350

⁹ Los resultados de los alumnos de 4 grado se comparan con los del 3 grado. Comparaciones análogas aplican para los alumnos de 5 y 6 grados.

¹⁰ Se identificaron los salones de clase utilizando la variable *grupo* en el mismo grado y escuela.

EVALUACIÓN DEL IMPACTO DE LOS JUEGOS EDUCATIVOS EN LÍNEA DE INOMA

Variable explicativa	Variable dependiente: cambio en la calificación estandarizada del examen 2011-12		
	4th	5th	6th
Niñas			
Tratamiento final			
Excluidas del tratamiento			
Abandonaron tratamiento			
Excluidas del control			
En la muestra PCA (control)			
R cuadrada			
Observaciones			
Conglomerados			

*p<.1; **p<.05; ***p<.01. Valores de p para pruebas de un solo lado usando errores estándar robustos (aglomerados por salón de clases). Todas las especificaciones incluyen como controles un polinomio de grado tres en la calificación estandarizada de 2011.

El primer resultado que destaca es que el coeficiente de la *dummy* “tratamiento final” es positivo y significativo en algunos casos. Un impacto positivo con 95% de confianza para niños de 5to grado y niñas de 6to grado. Tampoco puede ser rechazado un impacto positivo con 90% de confianza para niños de 6 grado. Las estimaciones puntuales que son significativas no son despreciables. Implican mejoras importantes: entre 0.114 y 0.180 de una desviación estándar en Enlace.

Las estimaciones para alumnos de escuelas excluidas de los grupos de control o de tratamiento y para las escuelas que abandonaron el tratamiento resultan desconcertantes. En cuatro casos son positivos y significativos con 95% de confianza, y las estimaciones puntuales varían entre 0.243 and 0.381. Es decir, en promedio los alumnos de escuelas que estaban originalmente en la PCA y que fueron excluidos posteriormente obtuvieron mejores resultados que los alumnos que estuvieron en el grupo de tratamiento final en algunos casos. Estos hallazgos ponen en duda la validez del diseño experimental.

Así mismo el cuadro 2 indica que la muestra para la PCA no es representativa de todas las escuelas en el estado de Puebla. El coeficiente de pertenecer a la muestra para la PCA es negativo y significativo con 95% de confianza para las niñas de los grados 5to y 6to.

En resumen, los resultados respecto al impacto de los juegos en línea de Inoma en Matemáticas no son concluyentes pero sí promisorios. Parece que haber recibido acceso a los juegos en línea de Inoma tuvo un impacto positivo.

El cuadro 3 muestra un análisis similar para Español. Si bien no hay una razón a priori para esperar un impacto en el desempeño en Español, los resultados muestran cierta evidencia de mejoras. La estimación del impacto es positiva y significativa con 95% de confianza para los niños de 6to grado, y con 90% de confianza para los niños de 5to grado y las niñas de 6to grado. También hay evidencia de estudiantes de 5to grado en las escuelas excluidas del

EVALUACIÓN DEL IMPACTO DE LOS JUEGOS EDUCATIVOS EN LÍNEA DE INOMA

tratamiento que tuvieron mejores resultados. El impacto en Español es sorprendente. En principio, es posible que los juegos de Inoma liberaran recursos del estudio de Matemáticas al estudio de Español.

Cuadro 3. Impacto en Español de Enlace como resultado de la exposición a los juegos de Inoma

Variable explicativa	Variable dependiente: cambio en la calificación estandarizada del examen 2011-12		
	4th	5th	6th
Niños			
Tratamiento final			
Excluidos del tratamiento			
Abandonaron tratamiento			
Excluidos del control			
En la muestra PCA (control)			
R cuadrada			
Observaciones			
Conglomerados			
Niñas			
Tratamiento final			
Excluidas del tratamiento			
Abandonaron tratamiento			
Excluidas del control			
En la muestra PCA (control)			
R cuadrada			
Observaciones			
Conglomerados			
*p<.1; **p<.05; ***p<.01. Valores de p para pruebas de un solo lado usando errores estándar robustos (aglomerados por salón de clases). Todas las especificaciones incluyen como controles un polinomio de grado tres en la calificación estandarizada de 2011.			

VII. Advertencias y preocupaciones

Existen varias advertencias y preocupaciones acerca de los resultados presentados en los cuadros 2 y 3. La primera y más importante es que no queda claro cuál fue la intensidad del tratamiento, aunque probablemente fue modesta. El tiempo de exposición fue corto—unas pocas sesiones con solo un puñado de juegos. La plataforma de Inoma fue una tecnología nueva y podría existir una curva de aprendizaje para los maestros que tratan de promoverla. Muy probablemente no hubo tiempo suficiente para explotarla apropiadamente.

La segunda es que no queda claro cómo se percibió todo el experimento por parte de los maestros y directores de las escuelas. ¿Lo percibieron como una política pública de cuyos resultados ellos serían considerados responsables o simplemente como un experimento académico sin consecuencias en el cual ellos jugaban un papel pasivo si no indiferente? No es claro qué expectativas tenían en término de las consecuencias de la evaluación. Por ejemplo, ¿continuarían teniendo la nueva responsabilidad de utilizar los juegos en línea de Inoma únicamente si los alumnos tenían un buen desempeño? ¿Se percibió como una estrategia para sustituir los métodos tradicionales que pondría en peligro el estatus de los

maestros actuales? Esos podrían ser factores que afecten la interpretación de los resultados como una guía de política.

La tercera es que el contacto probable entre los directores de las escuelas de los grupos de tratamiento y de control hace dudar de la pureza del experimento. Algunos directores de escuela eran conscientes de que estaban participando en “algún tipo de competencia”. Esa percepción pudo haber afectado su comportamiento, pero no es claro en qué sentido.

VIII. Discusión de los hallazgos

La comparación de los grupos finales de tratamiento y de control por sí misma da evidencia de un efecto positivo de la exposición a los juegos en línea de Inoma. A pesar del corto tiempo de exposición al tratamiento y su estatus no obligatorio, no puede rechazarse un impacto positivo con 95% de confianza para algunos estudiantes. Las estimaciones puntuales en esos casos no son despreciables: 0.114 y 0.180 desviaciones estándar en las calificaciones de Matemáticas.

Una amplia gama de intervenciones han sido estudiadas empíricamente en la literatura académica en términos de su impacto en los resultados educativos. Más allá de las intervenciones que afectan la disponibilidad de escritorios, el conocimiento por parte de los maestros de las asignaturas que enseñan y del ausentismo de los maestros, no hay mayor guía empírica acerca de lo que funciona en la realidad. De hecho, entre mayor es el rigor metodológico que tienen los estudios, lo más probable es que rechacen un impacto positivo de una intervención bajo estudio¹¹. En ese contexto la evidencia sobre los juegos en línea de Inoma es sumamente alentadora.

Inoma dio un paso gigantesco en la dirección correcta al intentar realizar una evaluación experimental del impacto de sus juegos en línea. Los resultados aquí presentados indican que no puede descartarse la existencia de importantes logros académicos resultado de jugar sus juegos en línea.

¹¹Ver “*School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010*” — de Paul W. Glewwe, Eric A. Hanushek, Sarah D. Humpage, y Renato Ravina.