

Desempeños diferenciados en una evaluación de matemáticas en Perú: entorno local y abstracción en escuelas rurales

César Guadalupe

Universidad del Pacífico
<https://orcid.org/0000-0002-2717-0814>
ca.guadalupe@up.edu.pe

Recibido: 15/01/2024
Aprobado: 18/12/2024

Resumen

Este artículo presenta un estudio exploratorio que se enfoca en identificar posibles factores que deben ser considerados para explicar diferencias de desempeño entre zonas urbanas y rurales en las evaluaciones estandarizadas de matemáticas en sexto grado de primaria. Los resultados sugieren que diferencias en la riqueza letrada y la demanda cognitiva presentes en cada entorno, incluyendo las identificadas en la dotación de docentes, deberían ser tomadas en cuenta para entender las diferencias en desempeño en desmedro de la población rural. Esto resulta particularmente relevante en el caso de las preguntas que suponen la ejecución de operaciones con mayor nivel de abstracción. Los resultados sugieren que existen dinámicas diferenciadas en los procesos de enseñanza-aprendizaje y en la relación de estos con el entorno local, las cuales deben ser consideradas para desarrollar acciones de mejora educativa.

Palabras clave: *evaluación de estudiantes, análisis comparativo, diferencias sociales, escuela rural.*

Abstract

This article presents an exploratory study focusing on identifying potential factors that should be considered to explain performance differences between urban and rural areas in standardised mathematics assessments at the sixth grade of primary education. The findings suggest that differences in the richness of literacy-related elements and cognitive demand present in each environment, including those identified in teacher allocation, should be taken into account to understand performance disparities to the detriment of the rural population. This is particularly relevant in questions involving operations with higher levels of abstraction. The results suggest that there are differentiated dynamics in teaching-learning processes and their relationship with the local environment, which must be considered when developing educational improvement initiatives.

Palabras clave: *student evaluation, comparative analysis, social differences, rural school.*

Introducción

Las mediciones estandarizadas de aprendizajes parten de un supuesto básico: son capaces de captar el atributo de interés (aprendizaje de un determinado constructo) de modo equivalente entre distintos individuos y poblaciones. Esto, sin embargo, no siempre se cumple, sea por errores en la construcción de las pruebas o de los ítems, o por circunstancias más complejas que resultan difíciles de anticipar en el diseño.

Por esta razón, la psicometría ha desarrollado técnicas destinadas a detectar y estimar la magnitud de posibles problemas en este terreno. Así, el análisis del funcionamiento diferencial de los ítems (o DIF, por sus siglas en inglés) (Zumbo, 1999; 2007) consiste en determinar si el o los parámetros con los que se caracteriza el comportamiento de cada ítem¹ se mantiene(n) constante(s) cuando se analizan segmentos diferentes de la población estudiada. Por ejemplo, este análisis permite explorar si se cumple que estudiantes de zonas urbanas y rurales con similar habilidad (estimada mediante el *pool* de ítems) tienen probabilidades similares de responder de modo correcto cada ítem en particular. Se asume que, si la prueba y los ítems miden la habilidad y solo esta, entonces, dos personas con habilidades similares deben contar con similares probabilidades de respuesta correcta a cada pregunta.

Sin embargo, no es lo mismo detectar y manejar estadísticamente posibles situaciones de funcionamiento diferencial de los ítems que explicar por qué se presentó dicha situación: ¿diferencias en la administración?, ¿una diferente interacción entre los procedimientos de administración y cada contexto?, ¿atributos diferentes de la población?

Existe literatura enfocada en determinar si algunas diferencias en las poblaciones se traducen en respuestas diferentes en las pruebas. Por un lado, se ha abordado el desarrollo de mecanismos que garantizan que la construcción de pruebas permita que estas midan, de forma equivalente, el rasgo de interés entre grupos lingüísticos y culturales diferentes (Dept et al., 2010; Hambleton, 2005). Ello permite el uso de una escala común, que es —a fin de cuentas— la *raison d'être* de las mediciones estandarizadas. Por otro lado, la literatura también se vincula con una reflexión iniciada por Luria (1976) sobre cómo las respuestas a determinadas preguntas se relacionan con las experiencias vitales que configuran la forma en la cual las personas representamos el mundo (nuestras formas y hábitos de pensamiento), y cómo desarrollamos habilidades vinculadas a, por ejemplo, orientaciones prácticas o esquemas clasificatorios abstractos, al pensamiento hipotético, entre otras.

1. La teoría de respuesta al ítem usa información empírica para caracterizar a cada ítem y a cada persona en la escala correspondiente al rasgo objeto de medición (la habilidad en el caso de las pruebas de logros de aprendizaje). Véase Baker (2001), o van der Linden y Hambleton (1997).

Es importante señalar que el estudio que ha dado pie a este artículo no se abocó a comprender o discutir las propiedades sicométricas de las pruebas usadas en el Perú, sino, más bien, a tomar evidencia sobre posibles comportamientos diferenciados de los ítems como punto de partida para explorar las maneras en las que estudiantes de zonas rurales pobres del país los resuelven. Para ello, tomamos los resultados del análisis estadístico de la Evaluación Nacional Peruana de 2013 con el fin de identificar características o fenómenos que podrían subyacer al comportamiento diferencial detectado, así como a las marcadas diferencias de desempeño entre estudiantes de zonas urbanas y rurales (Guadalupe et al., 2015; Guadalupe et al., 2017; Rotalde, 2023). Esto se realizó a través de un trabajo de campo en un número limitado de escuelas con características distintas (ya sea como escuelas en sí o en virtud de la localidad en la que se ubicaban). De hecho, la información sobre el funcionamiento diferencial de algunos ítems representó solo un punto de partida que permitió el inicio de la indagación de campo. Cabe añadir que tampoco pudo haber sido de otra forma, ya que la observación de campo no podía descansar en información (no existente) sobre los niveles de habilidad de los estudiantes efectivamente observados.

El trabajo de campo permitió identificar diferencias entre las maneras como los estudiantes interactúan con los distintos ítems. Encontramos aspectos vinculados a cómo los estudiantes operan con los saberes matemáticos contenidos en las pruebas, lo que brindó información sobre las diferencias (de habilidad y/o de funcionamiento de las pruebas) que deben ser tomadas en cuenta por la política educativa. Evidentemente, una administración estandarizada de pruebas no está diseñada para capturar lo que una observación detenida de la interacción entre las personas, los ítems y el contexto puede llegar a hacer visible. Esta, por su parte, puede constituir elementos claves para entender no solo los resultados que arrojan las pruebas, sino los problemas sustantivos que subyacen a los desempeños que las pruebas muestran. Una evidencia, por ejemplo, fue el caso de las pruebas de matemáticas en Inglaterra, analizado por Cooper y Dunne (1998; 2000).

Dicho todo lo anterior, en este artículo, abordé un área novedosa para la literatura especializada en el Perú. Su relevancia obedece no solo a los aspectos de validez de la información que las pruebas generan, sino también a la necesidad de promover un uso más cuidadoso de ella en la política educativa. Asimismo, el artículo brinda elementos para entender los atributos educativos de diversos ámbitos dentro del país. Específicamente, el estudio se enfocó en las pruebas de matemáticas administradas en sexto grado de primaria en noviembre de 2013 y la labor de campo fue conducida a lo largo del año 2016. La distancia temporal entre el trabajo de campo y la producción de este artículo obedece, principalmente, a un problema de fondo: la labor de campo sugirió que los problemas a analizar requerían de una reflexión teórica más compleja que la inicialmente prevista; por lo tanto, se requirió una exploración más profunda que forzó a posponer la producción de este texto. Si bien, en aquel

momento, escribí algunos borradores, estos, a mi propio juicio, no presentaban el problema de una manera lo suficientemente clara como para ser un aporte a la literatura.

Planteamiento del problema

Desde 1996, el Ministerio de Educación del Perú (Minedu) conduce evaluaciones estandarizadas de aprendizajes de los estudiantes de la educación básica tanto bajo la forma de esfuerzos nacionales² como en el marco de estudios internacionales³.

Las evaluaciones de aprendizaje de gran escala se han convertido en un área controversial debido al nexo entre ellas y el contexto político-institucional en el que se administran a nivel internacional, marcado por los debates sobre la gobernanza global de la educación (Bieber et al., 2015; Elfert y Ydesen, 2023; Grek, 2015, 2024; Martens, 2007; Martens y Niemann, 2013, 2013; Ydesen, 2019). Pese a ello, también se puede afirmar que dichas evaluaciones brindan información que —como toda información— puede resultar valiosa. Esto es posible si, por un lado, sus resultados son objeto de una lectura y tratamiento cuidadosos, y si, por otro lado, no se cae en el fetichismo de considerar que esta única pieza de información es, en algún sentido, lo único relevante para evaluar el desempeño de un sistema educativo o para definir la “calidad” de la educación⁴.

En la presente investigación, las pruebas estandarizadas han sido tomadas como evidencia de disparidades en los desempeños de nuestros estudiantes. Es decir, como evidencia de inequidades que, para efectos de este artículo, asociaremos al lugar de residencia, de modo que el foco de preocupación sea la desventaja educativa de los estudiantes de escuelas rurales. En ese sentido, nos preguntamos por los factores a considerar para entender el desempeño diferenciado de los estudiantes con el objetivo de informar los procesos de construcción y análisis de las pruebas, y de entender mejor las diferencias de contexto que afectan el desempeño escolar (tal y como es representado en las pruebas). Dichas diferencias deberían considerarse en el diseño y análisis de los resultados de las pruebas, así como en el diseño de las estrategias de enseñanza-aprendizaje en general y, en particular, en las zonas rurales.

-
2. Evaluaciones muestrales en 1996, 1998, 2001, 2004, 2013, 2017-2019, 2021-2023 que cubrieron diversos grados. Evaluaciones censales anuales de 2007 a 2016 en grado 2; 2018 y 2023, en grado 4; 2018 y 2019, en grado 8 (segundo de secundaria); así como cada dos años, en grado 4 para educación bilingüe.
 3. Estudios del Laboratorio Latinoamericano coordinado por Unesco en 1997, 2006, 2013 y 2019; el ya mencionado programa PISA, en 2001, 2012, 2015, 2018 y 2021; y el Estudio Internacional de Educación Cívica y Ciudadanía (ICCS), de la Asociación Internacional para la Evaluación del Logro Educativo (IEA) en 2016.
 4. Para una aproximación comprensiva a este concepto, véase Unesco/Santiago (2007).

El estudio fue planteado como exploratorio por dos motivos: por el carácter de la información sobre funcionamiento diferencial tomada como punto de partida y porque sería prematuro formular hipótesis específicas a ser investigadas máxime cuando los recursos disponibles no permitían una exploración más detallada de los entornos en los que viven los estudiantes.

La pregunta general que orientó —pero no limitó— el trabajo fue formulada así: ¿los patrones de funcionamiento diferencial de los ítems de la evaluación nacional peruana de 2013 pueden ser usados como punto de partida para detectar diferencias en los niveles de demanda cognitiva de las propias tareas *vis-à-vis* diferencias de esta naturaleza en los entornos (escolar y extraescolar) de los estudiantes?

Marco analítico

Las pruebas estandarizadas han mostrado gran capacidad para informar sobre los logros de aprendizaje y las brechas de equidad asociadas a estos, especialmente desde la publicación del emblemático Informe Coleman (Coleman et al., 1966). Asimismo, los métodos de medición se fueron sofisticando en la medida que la aplicación de pruebas en contextos diversos empezaba a plantear algunos desafíos (Hambleton, 2005; Hambleton et al., 1991; Harkness et al., 2010), así como por el uso de métodos basados en el análisis de variables latentes y la contribución de cada pregunta a ello (Baker, 2001; Lazarsfeld y Henry, 1968).

Un problema central es la pretensión de lograr resultados equivalentes entre poblaciones que son siempre heterogéneas. Ya Luria había observado que la manera como las personas respondían preguntas estandarizadas dependía directamente de los tipos de demanda cognitiva a los que estaban expuestas en su vida cotidiana (Luria, 1976), por lo que descartó el uso de pruebas estandarizadas en su estudio conducido en los años 30 (p. 16).

Esta vinculación entre entorno y desempeño en una prueba estandarizada ha sido observada de diversas formas. Por ejemplo, de esto se trata el llamado «efecto Flynn» (Flynn, 2007)⁵ y su reversión (Bratsberg y Rogeberg, 2018) *we categorize hypothesized causal factors by whether they accommodate the existence of within-family Flynn effects. Using administrative register data and cognitive ability scores from military conscription data covering three decades of Norwegian birth cohorts (1962-1991; o la relación entre clase social y des-*

5. Equiparando las pruebas de coeficiente intelectual usadas a lo largo del siglo XX en los Estados Unidos, Flynn (2007) muestra que la media normalizada ha aumentado en aproximadamente dos desviaciones estándar y que los mayores incrementos se verifican en las áreas que lidian con operaciones abstractas. Esto se explica, de acuerdo con Luria (1976), por cambios en el entorno: sociedades modernas donde el lenguaje clasificatorio, abstracto e hipotético de la ciencia está más presente en los asuntos cotidianos (entre otras razones, por la expansión de la escuela) que en sociedades con demandas cotidianas de un carácter más práctico o funcional.

empeño académico en pruebas estandarizadas (Cooper y Dunne, 1998, 2000); o entre la aplicación de una prueba internacional que usa información local en un contexto muy específico, como lo es la población nómada de Mongolia (Maddox, 2014, 2015).

Las observaciones subrayan que las pruebas estandarizadas no se comportan de modo estándar en situaciones concretas, lo que resulta crítico pues —de hecho— solo existen situaciones concretas. Por cierto, este es un problema general relacionado con la naturaleza de la labor cognoscitiva (que hace abstracción de la multiplicidad de elementos que componen una situación siempre singular); en ese sentido, debe ser considerado para ponderar con robustez el alcance y las limitaciones de cualquier forma de conocimiento, incluyendo las pruebas estandarizadas (Guadalupe, 2017). Asimismo, la investigación más reciente sobre cognición humana muestra que el desarrollo del pensamiento abstracto está íntimamente vinculado con las competencias lingüísticas de las personas en tanto el lenguaje es la “herramienta social” sobre la que reposa esta forma de pensamiento (Barca et al., 2019; Borghi, 2022, 2023; Borghi y Tummolini, 2020; Granito y Scrolli, 2015; Langland-Hassan et al., 2021).

En el contexto de profundas disparidades que marcan al sistema educativo peruano, cabe preguntarse acerca de los problemas asociados a diferencias de riqueza en la demanda cognitiva cotidiana que se expresarían en un desempeño diferenciado según los tipos de ítems (nivel de abstracción, tipo de operación, tipo de contexto, etc.) considerados en una prueba dada. Finalmente, esta investigación se entronca en una reflexión general sobre la relación entre contexto y formas de pensamiento y, dentro de esta, en el área de las matemáticas. Por un lado, la sociología (Bernstein, 1990; Bourdieu, 1997; Bourdieu y Passeron, 2000) ha destacado que las dificultades experimentadas por los estudiantes en la experiencia escolar están asociadas a su posición en la estructura social. Por otro lado, como ya se mencionó, la reflexión contemporánea sobre cognición y pensamiento abstracto destaca la centralidad del lenguaje y la socialidad en ello.

Diseño y métodos

La investigación partió del reporte de análisis de funcionamiento diferencial de ítems elaborado por el Minedu⁶. Con esta información, se tomó un conjunto de cuatro ítems (dos con DIF en contra de la población rural y dos sin DIF), y se administraron en seis escuelas del país. Luego, se condujo un grupo focal con los estudiantes participantes para identificar la manera como enfrentaron cada una de las preguntas.

El reporte DIF incluyó los 107 ítems usados en la prueba de matemáticas de la Evaluación Muestral de 2013. Se concluyó que estos brindaron, *en conjun-*

6. Si bien este reporte no ha sido publicado, el autor solicitó a la oficina correspondiente del Minedu acceso al mismo.

to, una representación sin mayores sesgos de las habilidades de los estudiantes de escuelas urbanas y rurales. Aunque se registró algunos casos con DIF, estos se presentaron en ambas direcciones, lo que canceló su efecto en el agregado de la prueba.

Dado que el interés de este trabajo no era la robustez de la prueba, sino explorar diferencias a nivel de ítems singulares, fue posible encontrar un número pequeño de ítems con DIF desfavorecedor de la población rural. De esos ítems, solo dos pertenecían al conjunto de ítems liberados⁷ que podíamos utilizar en nuestro trabajo.

La operación de campo cubrió los siguientes elementos:

- i. Se presentó el propósito de la investigación; se solicitó la colaboración de los directores, docentes y estudiantes; se expresó un compromiso de confidencialidad; y se solicitó la autorización (verbal) para proceder.
- ii. Los estudiantes respondieron individualmente las cuatro preguntas presentadas en un cuestionario en papel.
- iii. Luego, se sostuvo una conversación semiestructurada con ellos acerca de la manera como habían resuelto cada pregunta y se indagó sobre el tipo de dificultades que habían tenido, qué les resultó más fácil, etc.
- iv. Posteriormente, se hizo una observación de la escuela (condiciones materiales, organización física, disponibilidad de materiales impresos, organización del aula), y se conversó con los docentes; esta observación tuvo como propósito registrar la presencia o ausencia de materiales impresos y textos, así como explorar con los docentes posibles explicaciones o percepciones acerca del desempeño de los estudiantes.
- v. Finalmente, se hizo una observación general del entorno local con el propósito de identificar elementos textuales en el entorno próximo de la escuela.

Toda la información fue registrada en notas de campo y organizada de acuerdo con los siguientes elementos: tipo de dificultad reportada, tipo de ejercicio que presentaba mayor facilidad, naturalidad en la respuesta, grado de formalización escolar de las respuestas o estrategia de resolución del problema, y densidad del entorno letrado.

Los ítems usados (ver Anexo) pueden ser caracterizados de la siguiente manera:

7. La necesidad de reutilizar ítems para tener comparabilidad en el tiempo lleva a que solo se pueda acceder a un subconjunto “liberado” a efectos de ilustrar las características de las pruebas o para permitir su uso en investigaciones como la presente.

- El primer ítem (*Triángulo isósceles*) fue respondido satisfactoriamente por un estudiante con un desempeño típico en la prueba nacional.⁸ Sin embargo, mostró DIF en contra de la población rural. La formulación del ítem no hace referencia a contexto o propósito alguno, y solicita resolver una pregunta por el solo propósito de resolverla. Las figuras presentadas, además, no representan objetos reales de ningún tipo. Así, se trata de un ítem completamente desprovisto de contexto o, lo que es lo mismo, desnudamente algorítmico. Su solución requiere claridad sobre los dos conceptos que forman la pregunta (triángulo e isósceles), y detectar que, en los gráficos, existía un elemento de apoyo que simplificaba su resolución.
- El segundo ítem (*Banda de música*) resultó, en la prueba nacional, más difícil que la habilidad media (dificultad 539). Este apela a un contexto presumiblemente familiar por tratarse de una situación escolar que, sin embargo, no es necesariamente común (tener una banda de música). Al mismo tiempo, plantea un problema con fracciones que, como se verá más adelante, suscita reacciones particulares entre los estudiantes.
- El tercer ítem (*Ganchos para colgar*) es uno cuya dificultad se encuentra entre la de los dos anteriores (523 puntos). Asimismo, mostró DIF en contra de la población rural. Hace referencia al contexto escolar, aunque no es claro en qué medida los trabajos de los estudiantes son presentados de la forma que el ejercicio plantea (en las escuelas visitadas, los trabajos eran adheridos a paneles o paredes usando cinta adhesiva o pines). Además, combina una ayuda visual con información tabulada que sugiere cómo aproximarse a la respuesta.
- El cuarto ítem (*Cantidad de estudiantes*) presentó una dificultad ligeramente superior a la media nacional (505 puntos) y presenta un gráfico de barras superpuestas con información nuevamente referida al contexto escolar. El problema planteado es simple (determinar una diferencia) y demanda operaciones básicas: sumar dos grupos de seis cantidades y, luego, comparar los resultados; o hacer seis sustracciones y, luego, sumar algebraicamente los resultados parciales.

8. El parámetro de dificultad del ítem es idéntico a la media nacional (500). Nótese que el desempeño observado en 2013 está lejos de ser satisfactorio ya que solo 16 % (e.e.= 0,63) de los estudiantes logró un desempeño satisfactorio en matemáticas (Minedu, 2016, p. 62). La media nacional se encuentra aproximadamente a una desviación estándar (100 puntos en la escala) por debajo del umbral para considerar el desempeño como satisfactorio (p. 68). Los parámetros de dificultad de los ítems fueron proporcionados por la UMC como parte de la descripción de los ítems liberados.

Características de las escuelas y los estudiantes

El trabajo de campo se condujo en aulas de sexto grado de primaria de seis escuelas seleccionadas de la siguiente manera:

- Una escuela de gestión no estatal (particular pagada), prestigiosa y reconocida por sus buenos resultados educativos, ubicada en la capital del país (Lima) en una zona residencial de sectores de ingresos medios/altos, con una población escolar de dichos sectores. La intención fue usar las respuestas de estos estudiantes como un rasero de comparación para el análisis de las respuestas obtenidas en las escuelas de interés. Se seleccionó aleatoriamente a seis estudiantes de sexto grado de primaria.
- Una escuela de gestión estatal de la provincia de Huarochirí (Lima), ubicada en una zona urbana de limitados recursos, aunque con comunicación fluida a la ciudad de Lima. Se trabajó con los 16 estudiantes de la sección de sexto grado.
- Una escuela multigrado de gestión estatal (Sector Educación), ubicada en un centro poblado rural de la provincia de Huarochirí a cerca de 4000 metros sobre el nivel del mar y con limitado acceso mediante transporte público (un bus que conecta dos veces a la semana con una ciudad intermedia; la distancia a Lima es de aproximadamente seis horas en vehículo particular, cuatro de las cuales por caminos de tierra y cruzando un paso a aproximadamente 4900 metros sobre el nivel del mar). Se trabajó en la cocina de la escuela con los seis estudiantes de sexto grado.
- Una escuela completa (no multigrado) de gestión estatal (Sector Educación), ubicada en un centro poblado rural de la provincia de Yauyos (Lima) a cerca de 4000 metros sobre el nivel del mar y con limitado acceso mediante transporte público (condiciones de acceso similares a la descrita anteriormente). Se trabajó en el aula con los 10 estudiantes de sexto grado.
- Una escuela unidocente rural de gestión estatal (Sector Educación) de la provincia de Ambo (Huánuco), ubicada a varias horas a pie del punto caminero más próximo. Se trabajó con los dos estudiantes de sexto grado.
- Una escuela multigrado rural de gestión estatal (Sector Educación) de la provincia de Ambo (Huánuco), ubicada a varias horas a pie del punto caminero más próximo. Se trabajó con un estudiante de quinto grado, ya que no había estudiantes de sexto grado.

La observación de los entornos de trabajo sugiere una clara diferencia entre la escuela tomada como referencia, y las condiciones materiales y

simbólicas, incluyendo la poca densidad letrada⁹ de los demás entornos. En todos los casos, el castellano era la lengua materna de los estudiantes. El trabajo de campo se condujo en noviembre de 2016 a efectos de asegurar que el momento del ciclo escolar coincidiera con el previsto en el diseño de los ítems (administrados en noviembre de 2013).

Análisis y resultados

En la escuela de referencia, los estudiantes podían resolver los ejercicios con facilidad. Respondieron todos los ejercicios en menos de ocho minutos, y solo dos estudiantes cometieron un error cada uno. El diálogo con ellos se desarrolló con naturalidad y las explicaciones de sus respuestas eran naturales, es decir, no forzosamente formalizadas. Tanto chicos como chicas se expresaron con claridad y soltura.

La resolución del primer problema (*Triángulo isóceles*) descansó, en casi todos los casos, en utilizar la cuadrícula proporcionada en la pregunta como instrumento de medición para verificar lo observado. Se presentó un solo caso en el que se optó por utilizar una regla graduada (que el estudiante tenía consigo). Aquí, los estudiantes partieron por descartar, de modo inmediato, la primera figura por no ser un triángulo y rápidamente identificaron, por inspección visual, la figura *d* como la que potencialmente sería la respuesta, lo que verificaron con la medición. Debe notarse, también, que notaron que la inclinación de los lados de la figura respecto de la cuadrícula impedía que los lados se midiesen usando recuadros enteros. Esto obligaba a verificar la inclinación en los dos lados de modo que se pudiese asumir que eran equiparables.

La resolución del segundo problema (*Banda de música*) siguió dos estrategias diferentes: la primera consistió en multiplicar mentalmente el número de estudiantes (30) por $\frac{2}{5}$ bajo la forma de $30 \times 2 = 60$, y luego $60/5 = 12$. Sin embargo, una estudiante operó de una forma menos “escolar”: obtener primero el quinto de 30 (6) y, luego, duplicar ese resultado.

El tercer problema (*Ganchos para colgar*) fue resuelto de dos formas: dos de los estudiantes contaron (parcialmente) una a una la necesidad de ganchos adicionales; los demás, a los que se sumaron los dos que empezaron contando, identificaron que solo la primera hoja necesitaba cuatro ganchos y que cada hoja adicional requería dos, por lo que el total de ganchos era igual a cuatro más dos veces el número de hojas adicionales.

El cuarto problema (*Cantidad de estudiantes*) fue resuelto sumando las cantidades, excepto en un caso, donde la estudiante agregó algebraicamente las diferencias.

9. La limitada presencia de textos escritos suele, además, ser deficiente en términos ortográficos, uso de preposiciones o problemas de concordancia gramatical, los que obstaculizan la comprensión de los textos.

La primera diferencia entre la escuela de referencia y las demás se vincula al desempeño general de los estudiantes: en las últimas, siempre se utilizó un tiempo marcadamente mayor (entre 14 y 20 minutos), y todos los estudiantes tuvieron dificultades para responder al menos uno de los ítems.

Asimismo, el primer ítem (*Triángulo isósceles*) fue respondido con mucha dificultad y no siempre de modo certero. Un número reducido de estudiantes descartó la primera figura por no ser un triángulo, y todos ignoraron la presencia de las cuadrículas y apelaron a reglas graduadas para medir los lados. No hubo una preselección de alguna figura que pudiese ser un mejor candidato a respuesta correcta a partir de la inspección visual. Un número pequeño de estudiantes en la escuela más próxima a Lima mostró mayor facilidad para resolver el problema, ya que, efectivamente, descartó la primera figura y , sin poder expresar claramente su razones, descartó la figura b por su irregularidad. Cuando se les preguntó por qué no habían usado la cuadrícula, los estudiantes no expresaron razones, excepto en dos casos en los que se mencionó que dicha cuadrícula no tenía unidades, por lo que no habría certeza si se usaban como elemento de medida. En general, los estudiantes procedieron a medir los lados tan pronto leyeron la pregunta, sin detenerse a pensar en qué casos eso podría ser necesario.

El segundo ítem (*Banda de música*) fue prácticamente imposible de responder por una marcada dificultad para operar con fracciones: su presencia parecía ser ajena, desconocida o intimidante; asimismo, los estudiantes trataron de evocar el procedimiento formal para operar con fracciones sin tener mayor claridad del propósito. Los pocos estudiantes que trataron de formular la operación $30 \cdot \frac{2}{5}$, tuvieron dificultades para ir más allá del planteamiento tanto por no entender por qué debían usar esa formulación (lo que se expresa en no poder explicar) como por no recordar cómo operar con esta.

En la escuela de Yauyos, el investigador indagó oralmente acerca de algunas operaciones básicas que, para el sexto grado, deberían estar automatizadas:

¿Cuánto es la mitad de 30? Todos respondieron satisfactoriamente en automático.

¿Cuánto es un tercio de 30? Solo dos estudiantes respondieron correctamente, dos dudaron antes de dar una respuesta incorrecta y los demás no sabían cómo responder la pregunta.

¿Cuánto es un quinto de 30? Ningún estudiante pudo responder esa pregunta, y no parecía ser evidente la conexión entre estas preguntas y la resolución de este problema.

En la escuela más próxima a Lima, un número reducido de estudiantes resolvió exitosamente el problema luego de formalizarlo y operarlo; un par de ellos erró en las operaciones, pero lo había formulado correctamente, como quedó claro en el intercambio posterior.

El tercer ítem (*Ganchos para colgar*) fue resuelto en pocos casos y siempre apelando a una representación concreta del problema en el papel, sea completando y extendiendo la tabla prevista, o dibujando las 20 hojas y contando los ganchos. En ningún caso, hubo una solución que fuese formulada como una regla del tipo: cuatro para el primero, solo dos para cada hoja adicional.

El cuarto ítem (*Cantidad de estudiantes*) fue respondido, en la mayor parte de los casos, de modo correcto a partir de la suma de todos los valores observados. Ninguno operó con las diferencias (que hubiese permitido operar con magnitudes menores).

Debe anotarse que la mayor parte de los estudiantes mostraron dificultades para verbalizar una explicación acerca de cómo abordaron los problemas.

Discusión

Los resultados muestran un abismo entre la forma como respondieron los estudiantes de la escuela de referencia y la de los demás. Debe notarse que las preguntas no exigían un desempeño particularmente destacado o siquiera satisfactorio; por el contrario, los cuatro ejercicios usados eran simples¹⁰. En ese sentido, es esperable que un estudiante de finales del sexto grado de primaria los resuelva correctamente, mostrando ciertos niveles de automaticidad en la operación (lo que lleva a una economía del tiempo) y desplegando estrategias basadas en los recursos disponibles (como las cuadrículas en el primer ítem). Así, si bien es posible que los estudiantes de la escuela de referencia puedan mostrar desempeños particularmente destacados, estos no eran necesarios para resolver estos problemas. La pauta encontrada en esta escuela (con relación a las preguntas administradas) es una suerte de mínimo al que todo estudiante debería llegar en cualquier parte del país.

Resulta claro que las demás escuelas están lejos de esos desempeños mínimos esperables. Sin embargo, hay algunas diferencias en favor de aquella más próxima a Lima en el caso de los dos primeros ítems.

Las dificultades observadas en el intercambio con los estudiantes se asocian a varios problemas:

- a. *Dificultades relativas a la internalización de conceptos.* Por ejemplo, el primer ítem suponía entender los conceptos de triángulo e isósceles, y ambos no eran claros para la totalidad de estudiantes de las escuelas de interés.
- b. *Dificultades relativas al planteamiento de una estrategia de resolución de problemas.* Se apreció una tendencia a operar cada problema sin reflexionar antes, y la operación descansaba en la posibilidad de equiparar el problema a situaciones experimentadas en la vida escolar, como

10. Como lo muestra el valor de los correspondientes parámetros de dificultad, estos valores se encuentran en la categoría de desempeño “En proceso.”

recordar la “fórmula” para operar con fracciones; usar reglas graduadas; entre otras. Esto denota un bajo nivel de automatización de las operaciones y de capacidad de trasladar lo aprendido en un contexto a otro. Asimismo, los estudiantes se enfocaban más en la operación que en algún propósito explícito o implícitamente planteado en las preguntas, por lo que construir una respuesta descansaba fundamentalmente en la operación.

- c. *Dificultades relativas al desarrollo de estrategias propias para enfrentar un problema.* En general, se observó que, en ocasiones, los estudiantes fueron claros y explícitos al afirmar que no podían resolver algo porque no recordaban el procedimiento que habían visto en la escuela. Esto indica que, si bien esos procedimientos pueden haber sido trabajados, no se logra una apropiación de los mismos por parte de los estudiantes.

Asimismo, llama la atención que el ítem más fácil (*cantidad de estudiantes*) sea aquel caracterizado por ser muy concreto, referido al entorno escolar que es familiar a todos los estudiantes, y que requiere operaciones aritméticas muy simples que resultan evidentes en la propia presentación del problema.

Si bien no es posible demostrar que las diferencias en el desempeño de los estudiantes —y, eventualmente, en el comportamiento de los ítems— obedezca a diferencias en los entornos (lo rural versus lo urbano), sí es posible encontrar algunos elementos que sugieren que una explicación puede vincularse a ello. En particular, se puede mencionar lo siguiente:

En las comunidades menos accesibles desde entornos urbanos consolidados, se observan mayores dificultades en general y una práctica imposibilidad de resolver los dos ítems que mostraron DIF contra los estudiantes de escuelas rurales.

En efecto, los entornos rurales son mucho menos densos en términos de material escrito, y las actividades regulares (económicas, domésticas) suponen una menor exposición a elementos abstractos propios de la vida urbana. Asimismo, si bien son comunidades castellanohablantes, las conversaciones informales sostenidas entre el investigador y personas de la comunidad, así como los pocos textos disponibles, sugerían una escasa apropiación de las formas y recursos estándar del castellano. Todo ello explicaría (siguiendo a Luria y a Flynn, así como a Borghi) una mayor probabilidad de presentar dificultades con operaciones y conceptos abstractos.

Asimismo, los niveles de demanda cognitiva observados en las escuelas (especialmente en las de sectores sociales en desventaja) fueron particularmente bajos (Cueto et al., 2003). Esto derivaría en limitadas oportunidades de aprender y en el no desarrollo de habilidades más complejas que el propio entorno local tampoco favorece.

Por su parte, los docentes de las zonas rurales presentan mayores debilidades formativas que sus pares urbanos (Guadalupe et al., 2013; Guadalupe

et al., 2017). Ello se traduce en un menor capital educativo; de hecho, la propia evaluación muestral de 2013 muestra evidencia extremadamente preocupante al respecto. En efecto, a los docentes responsables de la enseñanza de matemática se les solicitó responder un cuestionario¹¹ que, en la sección segunda (páginas 7-12), incluyó 33 preguntas cerradas y dos abiertas sobre fracciones. Las preguntas fueron construidas atendiendo a lo que un estudiante debe lograr al término de la educación primaria; por lo mismo, cabría esperar que el 100 % de los estudiantes cuente con docentes de matemáticas que puedan resolver todas las preguntas planteadas. Lamentablemente, ese no fue el caso. No hubo una sola pregunta respondida adecuadamente por todos los docentes, y con algunas preguntas, se mostró que hasta tres de cada cuatro estudiantes de zonas rurales tenían docentes que no podían responderla de modo certero.^{12 13} Al mismo tiempo, los resultados no fueron mucho mejores en el ámbito urbano.

Ahora bien, ninguno de los factores mencionados es consustancial a la ruralidad. Estos son el resultado de la situación de desventaja en la que se encuentra el mundo rural en el Perú y lo que observamos estaría vinculado a la reproducción de dichas desigualdades.

En adición a lo anterior, el trabajo de campo encontró también dos fenómenos complejos que caben anotarse:

- i. En los grupos focales, se experimentó una situación de inhibición en la participación de las niñas muy marcada en la escuela de Yauyos.
- ii. Los estudiantes presentaron dificultades para verbalizar las explicaciones de los procedimientos que habían seguido (en todas las escuelas, excepto en la de referencia, y con menor intensidad en aquella próxima a Lima); usualmente, formulaban una frase o dos, no necesariamente bien conectadas; cuando el investigador repreguntaba para lograr mayor elaboración, los estudiantes tendían a repetir lo dicho, a empezar una nueva frase que quedaba trunca o a permanecer en silencio.

11. Disponible en <http://umc.minedu.gob.pe/wp-content/uploads/2016/07/CUESTIONARIO-DOCENTE-MATEMATICA.pdf>.

12. El diseño muestral tiene a los estudiantes como unidad de observación. Así, los resultados de los docentes no permiten inferencias sobre el universo de docentes, sino en función del número de estudiantes (y la ponderación de estos) que cada docente atiende. Así, los resultados se expresan bajo la forma general: “La pregunta x fue respondida de modo correcto/incorrecto/en blanco por los docentes que atienden al y por ciento de estudiantes”.

13. Por ejemplo, la pregunta 8.3 requería que el docente pudiera formalizar como fracciones tres ejemplos de subconjuntos. Esta pregunta fue respondida de modo erróneo por los docentes de dos tercios de los estudiantes de escuelas estatales de zonas rurales y de la mitad sus pares urbanos (a lo que habría que añadir un 7 % que la dejó en blanco en ambas zonas).

Si bien la lengua materna de todos los participantes del estudio era el castellano, pareciera que el dominio de esta es bastante limitado, lo que deviene en un impedimento para construir explicaciones o elaborar ideas. En ese sentido, tras los problemas de desempeño escolar, parece existir un problema anterior de competencia lingüística que la exposición a la vida cotidiana del país, incluyendo sus medios de comunicación, sugiere que es un fenómeno extendido.

Reflexión final

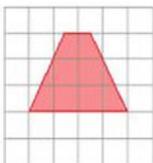
La evidencia generada por el trabajo de campo permite sugerir que, efectivamente, existen elementos del entorno que podrían estar explicando el menor desempeño rural; así mismo, que esos elementos son claves para entender la dinámica de los procesos de enseñanza y aprendizaje que se viven de diversas maneras dentro del país. Por otro lado, la evidencia muestra que existen condiciones (seguramente definidas por la conjunción de la marginación del mundo rural, con atributos propios del sistema educativo) que llevarían a los estudiantes a experimentar dificultades en el desempeño matemático. Esto, no solo porque “no saben”, sino también porque el conocimiento adquirido está apegado a formatos rígidos de operación más que a propósitos generales. Ello podría vincularse a la ausencia de situaciones locales en las que lo explorado en clase pueda adquirir una concreción familiar que impela al estudiante a no almacenar en un compartimento estanco (“solo para fines escolares”) lo aprendido, sino, por el contrario, a movilizar esos aprendizajes en los contextos en los que se desenvuelven las actividades cotidianas.

Anexo.
Ítems utilizados en el trabajo de campo, reproducidos con la autorización de la Oficina de Medición de la Calidad de los Aprendizajes (UMC) del Minedu.

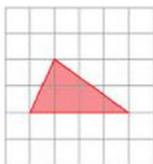
1. Triángulo isósceles

27. Un triángulo es isósceles cuando dos de sus lados tienen la misma medida.
 ¿Cuál de las siguientes figuras sombreadas es un triángulo isósceles?

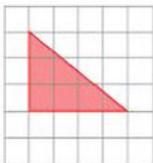
a



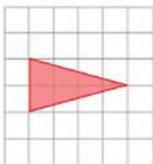
b



c



d



MAP00464

2. Banda de música

11. En la banda de música de la escuela hay 30 estudiantes, de los cuales los $\frac{2}{5}$ tocan tambor. ¿Cuántos estudiantes de la banda tocan tambor?

- a) 60 estudiantes.
 b) 15 estudiantes.
 c) 12 estudiantes.
 d) 6 estudiantes.

MABP0478

3. Ganchos para colgar

2. Cecilia observa que su maestra utiliza ganchos para colgar las hojas de trabajo de sus compañeros de la siguiente manera.



Luego, Cecilia elabora la siguiente tabla:

Cantidad de hojas de trabajo	1	2	3	4
Cantidad de ganchos utilizados	4	6	8

Si ella sabe que la maestra va a colgar 20 hojas de trabajo, ¿cuántos ganchos utilizará?

- a) 80 ganchos.
 b) 50 ganchos.
 c) 42 ganchos.
 d) 12 ganchos.

MABP0329

4. Cantidad de estudiantes

27. El siguiente gráfico muestra la cantidad de estudiantes de una escuela de acuerdo con el grado y el sexo. Observa:



Según el gráfico, en esta escuela ¿hay más estudiantes varones o hay más estudiantes mujeres?

Escribe aquí tu procedimiento.

Respuesta: _____

MAEP00554

Referencias

- Baker, F. B. (2001). *The basics of item response theory* (2a. Ed.). ERIC Clearinghouse on Assessment and Evaluation.
- Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., Tummolini, L., y Borghi, A. M. (2019). Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of Life Reviews*, 29, 120–153. <https://linkinghub.elsevier.com/retrieve/pii/S1571064518301271>
- Bernstein, B. (1990). *The structuring of pedagogic discourse*. Routledge.
- Bieber, T., Martens, K., Newmann, D., y Teltemann, J. (2015). Towards a global model in education? International student literacy assessments and their impact on policies and institutions. En M. Hamilton, B. Maddox, y C. Addey (Eds.), *Literacy as Numbers: Researching the Politics and Practices of International Literacy Assessment Regimes* (pp. 165–186). Cambridge University Press.
- Borghi, A. M. (2022). Concepts for Which We Need Others More: The Case of Abstract Concepts. *Current Directions in Psychological Science*, 31(3), 238–246. SAGE Publications Inc. <https://doi.org/10.1177/096372142211079625>
- Borghi, A. M. (2023). *The Freedom of Words: Abstractness and the Power of Language*. Cambridge: Cambridge University Press.
- Borghi, A. M., y Tummolini, L. (2020). Touch me if you can: The intangible but grounded nature of abstract concepts. *Behavioral and Brain Sciences*, 43, e123. Cambridge University Press. <https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/touch-me-if-you-can-the-intangible-but-grounded-nature-of-abstract-concepts/8B564BE2696574E688CA701B8ED697F1>
- Bourdieu, P. (1997). *Capital cultural, escuela y espacio social*. Siglo XXI.
- Bourdieu, P., y Passeron, J.-C. (2000). *Reproduction in Education, Society and Culture* (2a. Ed.). Thousand Oaks, New Delhi: Sage.
- Bratsberg, B., y Rogeberg, O. (2018). Flynn effect and its reversal are both environmentally caused. *Proceedings of the National Academy of Sciences of the United States of America*, 115(26), 6674–6678. National Academy of Sciences. <http://www.ncbi.nlm.nih.gov/pubmed/29891660>
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., y York, R. L. (1966). *Equality of educational opportunity*. US Government Printing Office.
- Cooper, B., y Dunne, M. (1998). Anyone for Tennis? Social Class Differences in Children's Responses to National Curriculum Mathematics Testing. *The Sociological Review*, 46(1), 115–148.
- Cooper, B., y Dunne, M. (2000). *Assessing children's mathematical knowledge. Social class, sex and problem-solving*. Open University Press.

- Cueto, S., Ramírez, C., León, J., y Pain, O. (2003). *Oportunidades de aprendizaje y rendimiento en matemática en una muestra de estudiantes de sexto grado de primaria de Lima* (No. 43). GRADE.
- Dept, S., Ferrari, A., y Wäyrynen, L. (2010). Developments in Translation Verification Procedures in Three Multilingual Assessments: A Plea for an Integrated Translation and Adaptation Monitoring Tool. En J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B.-E. Pennell, et al. (Eds.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 157–173). John Wiley & Sons, Inc.
- Elfert, M., y Ydesen, C. (2023). *Global Governance of Education. The Historical and Contemporary Entanglements of UNESCO, the OECD and the World Bank*. Educational Governance Research. Springer.
- Flynn, J. (2007). *What is intelligence?: Beyond the Flynn effect*. Cambridge University Press.
- Granito, C., y Scorolli, C. (2015). Naming a Lego World. The Role of Language in the Acquisition of Abstract Concepts. *PLOS ONE*, 10(1), e0114615. <https://dx.plos.org/10.1371/journal.pone.0114615>
- Grek, S. (2015). Transnational education Policy-making: International assessments and the formation of a new institutional order. En M. Hamilton, B. Maddox, y C. Addey (Eds.), *Literacy as Numbers: Researching the Politics and Practices of International Literacy Assessment Regimes* (pp. 35–52). Cambridge University Press.
- Grek, S. (2024). *The New Production of Expert Knowledge: Education, Quantification and Utopia*. Cham, Switzerland: Springer Nature. <https://library.oapen.org/handle/20.500.12657/86096>
- Guadalupe, C. (2017). Standardisation and diversity in international assessments: Barking up the wrong tree? *Critical Studies in Education*, 58(3), 326–340.
- Guadalupe, C., Burga, A., Miranda, L., y Castillo, L. (2015). Brechas de equidad en la Evaluación Censal de Estudiantes 2007-2014: Tres aproximaciones a su medición. *Persona*, 18, 47–68.
- Guadalupe, C., León, J., y Cueto, S. (2013). *Charting progress in learning outcomes in Peru using national assessments*. Paper commissioned for the EFA Global Monitoring Report 2013/4, Teaching and learning: Achieving quality education for all. Paris.
- Guadalupe, C., León, J., Rodríguez, J., y Vargas, S. (2017). *Estado de la educación en el Perú. Análisis y perspectivas de la educación básica*. GRADE. <http://www.grade.org.pe/publicaciones/estado-de-la-educacion-en-el-peru-analisis-y-perspectivas-de-la-educacion-basica/>
- Hambleton, R. K. (2005). Issues, Designs and Technical Guidelines for Adapting Tests Into Multiple Languages and Cultures. En R. K. Hambleton, P.

- Merenda, y C. Spielberger (Eds.), *Adapting Psychological and Educational Tests for Cross-Cultural Assessment* (pp. 3–38). Lawrence Erlbaum.
- Hambleton, R. K., Swaminathan, H., y Rogers, J. (1991). *Fundamentals of Item Response Theory*. Sage.
- Harkness, J. A., Braun, M., Edwards, B., Johnson, T. P., Lyberg, L., Mohler, P. Ph., Pennell, B.-E., et al. (Eds.). (2010). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. Wiley.
- Langland-Hassan, P., Faries, F. R., Gatyas, M., Dietz, A., y Richardson, M. J. (2021). Assessing abstract thought and its relation to language with a new nonverbal paradigm: Evidence from aphasia. *Cognition*, 211, 104622. <https://linkinghub.elsevier.com/retrieve/pii/S001002772100041X>
- Lazarsfeld, P., y Henry, N. (1968). *Latent structure analysis*. Houghton Mifflin.
- van der Linden, W., y Hambleton, R. K. (Eds.). (1997). *Handbook of Modern Item Response Theory*. Springer.
- Luria, A. (1976). *Cognitive development, its cultural and social foundations*. Harvard University Press.
- Maddox, B. (2014). Globalising Assessment: An Ethnography of Literacy Assessment, Camels and Fast Food in the Mongolian Gobi. *Comparative Education*, 50(4), 474–489.
- Maddox, B. (2015). Inside the Assessment Machine: The Life and Times of a Test Item. En M. Hamilton, B. Maddox, y C. Addey (Eds.), *Literacy as Numbers: Researching the Politics and Practices of International Literacy Assessment Regimes* (pp. 129–146). Cambridge University Press.
- Martens, K. (2007). How To Become an Influential Actor: The “Comparative Turn” in OECD Education Policy. En K. Martens, A. Rusconi, y K. Leuze (Eds.), *New Arenas of Education Governance. The Impact of International Organizations and Markets on Educational Policy Making* (pp. 40–56). Palgrave Macmillan.
- Martens, K., y Niemann, D. (2013). When Do Numbers Count? The Differential Impact of the PISA Rating and Ranking on Education Policy in Germany and the US. *German Politics*, 22(3), 314–332.
- Ministerio de Educación. (2016). *¿Cuánto aprenden nuestros estudiantes al término de la educación primaria? Informe de logros de aprendizaje y sus factores asociados en la Evaluación Muestral 2013*. Minedu.
- Rotalde, D. (2023). *Desde el corazón de la educación rural* | Penguin Libros. Debate.
- Unesco/Santiago. (2007). *Educación de calidad para todos: Un asunto de derechos humanos*. Unesco. <http://unesdoc.unesco.org/images/0015/001502/150272s.pdf>
- Ydesen, C. (Ed.). (2019). *The OECD’s Historical Rise in Education. The Formation of a Global Governing Complex*. Global Histories of Education. Palgrave Macmillan.

- Zumbo, B. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-type (Ordinal) Item Scores*. Directorate of Human Resources Research and Evaluation, Department of National Defense.
- Zumbo, B. (2007). Three Generations of DIF Analyses: Considering Where It Has Been, Where It Is Now, and Where It Is Going. *Language Assessment Quarterly*, 4(2), 223–233. Taylor & Francis Group.