Informe final publicable de proyecto

Prueba nacional y estandarizada para la evaluación temprana de competencias en lengua y matemáticas: una clave para el acompañamiento de trayectorias educativas

Código de proyecto ANII: FSED_2_2022_1_174584

Fecha de cierre de proyecto: 15/12/2024

ZUGARRAMURDI GARCIA, Camila (Responsable Técnico - Científico)

MAICHE MARINI, Alejandro (Co-Responsable Técnico-Científico)

NOGUEIRA GONZALEZ, Jenny Natali (Investigador)

CABANA FAJARDO, Álvaro Jesús (Investigador)

DE LEÓN PIÑEIRO, Dinorah (Investigador)

LUZARDO VERDE, Mario Javier (Investigador)

MARCONI FERRARI, Cecilia (Investigador)

RONQUI LEITES, Valentina (Investigador)

SOLEDAZ ASSIS, María (Investigador)

VALLE LISBOA ASURABARRENA, Juan Carlos (Investigador)

UNIVERSIDAD DE LA REPÚBLICA. CENTRO INTERDISCIPLINARIO EN COGNICIÓN PARA LA ENSEÑANZA Y EL APRENDIZAJE (Institución Proponente) \\
SOCIEDADES CIENTÍFICO-TECNOLÓGICAS. SOCIEDAD URUGUAYA DE CIENCIAS COGNITIVAS Y DEL COMPORTAMIENTO \\
SOCIEDADES CIENTÍFICO-TECNOLÓGICAS. SOCIEDAD URUGUAYA DE CIENCIAS COGNITIVAS Y DEL COMPORTAMIENTO

Resumen del proyecto

Las trayectorias educativas muestran comportamientos persistentes. Los alumnos que comienzan su vida escolar con buenos logros de aprendizaje, culminan su educación con altos niveles de desempeño. En contraposición, aquellos alumnos que comienzan su vida escolar mostrando perfiles de riesgo educativo, se enfrentan a dificultades persistentes a lo largo de su trayectoria, que en muchos casos resultan en la desvinculación del sistema educativo. Sin embargo, muchas de las habilidades y competencias que se ponen en juego durante la vida escolar comienzan a desarrollarse tempranamente, durante la educación inicial. Dentro de estas habilidades y competencias, las del dominio de lengua y matemática son especialmente fundacionales para el futuro desempeño académico. Cuando estas son evaluadas de forma temprana y exhaustiva, permiten implementar intervenciones focalizadas y oportunas que, en muchos casos, son capaces de prevenir trayectorias educativas de riesgo. El presente proyecto tiene como objetivo brindar al sistema educativo una prueba nacional y estandarizada de competencias tempranas en lengua y matemáticas orientadas a niños y niñas de 5 y 6 años. Esta posee formato digital de videojuego, es autoadministrada y puede aplicarse grupalmente en contexto de aula. Asimismo, brinda reportes de resultados de cada estudiante y a nivel grupal. En consecuencia, puede ser utilizada como una evaluación formativa que funcione como un insumo para docentes, aportando información valiosa sobre competencias tempranas en estos dominios. Esta permite guiar el diseño y la implementación de intervenciones oportunas que favorezcan el acompañamiento de trayectorias educativas, promoviendo la inclusión y reduciendo el abandono.

Ciencias Sociales / Ciencias de la Educación / Educación General (incluye entrenamiento, pedagogía y didáctica) / Ciencias cognitivas y evaluación educativa

Palabras clave: evaluación estandarizada / lectura / matemáticas /

Antecedentes, problema de investigación, objetivos y justificación.

La evaluación de los aprendizajes es un área central de la práctica educativa. Las evaluaciones pueden distinguirse según su objetivo en sumativas o formativas. Las evaluaciones sumativas están orientadas a brindar información al sistema educativo sobre los niveles de desempeño de niños y niñas en relación a alguna escala estandarizada. Las evaluaciones formativas buscan brindar información a docentes sobre las fortalezas y debilidades en los logros de aprendizaje, de modo tal que esta información pueda ser utilizada como guía para la planificación de su práctica docente (Bennett, 2015). En este sentido, las evaluaciones formativas son clave para el acompañamiento de trayectorias educativas ya que brindan información al docente de los logros de aprendizaje de los alumnos, la cual permite identificar las necesidades individuales de cada uno de ellos.

Para que las evaluaciones formativas cumplan el rol de acompañar las trayectorias educativas, es fundamental que evalúen efectivamente aquello que se proponen evaluar. Esto se conoce, en el ámbito de la Psicología, como la validez de un instrumento de medida. En el ámbito educativo, el conocimiento que los docentes tienen sobre las características de la evaluación educativa y los requisitos para que esta sea válida se conoce como alfabetización en evaluación (Xu & Brown, 2016). Si bien no hay datos disponibles sobre la formación docente en Uruguay en este dominio en particular, en el mundo es escasa o nula (Popham, 2009), y cabe suponer que el caso de Uruguay no es distinto. A pesar de esto, los docentes suelen identificar correctamente el nivel de desempeño de los niños en relación con su grupo, en disciplinas relevantes como la lengua o la matemática (ver, por ejemplo, Cabell et al., 2009). Sin embargo, la evaluación que los docentes hacen de los logros de sus alumnos podría describirse como "a grandes rasgos". Esto es, son capaces de identificar correctamente quiénes están por encima de la media del grupo, y quiénes están por debajo de la media, y en especial de identificar los casos extremos (Kilday et al., 2012), pero no de brindar una descripción más fina de los niveles de desempeño o logros en el aprendizaje. A esto se suman los sesgos subyacentes a cualquier evaluación subjetiva. En particular, en la evaluación que los docentes hacen de sus alumnos interfieren sesgos que cuestionan en algunos casos la validez de los resultados. Por ejemplo, un estudio en que 210 docentes evaluaron a 711 niños y niñas a su cargo en nivel inicial mostró que las evaluaciones que los docentes hacían de sus alumnos estaban fuertemente influenciadas por los años de experiencia, la raza/etnicidad del docente, el tamaño del grupo y la propia percepción de autoeficacia del docente (Mashburn et al., 2006). En suma, las evidencias apuntan tanto a la validez de la evaluación de los docentes en términos de niveles de desempeño, a grandes rasgos, en relación con su grupo, como a los sesgos inherentes a esta misma evaluación y a la necesidad de contar con herramientas de evaluación más precisas. Por lo tanto, es deseable contar con sistemas de evaluación formativa que complementen la evaluación que los propios docentes hacen. Una evaluación formativa estandarizada y validada brindará una fuente adicional de información que provea un perfil más exhaustivo de logros de aprendizaje de niños y niñas. Además, una evaluación formativa estandarizada independiente del docente puede ser un aliado para sustentar sus apreciaciones, presentarlas y fundamentarlas ante directores y familias.

El sistema educativo uruguayo se destaca en la región por contar con pruebas sistemáticas y estandarizadas del desarrollo infantil y los aprendizajes. En este sentido, existe la prueba INDI que se aplica en estudiantes de nivel inicial y que evalúa el nivel de desarrollo en las áreas cognitiva, motora y socioemocional. Más adelante en su escolarización, los estudiantes son evaluados en sus aprendizajes en lectura, escritura y oralidad mediante la prueba LEO en segundo de educación primaria. ANEP dispone también de las pruebas SEA a partir de tercer año, con carácter formativo. Adicionalmente, en tercero y sexto de educación primaria, el Instituto Nacional de Evaluación Educativa (INEEd) aplica las pruebas ARISTAS que evalúa aprendizajes en matemática y lengua. Sin embargo, no existe al momento ninguna prueba a nivel nacional que evalúe de modo formativo las habilidades tempranas en lengua y matemática.

Cada uno de los instrumentos descritos anteriormente cumplen un rol valioso en brindar información sobre el estado del sistema educativo y de los logros de aprendizaje y desempeño de niños y niñas que asisten a la educación pública. Sin embargo, para que esta evaluación cumpla un rol en el acompañamiento de trayectorias educativas, debe tener algunas características particulares que no todos los instrumentos descritos anteriormente comparten. En especial, debe ser posible de ser aplicada en contexto de aula, por el maestro, y proporcionar información sobre el desempeño del niño o niña. De los instrumentos descritos anteriormente, el INDI y las pruebas LEO cumplen con estos criterios. Sin embargo, el alcance de estos instrumentos posee dos limitaciones que se buscan complementar. Por una parte, el INDI es un instrumento de "amplio espectro" en el desarrollo infantil, y por lo tanto no busca ahondar en dominios específicos como la matemática y la lengua, sino que las evalúa someramente (posee 7 ítems de lenguaje y 6 ítems de matemática). Sin embargo, las competencias en lengua y matemática son fundacionales del futuro desempeño personal y académico (Arnold et al., 2005; Jordan et al., 2009). Por otra parte, las pruebas LEO, si bien se centran en el dominio de la lengua, se aplican recién en segundo año, cuando algunas de las dificultades en el aprendizaje de la lectoescritura ya se están instalando, y las intervenciones de remediación son menos eficaces y efectivas (Wanzek & Vaughn, 2007).

En consecuencia, el presente trabajo tiene como objetivo brindar al sistema educativo una prueba nacional y estandarizada de evaluación digital, individual, gratuita y aplicable en contexto de aula, de competencias tempranas en lengua y matemáticas en niños de 5 y 6 años. La evaluación temprana es crucial para prevenir trayectorias de riesgo educativo, especialmente en las áreas de lengua y matemática, ya que son fundacionales en los aprendizajes futuros y determinan el futuro desarrollo personal y académico (Arnold et al., 2005; Jordan et al., 2009).

Las trayectorias educativas son persistentes. Múltiples estudios muestran que cuando los niños y niñas obtienen buenas puntuaciones al inicio de su vida escolar, su desempeño suele mantenerse a lo largo de su trayectoria educativa (Cardozo et al., 2022; Jordan et al., 2009; McCordle et al., 2001). Resultados análogos se observan cuando se miden las competencias tempranas en lengua y matemática; estas tienen un alto valor predictivo sobre el desempeño académico futuro en general, y sobre los aprendizajes futuros en lengua y matemática en particular (Duncan et al., 2007). Las competencias en lengua y matemática son centrales en los primeros años de escolarización. Según datos aún no publicados de nuestro equipo de investigación, al final de Primer año de educación primaria, el 55% de la varianza en las notas de promoción de los estudiantes es explicada por el desempeño en pruebas de lectura de palabras, y el 35% de la varianza es explicada por el desempeño en pruebas de aritmética. Estos valores reflejan la importancia que los docentes dan, explícita o implícitamente, a estas competencias durante el inicio de la Educación Primaria.

Si bien solemos pensar los aprendizajes en lengua –en particular, en lectoescritura– y matemática como aprendizajes formales que ocurren a partir del inicio de Primaria, las bases de estos aprendizajes comienzan a asentarse en años anteriores, durante la Educación Inicial, e incluso antes. En lengua, estas incluyen la conciencia fonológica, el conocimiento de letras y la memoria verbal de corto plazo. La conciencia fonológica refiere a la capacidad de identificar y manipular los sonidos de la lengua y se mide en tareas como señalar cuántos sonidos conforman una palabra. El conocimiento de las letras refiere a conocer las reglas de conversión de grafemas a fonemas y de fonemas a grafemas. Por ejemplo, que a la letra *c* le corresponden dos sonidos: /*s*/ y /*k*/, y que al sonido /*s*/ le corresponden las letras *c*, *s* y *z*, dependiendo de la palabra. La memoria verbal de corto plazo es aquella que involucra retener en memoria por períodos breves de tiempo una secuencia de elementos verbales, por ejemplo un número de teléfono o una lista de compras (National Reading Panel, 2000). Entre las habilidades matemáticas tempranas son centrales el conocimiento sobre cantidades no simbólicas (capacidad para aproximar), el conteo, los principios de cardinalidad y correspondencia, el mapeo símbolo-palabra-cantidad, y el cálculo entendido como la habilidad de combinar o separar conjuntos, así como otras habilidades más de tipo espacial como por ejemplo el reconocimiento de patrones o formas (Rittle-Johnson et al., 2017). Dado que existen habilidades y competencias tempranas en lengua y matemáticas, que estas sientan las bases del desempeño futuro, y que pueden evaluarse en los niveles de educación inicial, no es necesario esperar a que efectivamente los niños y niñas muestren signos de dificultades en el aprendizaje de lengua y matemáticas para implementar intervenciones oportunas. Abundante literatura señala que las intervenciones oportunas son más eficaces y eficientes que los programas de remediación (ver, por ejemplo, Fletcher & Vaughn, 2009). Por lo tanto, el abordaje más sensato para reducir las trayectorias educativas de riesgo es la prevención, implementando intervenciones focalizadas lo antes posible, desde la Educación Inicial (Lyytinen et al., 2015; Ozernov-Palchik & Gaab, 2016; Raudenbush et al., 2020).

Lexiland y PUMA han sido especialmente diseñadas para evaluar estas habilidades y competencias tempranas y brindar información exhaustiva y de calidad para el diseño de intervenciones focalizadas y oportunas.

Por lo tanto, el objetivo de este proyecto es el de disponer de una evaluación estandarizada de competencias tempranas en matemática y lengua. Esta está destinada a una población (alumnos de Nivel 5 y Primero) para la cual no existen pruebas estandarizadas de estas competencias en nuestro sistema educativo. Además, se realiza en formato digital de forma grupal, lo que la hace aplicable al contexto de aula y genera un informe de resultados a nivel grupal e individual. Por estas razones, estas herramientas tienen el potencial y las características para convertirse en una herramienta de apoyo a la tarea docente, complementando las evaluaciones disponibles actualmente.

Metodología/Diseño del estudio

Muestra

A los efectos de calcular el tamaño muestral necesario se recurrió a la estimación del error de muestreo en la estimación de los percentiles. Se permitió un error del 5% en el valor del cuantil. Con este valor se obtiene que el error absoluto a tener en el percentil 10 (o 90) es de 0.05 y el tamaño muestral es de 752 alumnos en cada nivel (Nivel Inicial 5 y Primero). A los efectos de obtener una muestra nacional se utilizó en primera instancia un muestreo de los departamentos. Para ello se seleccionó Montevideo y se calculó la proporción de la población de alumnos de Primer año y Nivel Inicial 5 para el resto. A partir de esto, se le asignaron probabilidades de selección a cada departamento y se seleccionaron de forma aleatoria pero ponderada según sus probabilidades a los departamentos de Canelones, Colonia, San José, Salto y Soriano.

Para seleccionar las escuelas dentro de cada departamento se realizó un muestreo de conglomerados. Para saber cuántos grupos por departamento seleccionar se usó la asignación proporcional y luego la estratificación por nivel socioeconómico. La cantidad de grupos se seleccionó a partir de los tamaños promedio de los grupos de cada nivel. Solo se tuvieron en cuenta grupos con 10 o más alumnos. El tamaño promedio de los grupos con 10 o más alumnos es de 21 niños en Nivel Inicial 5, y 20 niños en Primer año.

Adicionalmente, los grupos de cada departamento se asignaron usando muestreo estratificado proporcional, estratificando por nivel socioeconómico. Se usaron los 5 estratos que considera la ANEP para la clasificación de escuelas. A partir de la clasificación en cada estrato socioeconómico y departamento se sortearon escuelas con probabilidades proporcionales a la cantidad de alumnos de Nivel Inicial 5 y Primer año. Se sortearon las escuelas necesarias para completar los grupos requeridos.

Para cada departamento y cada quintil se seleccionaron al azar las escuelas. Por razones logísticas, se incluyeron en la muestra solamente escuelas ubicadas en capitales departamentales y en localidades con más de 5000 habitantes. Para Nivel Inicial 5 se seleccionaron por sorteo 29 grupos de 28 escuelas distintas, ya que en una de ellas se seleccionaron (al azar) dos grupos. Para Primer año se seleccionaron 33 grupos de 32 escuelas distintas, ya que en una de ellas se seleccionaron (al azar) dos grupos. Las escuelas sorteadas se distribuyen por quintil socioeconómico de la siguiente manera: 6 escuelas de quintil 1, 6 de quintil 2, 6 de quintil 3, 5 de quintil 4 y 9 de quintil 5. La distribución no uniforme del número de escuelas (y alumnos) por quintil replica la distribución no uniforme existente en la población, de acuerdo a los datos disponibles en la base GURI.

A partir de esta selección inicial, en la etapa de coordinación con las escuelas se realizaron algunas modificaciones. Debido a cuestiones de disponibilidad y organización de los centros educativos se reemplazaron algunas escuelas por otras del mismo departamento y quintil.

Finalmente, se evaluaron, durante agosto, setiembre y octubre del 2023, 28 grupos de Nivel Inicial 5 de 26 escuelas y 34 grupos de Primer año de 30 escuelas, 31 escuelas en total. Estas se distribuyeron según quintil socioeconómico de la siguiente manera: 6 escuelas de quintil 1, 6 de quintil 2, 6 de quintil 3, 5 de quintil 4 y 8 de quintil 5. Finalmente, la muestra quedó compuesta por 1380 alumnos de Primer año y Nivel Inicial 5.

De estos, 269 alumnos (19%) estuvieron ausentes el día de la evaluación y no realizaron la evaluación. La media de edad de los participantes fue de 7 años y 2 meses para Primer año, y 6 años y 2 meses para Nivel Inicial 5. En cuanto a la distribución por departamentos, la mayor parte de los participantes, el 52.66%, fueron alumnos de Montevideo, el 19,3 % de Canelones, 10,4 % de Salto, 7,0 % de San José, 5,6 % de Colonia y el 5,0 % de Soriano. Estos porcentajes son proporcionales a la población escolar por departamento.

Durante el año 2024 se realizó un seguimiento de la muestra en 5 escuelas de las evaluadas durante el 2023. Fueron seleccionadas, por conveniencia, una escuela por quintil socioeconómico, tres de Montevideo y dos de Canelones. La muestra estuvo compuesta por 353 alumnos del primer y segundo año de Educación Primaria.

Procedimiento

En el año 2023 se aplicaron las pruebas Lexiland y PUMA en grupos de Nivel 5 de Educación Inicial y de primero de Educación Primaria.

La prueba PUMa (Prueba Uruguaya de Matemática), en su versión breve diseñada para este proyecto, consta de nueve subpruebas que miden las siguientes competencias: sistema numérico aproximado, serie numérica progresiva, serie numérica regresiva, conteo, transcodificación, rotación mental, patrones, cálculo mental, composición y descomposición. Estas subpruebas se clasifican entre las que miden componentes simbólicos y no simbólicos de la matemática.

Lexiland es una prueba de identificación temprana del riesgo lector, evalúa reconocimiento de palabras y habilidades pre lectoras en tres dimensiones: conciencia fonológica, conocimiento de las letras y memoria no verbal.

Ambas pruebas son digitales y gamificadas, donde se sigue una narrativa que va introduciendo las distintas subpruebas, a través de personajes y desafíos. Las consignas son explicadas oralmente, se presentan ejemplos e intentos de práctica. Por lo tanto, son pruebas autoadministradas, donde cada niño juega de forma autónoma con un dispositivo.

La validez de criterio de Lexiland ha sido corroborada a través de un estudio previo, longitudinal, de aproximadamente 600 niños en que se estableció la relación entre habilidades prelectoras (conciencia fonológica, conocimiento de letras y memoria no verbal) y el futuro desempeño lector. A partir de los datos obtenidos se construyó un modelo estadístico que, a partir de los datos medidos durante Nivel Inicial 5, predice el desempeño lector al final de Primer año en dos categorías: con riesgo lector y sin riesgo lector. La precisión en la clasificación del modelo se evaluó a través de validación cruzada, entrenando al modelo con el 70% de los datos y testeándolo en el 30% de los datos restantes. A través de este método se estimó la sensibilidad y la especificidad de la prueba. Los valores obtenidos para sensibilidad y especificidad son del 90 y 76 por ciento, respectivamente. Esto quiere decir que Lexiland identifica correctamente a 9 de cada 10 niños que mostrarán riesgo lector, e identifica correctamente a casi 8 de cada 10 niños que no mostrarán riesgo lector. Estos valores reflejan un alto grado de precisión en la clasificación que el modelo logra, equivalentes a los obtenidos en evaluaciones individuales llevadas a cabo por técnicos capacitados (Zugarramurdi, Fernández, Lallier, Carreiras, et al., 2022).

La validez de constructo de la prueba PUMa ha sido demostrada en relación al Test de Competencia Matemática Básica TEMA-3 (Ginsburg et al., 2007). Para una submuestra de 184 participantes, se encontró una correlación de Pearson entre PUMa y TEMA-3 de $r = .77$ ($p < .01$). En relación a su validez interna, se obtuvo un valor de alfa de Cronbach de 0.82. Estos resultados sugieren que PUMa efectivamente es capaz de medir competencias matemáticas, incluso cuando se aplica grupalmente en contexto de aula, y que sus mediciones son estables (Marconi, et al., en revisión).

Para la aplicación de las pruebas se proporcionó una Tablet y auriculares para cada alumno. Se comenzó con una explicación oral por parte de los asistentes de investigación donde se contó la narrativa, se presentó a los personajes y se explicaron aspectos prácticos. Posteriormente, cada alumno jugó de forma individual. En cada grupo, una mitad comenzó jugando a Lexiland y la otra mitad a PUMa, y luego se alternaron. Durante la aplicación se realizaron intervenciones mínimas para asegurar la autonomía en la resolución de tareas.

Durante el año 2024, en una submuestra de cinco escuelas, entre los meses de abril y mayo, se aplicó la prueba PUMa de forma grupal y la prueba TEMA 3 de forma individual. Posteriormente, entre setiembre y noviembre se aplicó la prueba Lexiland a nivel grupal y de forma individual se evaluó lectura de palabras, comprensión de enunciados y fluidez lectora.

Resultados, análisis y discusión

En la fase de baremación de los instrumentos se establecieron percentiles para interpretar los resultados obtenidos. Los percentiles permiten entender la posición relativa de un individuo dentro de un grupo de referencia. En este caso, se determinaron los valores correspondientes al percentil 20 y al percentil 80. El percentil 20 indica que el 20% de la población evaluada tiene puntajes iguales o inferiores a dicho valor, lo que significa que estos niños están en el nivel más bajo de rendimiento. Por otro lado, el percentil 80 señala que el 80% de los niños tienen puntajes iguales o inferiores a este valor, ubicando al 20% restante en el nivel más alto de rendimiento. Esta información es crucial para identificar a los niños que requieren apoyo adicional en el desarrollo de sus habilidades matemáticas o de lectura y aquellos que muestran un desempeño sobresaliente, facilitando así la toma de decisiones educativas y pedagógicas informadas. En las Tabla 1 y Tabla 2 (Ver Anexo) se presentan los estadísticos descriptivos de los puntajes para cada una de las pruebas de Lexiland para Nivel 5 y Primero, respectivamente. En la Tabla 3 y Tabla 4 (ver Anexo) se presentan los estadísticos descriptivos de los puntajes para cada una de las pruebas de PUMa para Nivel 5 y Primero.

Se realizó un análisis de consistencia interna utilizando el coeficiente Alfa de Cronbach para evaluar la fiabilidad del instrumento. En la Tabla 5 (ver Anexo) se presentan los resultados para cada una de las pruebas de Lexiland y en la Tabla 6 (ver Anexo) para las pruebas de PUMa. En ambas pruebas los resultados revelan un alto nivel de consistencia entre los ítems de las diferentes tareas, indicando una alta correlación entre ellos.

Se evaluó la validez estructural de ambos instrumentos mediante un Análisis Factorial Confirmatorio y se analizan los valores de RMSEA y de CFI. El valor de RMSEA (Root Mean Square Error of Approximation) evalúa cuán bien se ajusta un modelo especificado a los datos observados. Los valores entre 0.05 y 0.08 son considerados aceptables. El CFI (Comparative Fit Index) compara el ajuste de un modelo propuesto con el ajuste de un modelo nulo que asume que no hay relaciones entre las variables observadas. Valores de CFI cercanos a 1 indican un buen ajuste del modelo. En el caso de PUMa, se determinaron a priori dos factores: pruebas simbólicas y pruebas no simbólicas. Se obtuvo un valor de RMSEA de 0.075 y un CFI de 0.97. Esto es un buen ajuste del modelo. Para el caso de Lexiland, se determinaron tres factores a priori: las pruebas de conciencia fonológica, las pruebas de conocimiento de las letras y la prueba de memoria. Se obtuvo un RMSEA de 0.062 y un valor de CFI 0.96, lo que muestra un ajuste de aceptable a bueno.

Discusión

Los resultados expuestos hasta aquí, muestran que ambos instrumentos resultan fiables y válidos para lo que buscan evaluar. Por lo tanto, ambas pruebas están en condiciones de ser utilizadas en el ámbito educativo para brindar evaluaciones de calidad. En cuanto a su aplicabilidad en las aulas, en este proyecto se realizaron instancias de formación con docentes de primaria, directores e inspectores, así como instancias de devolución de resultados. En estas se vio interés por parte de los docentes, quienes consideraron que ambas herramientas podían ser útiles para sus prácticas. Asimismo, las instancias de devolución de resultados fueron instancias muy ricas de trabajo con docentes donde se pudo realizar una interpretación conjunta de los datos. Este trabajo con los datos resulta de sumo interés tanto para docentes como para investigadores. En esta retroalimentación, los docentes pueden conocer más en profundidad el desempeño de sus estudiantes, así como tomar decisiones y objetivar sus percepciones. Al mismo tiempo, por el otro lado, supone para los desarrolladores de la herramienta una instancia de análisis de esta, sus posibilidades y áreas de mejora.

Por este motivo, hacia el final del presente proyecto, se puso en cuestionamiento la intención original de que las herramientas generen un informe automático para docentes, que pueda estar descontextualizado de los cometidos de las herramientas. Se definió que la mejor opción es realizar encuentros de devolución (presenciales o virtuales) con los equipos de las escuelas, ya que se entiende que este intercambio es parte de la formación a docentes en esta área y permite contextualizar los resultados y garantizar una interpretación válida.

Justamente, otro de los objetivos de este proyecto fue el de capacitar al sistema educativo en el uso de las aplicaciones. En este sentido, se realizaron capacitaciones con docentes de distintos departamentos del país, tanto virtuales como presenciales. Más allá de la formación acerca del conocimiento y uso de las aplicaciones, estas fueron instancias donde se trabajó el marco teórico que sustenta ambas herramientas. Esto puede dar posibilidades de los docentes de utilizarlas mejor, interpretar mejor los resultados e incorporar prácticas en el aula consistentes con estos marcos teóricos.

Por otra parte, este estudio permitió confirmar que estas herramientas pueden ser aplicadas en contextos ecológicos. Durante el trabajo de campo realizado se aplicaron en aulas reales, adaptándose a los horarios y rutinas de cada centro educativo. En este sentido, se vio que la evaluación conjunta de un grupo entero de alumnos en su aula es posible y arroja resultados confiables. Entre las dos pruebas el tiempo utilizado fue, en promedio, de unos 40 minutos. Esto es de suma importancia para que sean utilizadas en el sistema educativo, sin requerir cambios de dinámica, horarios o rutinas y sin necesidad del abordaje uno a uno del docente. Esto hace que se requiera de un solo evaluador, que puede ser el docente, y que las evaluaciones a nivel grupal sean más viables. Sumado a esto, el procesamiento de los datos no está a cargo del docente, sino que el hecho de que sean evaluaciones digitales hace que esto se realice de forma automática.

Para sumar a esto, se están analizando actualmente datos del trabajo de campo 2024 donde se evaluó a una submuestra con ambas evaluaciones digitales y a su vez con evaluaciones individuales que miden lo mismo que las digitales. De esta manera se obtendrá aún más evidencia sobre la validez de las herramientas, así como datos longitudinales que permitan ahondar en la caracterización de las trayectorias educativas en estas dos áreas. En el caso de PUMa, el equipo se plantea la posibilidad de identificar las competencias matemáticas iniciales que mejor predicen el avance de los niños entre inicial y primero y entre primero y segundo. Este enfoque resulta novedoso ya que no se fija en el desempeño sino en el grado de avance que resulta sustancial para el proceso formativo en los años iniciales: potenciar las posibilidades de mejora de los resultados de cada niño en relación a su punto de partida.

Este trabajo se encuentra aún en proceso y se prevé enviar el manuscrito a publicación en marzo de 2025.

Conclusiones y recomendaciones

Como conclusión de este proyecto se cuenta con dos herramientas para el sistema educativo uruguayo disponibles en sus plataformas web: <https://puma.cicea.uy> y <https://lexiland.com.uy>. Estas son herramientas fiables y válidas para evaluar competencias tempranas en lengua y matemática. Es decir, estas herramientas pueden acompañar las trayectorias educativas desde el inicio de estos aprendizajes básicos de la escolarización, y por lo tanto, pueden ayudar a diseñar acciones para prevenir dificultades futuras. Como recomendaciones se considera necesario continuar el intercambio con el sistema educativo, no solo en el uso de estas herramientas, sino en la evaluación de aprendizajes en general y en los marcos teóricos que lo sustentan. Creemos que la evaluación temprana de desempeños es una herramienta clave para el acompañamiento de trayectorias educativas, y que este es un esfuerzo viable y necesario para prevenir trayectorias de riesgo. Los resultados del presente proyecto dan cuenta de esta realidad y son un impulso para avanzar por este camino.

Productos derivados del proyecto

Tipo de producto	Título	Autores	Identificadores	URI en repositorio de Silo	Estado
Artículo científico	An early numeracy digital brief assessment: parametric and non-parametric IRT models	Marconi, Cecilia; De León, Dinorah; Luzardo, Mario; Maiche, Alejandro		https://hdl.handle.net/20.500.12381/3905	Finalizado
Parte de libro	Prueba Uruguay de Matemática. en Instrumentos de evaluación psicológica en el Uruguay: sistematización de las evidencias	De Leon, Dinorah; Araujo, Pablo; y Maiche, Alejandro			En proceso
Parte de libro	Lexiland: un instrumento digital para la identificación del riesgo lector en el ámbito escolar	Ronqui, Valentina; Valle-Lisboa, Juan; Zugarramurdi, Camila		https://hdl.handle.net/20.500.12381/3900	Finalizado
Póster	Enhancing early math skills through digitally adapted social learning in the classroom	Maiche, Alejandro		https://hdl.handle.net/20.500.12381/3904	Finalizado
Presentación en evento	De Lexiland a Anacahuíta: diseño y validación de herramientas para evaluar el desempeño lector y estimular la alfabetización inicial	Soledad Assis, Valentina; Ronqui, Camila; Zugarramurdi, Juan Valle; Lisboa		https://hdl.handle.net/20.500.12381/3908	Finalizado
Presentación en evento	Aportes de las ciencias cognitivas a una educación sin exclusiones	Camila Zugarramurdi		https://hdl.handle.net/20.500.12381/3907	Finalizado
Nota de prensa	Videojuegos para detección temprana de problemas de lectura	Juan Valle Lisboa	https://enperspectiva.uy/en-perspectiva-programa/entrevistas/videojuegos-para-deteccion-temprana-de-problemas-de-lectura-con-juan-carlos-valle-neurocientifico-y-uno-de-los-desarrolladores-de-la-app-lexiland/		Finalizado

Tipo de producto	Título	Autores	Identificadores	URI en repositorio de Silo	Estado
Nota de prensa	Crearon videojuegos para niños que permiten detectar futuros problemas de lectura antes de ir a la escuela	Camila Zugarramurdi, Juan Valle Lisboa	https://www.elpais.com.uy/informacion/educacion/crearon-videojuegos-para-ninos-que-permiten-detectar-futuros-problemas-de-lectura-antes-de-ir-a-la-escuela		Finalizado
Presentación en evento	Cognición matemática temprana: el surgimiento de la capacidad simbólica	Alejandro Maiche, Paola Garcia, Marcela Mena		https://hdl.handle.net/20.500.12381/3906	Finalizado
Nota de prensa	PUMa	Alejandro Maiche	https://drive.google.com/file/d/1vqyu9DLSGf8PFYaUIRjPWIhaDeZseyPx/view		Finalizado
Nota de prensa	Alejandro Maiche Aprendices La Red 2024	Alejandro Maiche	https://aprendices.edu.uy/t4/e7/alejandro-maiche		Finalizado
Nota de prensa	Matemáticas tempranas para forjar sociedades más justas	Alejandro Maiche	https://profuturo.education/observatorio/conversaciones/matematicas-tempranas-para-forjar-sociedades-mas-justas/		Finalizado

Referencias bibliográficas

- Alves, M. V., Ekuni, R., Hermida, M. J., & Valle-Lisboa, J. (2022). The Importance of Bringing the Latin American Perspective to Cognitive Sciences and Education. En M. V. Alves, R. Ekuni, M. J. Hermida, & J. Valle-Lisboa (Eds.), *Cognitive Sciences and Education in Non-WEIRD Populations: A Latin American Perspective* (pp. 3–10). Springer International Publishing. https://doi.org/10.1007/978-3-031-06908-6_1
- ANEP. (2002). Los Niveles De Desempeño Al Inicio De La Educación Primaria.
- ANEP. (2021). LEO. Evaluación autónoma de Lectura, Escritura y Oralidad en segundo año de educación primaria. Documento conceptual.
- ANEP (2022, November 17). Seminario de trayectorias educativas. 20 años. YouTube. <https://www.youtube.com/watch?v=c8HDeg9mHIQ>
- Arnold, E. M., Goldston, D. B., Walsh, A. K., Reboussin, B. A., Daniel, S. S., Hickman, E., & Wood, F. B. (Directores). (2005). Severity of Emotional and Behavioral Problems Among Poor and Typical Readers. En *Journal of Abnormal Child Psychology* (Vol. 33, pp. 205–217). <https://doi.org/10.1007/s10802-005-1828-9>
- Bakker, M., Torbeyns, J., Wijns, N., Verschaffel, L., & De Smedt, B. (2019). Gender equality in 4- to 5-year-old preschoolers' early numerical competencies. *Developmental Science*, 22(1), e12718. <https://doi.org/10.1111/desc.12718>
- Bennett, R. (2015). The Changing Nature of Educational Assessment. *Review of Research in Education*, 39(1), 370–407. <https://doi.org/10.3102/0091732X14554179>
- Butterworth, B. (2005). Developmental dyscalculia. *Handbook of mathematical cognition*, 93, 455e467.
- Cabell, S. Q., Justice, L. M., Zucker, T. A., & Kilday, C. R. (2009). Validity of teacher report for assessing the emergent literacy skills of at-risk preschoolers. *Language, Speech, and Hearing Services in Schools*, 40(2), 161–173. [https://doi.org/10.1044/0161-1461\(2009/07-0099\)](https://doi.org/10.1044/0161-1461(2009/07-0099))
- Cardozo, S., Silveira, A., & Fonseca, B. (2022). Detección temprana del riesgo escolar. Predicción de trayectorias de rezago en la educación primaria en Uruguay mediante técnicas de machine learning. *Revista Latinoamericana de Estudios Educativos*, 52(2), 297–326. <https://doi.org/10.48102/rlee.2022.52.2.391>
- Ceci, S. J., & Williams, W. M. (2010). Sex Differences in Math-Intensive Fields. *Current Directions in Psychological Science*, 19(5), 275–279. <https://doi.org/10.1177/0963721410383241>
- Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, 135, 218–261. <https://doi.org/10.1037/a0014412>
- Cuetos, F., Rodríguez, B., Ruano, E., & Arribas, D. (2007). Prolec-R, Batería de evaluación de los procesos lectores, Revisada. Madrid: TEA.
- de León, D., Sánchez, I., Koleszar, V., Cervieri, I., & Marini, A. M. (2021). Actividades numéricas en el hogar y desempeño matemático en niños preescolares. *Revista Argentina de Ciencias del Comportamiento (RACC)*, 13(3), 49-58
- Delgado, A., Moreira, N., Goñi Mazzitelli, M., & Ortega, M. (2019). Derribando barreras: Por más mujeres en las áreas de ciencia, tecnología, ingeniería y matemáticas (STEM). <https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/27103>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L.,

- Engel, M., Brooks-Gunn, J., Sexton, H., & Duckworth, K. (2007). School Readiness and Later Achievement. 20.
- Dunlosky, J. (2013). Strengthening the Student Toolbox. *American Educator*, 37(3), 12–21.
- Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*, 54, 408–423. <https://doi.org/10.1037/0003-066X.54.6.408>
- Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives*, 3(1), 30–37. <https://doi.org/10.1111/j.1750-8606.2008.00072.x>
- Girelli, L. (2022). What does gender has to do with math? Complex questions require complex answers. *Journal of Neuroscience Research*, jnr.25056. <https://doi.org/10.1002/jnr.25056>
- Ginsburg, Herbert., Baroody, A. J., del Río, M. C., & Guerra, I. (2007). TEMA 3: Test de competencia matemática básica: Manual. En TEMA 3: Test de competencia matemática básica: Manual. TEA.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and brain sciences*, 33(2–3), 61–83; discussion 83-135. <https://doi.org/10.1017/S0140525X0999152X>
- Hutchison, J. E., Lyons, I. M., & Ansari, D. (2019). More Similar Than Different: Gender Differences in Children's Basic Numerical Skills Are the Exception Not the Rule. *Child Development*, 90(1), e66–e79. <https://doi.org/10.1111/cdev.13044>
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender Similarities Characterize Math Performance. *Science*, 321(5888), 494–495. <https://doi.org/10.1126/science.1160364>
- INEEd. (2022). Aristas Primaria 2020: Alineación entre currículo intencional, implementado y logrado.
- Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, 45(3), 850–867. <https://doi.org/10.1037/a0014939>
- Kersey, A. J., Braham, E. J., Csumitta, K. D., Libertus, M. E., & Cantlon, J. F. (2018). No intrinsic gender differences in children's earliest numerical abilities. *Npj Science of Learning*, 3(1), Art. 1. <https://doi.org/10.1038/s41539-018-0028-7>
- Kilday, C. R., Kinzie, M. B., Mashburn, A. J., & Whittaker, J. V. (2012). Accuracy of Teacher Judgments of Preschoolers' Math Skills. *Journal of Psychoeducational Assessment*, 30(2), 148–159. <https://doi.org/10.1177/0734282911412722>
- Koleszar, V., Zugarramurdi, C., Luzardo, M. (2019). A novel approach to estimate the approximate number system. International Meeting of the Psychometric Society. Santiago de Chile, Chile.
- Koleszar, V., de León, D., Díaz-Simón, N., Fitipalde, D., Cervieri, I., & Maiche, A. (2020). Numerical Cognition in Uruguay: from clinics and laboratories to the classroom (Cognición numérica en Uruguay: de la clínica y los laboratorios al aula). *Studies in Psychology*, 41(2), 294–318.
- López, F. M., de León, D., Díaz-Simón, N., & Maiche, A. (2022). Development of Mathematical Cognition: The Role of Technology in Low-SES Populations. In *Cognitive Sciences and Education in Non-WEIRD Populations* (pp. 169-184). Springer, Cham
- Lopez, A. Y., & Wilms, J. D. (2020). A National Evaluation of Kindergarten Outcomes: Findings from Uruguay. En J. Hall, A. Lindorff, & P. Sammons (Eds.), *International Perspectives in Educational Effectiveness Research* (pp. 361–381). Springer International Publishing. https://doi.org/10.1007/978-3-030-44810-3_13
- Luzardo, M. (2019). Item Selection Algorithms in Computerized Adaptive Test Comparison Using Items Modeled with Nonparametric Isotonic Model. En M. Wiberg, S. Culpepper, R. Janssen, J. González, & D. Molenaar (Eds.), *Quantitative Psychology* (pp. 95–105). Springer International Publishing. Lyytinen, H., Erskine, J., Hämmäläinen, J., Torppa, M., & Ronimus, M. (2015). Dyslexia—Early Identification and Prevention: Highlights from the Jyväskylä Longitudinal Study of Dyslexia. *Current Developmental Disorders Reports*, 2(4), 330–338. <https://doi.org/10.1007/s40474-015-0067-1>
- Marconi, C. & de León, D. & Luzardo, M. & Maiche, A. (2024, December 27). An early numeracy digital brief assessment: parametric and non-parametric IRT models.
- Mashburn, A. J., Hamre, B. K., Downer, J. T., & Pianta, R. C. (2006). Teacher and Classroom Characteristics Associated With Teachers' Ratings of Prekindergartners' Relationships and Behaviors. *Journal of Psychoeducational Assessment*, 24(4), 367–380. <https://doi.org/10.1177/0734282906290594>
- McCardle, P., Scarborough, H. S., & Catts, H. W. (2001). Predicting, Explaining, and Preventing Children's Reading Difficulties. *Learning Disabilities Research and Practice*, 16(4), 230–239. <https://doi.org/10.1111/0938-8982.00023>
- National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (p. 35).
- Ozernov-Palchik, O., & Gaab, N. (2016). Tackling the "dyslexia paradox": Reading brain and behavior for early markers of developmental dyslexia. *Wiley Interdisciplinary Reviews: Cognitive Science*, 7(2), 156–176. <https://doi.org/10.1002/wcs.1383>
- Popham, W. J. (2009). Assessment Literacy for Teachers: Faddish or Fundamental? *Theory Into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>
- Raudenbush, S. W., Hernandez, M., Goldin-Meadow, S., Carrazza, C., Foley, A., Leslie, D., Sorkin, J. E., & Levine, S.C. (2020). Longitudinally adaptive assessment and instruction increase numerical skills of preschool children. *Proceedings of the National Academy of Sciences*, 117(45), 27945–27953. <https://doi.org/10.1073/pnas.2002883117>
- Ripani, M. F. (2020). Uruguay: Ceibal en Casa (Ceibal at home).
- Rittle-Johnson, B., Fyfe, E. R., Hofer, K. G., & Farran, D. C. (2017). Early Math Trajectories: Low-Income Children's Mathematics Knowledge From Ages 4 to 11. *Child Development*, 88(5), 1727–1742. <https://doi.org/10.1111/cdev.12662>
- Schmader, T., Johns, M., & Barquissau, M. (2004). The Costs of Accepting Gender Differences: The Role of Stereotype Endorsement in Women's Experience in the Math Domain. *Sex Roles*, 50(11), 835–850. <https://doi.org/10.1023/B:SERS.0000029101.74557.a0>
- Stoet, G., & Geary, D. C. (2018). The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education. *Psychological Science*, 29(4), 581–593. <https://doi.org/10.1177/0956797617741719>
- Valle-Lisboa, J., Cabana, Á., Eisinger, R., Mailhos, Á., Luzardo, M., Halberda, J., & Maiche, A. (2016). Cognitive abilities that mediate SES's effect on elementary mathematics learning: The Uruguayan tablet-based intervention. *Prospects*, 46(2), 301–315. <https://doi.org/10.1007/s11125-017-9392-y>
- Vásquez-Echeverría, A. (2020). El Inventario de Desarrollo Infantil y la evaluación sistemática del desarrollo en contextos educativos. Teoría, creación e implementación. [Preprint]. *PsyArXiv*. <https://doi.org/10.31234/osf.io/xg2hj>
- Wanzek, J., & Vaughn, S. (2007). Research-Based Implications From Extensive Early Reading Interventions. *School Psychology Review*, 36(4), 541–561.
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Zugarramurdi, C., Assis, M. S., & Valle-Lisboa, J. (2022). Technology on Our Side: Using Technology for Transferring Cognitive Science to Education. In *Cognitive Sciences and Education in Non-WEIRD Populations* (pp. 287–303). Springer.
- Zugarramurdi, C., Fernández, L., Lallier, M., Carreiras, M., & Valle-Lisboa, J. C. (2022). Lexiland: A Tablet-based Universal Screener for Reading Difficulties in the School Context. *Journal of Educational Computing Research*, 0(0), 073563312210743. <https://doi.org/10.1177/07356331221074300>
- Zugarramurdi, C., Fernández, L., Lallier, M., Valle-Lisboa, J. C., & Carreiras, M. (2022). Mind the Orthography: Revisiting the Contribution of Prereading Phonological Awareness to Reading Acquisition. *Developmental Psychology*, 58(6), 1003–1016. <https://doi.org/10.1037/dev0001341>

Licenciamiento

Reconocimiento-NoComercial-Compartir Igual 4.0 Internacional. (CC BY-NC-SA)